UNIVERSIDADE NOVE DE JULHO – UNINOVE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
MARÍA ALEJANDRA GUERRERO HERNÁNDEZ

INVESTIGAÇÃO DO COMPORTAMENTO DE UM SISTEMA DE TRANSPORTE EM MASSA USANDO SIMULAÇÃO DE EVENTOS DISCRETOS E DELINEAMENTO DE EXPERIMENTOS

São Paulo
2014

MARÍA ALEJANDRA GUERRERO HERNÁNDEZ

INVESTIGAÇÃO DO COMPORTAMENTO DE UM SISTEMA DE TRANSPORTE EM MASSA USANDO SIMULAÇÃO DE EVENTOS DISCRETOS E DELINEAMENTO DE EXPERIMENTOS

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Nove de Julho – UNINOVE, como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção.
Orientador: Prof. Dr. André F. Henriques Librantz

São Paulo
2014

Hernández, María Alejandra Guerrero.

Investigação do comportamento de um sistema de transporte em massa usando simulação de eventos discretos e delineamento de experimentos. /María Alejandra Guerrero Hernández. 2014.

160 f.

Dissertação (mestrado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2014.

Orientador (a): Prof. Dr. André Felipe Henriques Librantz.

UNINOVE
Universidade Nove de Julho

São Paulo, 24 de janeiro de 2014.

TERMO DE APROVAÇÃO

Aluna: María Alejandra Guerrero Hernández

Título da Dissertação: INSVESTIGAÇÃO DO COMPORTAMENTO DE UM SISTEMA DE TRANSPORTE EM MASSA USANDO SIMULAÇÃO DE EVENTOS DISCRETOS E DELINEAMENTO DE EXPERIMENTOS.

Presidente: PROF. DR. ANDRÉ FELIPE HENRIQUES LIBRANTZ_____

Membro: PROF. DR. MAURO DE MESQUITA SPINOLA_____

Membro: PROF. DR. FABIO HENRIQUE PEREIRA_____

Dedico este trabalho a meu marido Juan Carlos,
a minha mãe Melba e a meu pai José Ignacio.

AGRADECIMENTOS

A Deus, por me permitir atingir meus ideais.

Ao Prof. Dr. André Felipe Henriques Librantz; por todo o seu tempo investido na pesquisa, pela sua paciência e valiosas orientações.

A UNINOVE, pelo apoio financeiro.

A PROSUP-CAPES, pelo apoio financeiro.

Ao meu marido Juan Carlos, pela força e constante apoio, durante a realização desta pesquisa.

Aos meus pais, que mesmo de longe, estão sempre presentes no meu caminho.

All our science compared with the reality, is primitive and infantile
... and nevertheless it is the most valued thing that we have.
ALBERT EINSTEIN

RESUMO

O problema do planejamento das operações metroviárias é particularmente importante entre os sistemas de planejamento dos transportes, pois o crescimento acelerado das cidades pode não ser acompanhado linearmente pelas modernizações na infraestrutura. Este trabalho apresenta uma investigação do comportamento das variáveis: headway (tempo entre a saída dos trens desde a plataforma), tempo de abertura de portas e velocidade, envolvidas no funcionamento de um sistema em transporte em massa semelhante às linhas 1, 2 e 3 da cidade de São Paulo, usando simulação de eventos discretos. O trabalho teve como objetivo identificar o impacto dessas variáveis no tempo de viagem dos passageiros, no tamanho das filas nas plataformas, na ocupação do trem e no tempo de espera para embarcar. O modelo computacional desenvolvido no Software ProModel foi acompanhado de uma validação apoiada em técnicas quantitativas, o que permitiu uma precisão de 95% nos resultados do modelo. As variáveis foram analisadas com o apoio da metodologia de delineamento de experimentos. Este trabalho permitiu encontrar as variáveis críticas do sistema e a sua relação com o tipo de demanda da linha, o que pode auxiliar os tomadores de decisão na geração de novas estratégias de operações, com vistas a maior eficiência do transporte, sem a necessidade de investimentos na rede metroviária.

Palavras-chave: Sistema Metroviário, Simulação de Eventos Discretos, Validação de Modelos de Simulação, Delineamento de Experimentos.

ABSTRACT

The problem of operations planning is particularly important for subway transportation systems planning; the accelerated growth of the cities; cannot be linearly followed by the infrastructure. This research is based in the discrete event simulation, of the variables headway (time between the train's exit from the platform), time for opening doors and speed involve at a massive transportation systems operation like the lines 1, 2 and 3 of the city of São Paulo, using discrete events simulation. This research had the goal of (1) understand their impact on the travel time of passengers, (2) the size of the queues on platforms, (3) the occupation of the train and the (3) waiting time before boarding. The computational model developed with the software ProModel, was accompanied by quantitative' validation techniques, which ensure a statistical significance of 95% in the model results. The variables were analyzed with the basis of the methodology of design of experiments. This work allowed to find the most meaningful variables of the system and their relations to the type of demand line, this will lead to managers in taking decisions, it is possible to generate new strategies of operation in a short and medium term to improve transport efficiency, without the necessity in large investments within the infrastructure.

Keywords: Subway public transportation, Discrete Event Simulation, Simulation Model Validation , Design of Experiments

LISTA DE FIGURAS

Figura 1 - Metodologia para elaboração de Projetos de Simulação.....31
Figura 2 - Guia para Validação Operacional da Simulação.....35
Figura 3 – Método Boxplot.....42
Figura 4 - Histograma.....44
Figura 5 - Diagrama de probabilidade para distribuição normal.....46
Figura 6 - Igualdade de Médias comprovado com Anova.....52
Figura 7 - Diferença de Médias comprovada com Anova.....52
Figura 8 - Mapa com as Linhas Azul, Verde e Vermelha.....60
Figura 9 - Modelo Conceitual Processo de Transporte Metroviário.....62
Figura 10 - Modelo conceitual das variáveis de entrada e saída para um modelo de simulação de um sistema metroviário.....63
Figura 11 - Lógica da Programação em ProModel da sequência de atividades de um sistema metroviário.....72
Figura 12 - Diagrama de Causa – Efeito.....74
Figura 13 - Dados ajustados na Distribuição *Weibull* usando o Software Statfit.....81
Figura 14 - Resultado dos Testes Estatísticos obtidos com o Software Statfit.....82
Figura 15 - Análise de Normalidade dos dados.....84
Figura 16 - Resultados do Teste *Anova* em *Minitab*.....85
Figura 17 - Exemplo da metodologia de validação por animação.....88
Figura 18 - Gráfico de Probabilidade Normal para estação 1.....93

LISTA DE TABELAS

Tabela 1 - Intervalo de confiança considerando o número de observações 41
Tabela 2 - Distância em metros entre as estações das Linhas analisadas 64
Tabela 3 - Demanda de Passageiros 65
Tabela 4 - Matriz origem destino dos passageiros 65
Tabela 5 - Matriz de experimentação, experimento fatorial rotacional 2^k 75
Tabela 6 - Valores definidos para os fatores de controle 76
Tabela 7 - Cálculos em Excel para detecção de *outliers* 79
Tabela 8. Entrada de Passageiros: número de dados resultantes depois da eliminação de outliers 79
Tabela 9 - Cálculos em Excel para detecção de *outliers* 80
Tabela 11 - Agrupamento das Estações por tempo de abertura das portas usando teste de *Tukey* 86
Tabela 12 - Distribuição Estatística Abertura de Portas nas Estações 87
Tabela 13 - Entradas e saídas do sistema 88
Tabela 14 - Taxa de Ocupação dos trens da Linha Azul em horário vale 89
Tabela 15 - Taxa de Ocupação dos trens da Linha Vermelha 90
Tabela 16 - Taxa de Ocupação dos trens da Linha Verde 90
Tabela 17 - Tempo Médio de Espera e Tamanho da Médio da Fila por Linha 90
Tabela 18 - Cálculo do Tempo da Viagem Real 91
Tabela 19 - Erro no tempo da volta Real vs Simulação 91
Tabela 20 - Cálculo do Tempo de Meia Volta Teórico 92
Tabela 21 - Erro no tempo da Meia volta Teórico vs Simulação 92
Tabela 22 - Estações onde a $H0$ é aceita = Os dados simulados e os dados reais têm variâncias similares 93
Tabela 23 - Estações onde a $H0$ foi rejeitada e confirmada a $H1$= Não existe nenhuma evidência da igualdade entre as variâncias 94
Tabela 24 - Teste de Hipóteses, comparação das médias com dados reais e simulados usando Teste t de *student* 96
Tabela 25 - Taxa de Ocupação SAHP dos trens da Linha Azul 97
Tabela 26 - Taxa de Ocupação SAHP dos trens Linha Vermelha 98
Tabela 27 - Taxa de Ocupação SAHP dos trens da Linha Verde 98
Tabela 28 - Normalização dos dados da Linha Azul 99

Tabela 29 - Normalização dos dados da Linha Verde 99
Tabela 30 - Normalização dos dados da Linha Vermelha 99
Tabela 31 - Matriz de experimentação com valores, experimento fatorial rotacional 2k 99
Tabela 32 - Classificação da Demanda de Passageiros 100
Tabela 33 - Resultado dos Cenários Simulados na Linha Azul 101
Tabela 34 - Resultado dos Cenários Simulados na Linha Vermelha 102
Tabela 35 - Resultado dos Cenários Simulados na Linha Verde 103
Tabela 36 - Fatores que afetam o Tempo de Espera da Linha Verde 104
Tabela 37 - Fatores que afetam o Tempo da Viagem da Linha Verde 105
Tabela 38- Fatores que afetam a Lotação da Linha Verde 105
Tabela 39 - Fatores que afetam o Tamanho da Fila da Linha Verde 106
Tabela 40 - Fatores que afetam o Tempo de Espera na Linha Azul 106
Tabela 41 - Fatores que afetam o Tempo da Viagem na Linha Azul 107
Tabela 42 - Fatores que afetam a Lotação na Linha Azul 108
Tabela 43 - Fatores que afetam o Tamanho da Fila na Linha Azul 108
Tabela 44 - Fatores que afetam o Tempo de Espera da Linha Vermelha 109
Tabela 45 - Fatores que afetam o Tempo da Viagem da Linha Vermelha 110
Tabela 46 - Fatores que afetam a Lotação da Linha Vermelha 110
Tabela 47 - Fatores que afetam o Tamanho da Fila da Linha Vermelha 111
Tabela 48 – Identificação das variáveis críticas obtidas a partir da análise de cenários usando ANOVA 111

SUMÁRIO

1. INTRODUÇÃO 16
1.1. JUSTIFICATIVA E CONTRIBUIÇÕES 17
1.2. OBJETIVOS 18
1.2.1. Objetivo Geral 18
1.2.2. Objetivos Específicos 19
1.3. DELIMITAÇÃO DO ESTUDO 19
1.4. REVISÃO DA LITERATURA 19
1.5. TRABALHOS PUBLICADOS RESULTANTES DA DISSERTAÇÃO 24
1.6. ESTRUTURA DA DISSERTAÇÃO 24
2. FUNDAMENTAÇÃO TEÓRICA 26
2.1. SISTEMAS DE TRANSPORTE EM MASSA 26
2.2. SIMULAÇÃO COMPUTACIONAL 26
2.2.1. Simulação de Eventos Discretos 27
2.3. SISTEMAS E MODELOS 27
2.4. SOFTWARES DE SIMULAÇÃO 29
2.5. VANTAGENS E DESVANTAGENS DA SIMULAÇÃO 29
2.6. PROCESSO DE SIMULAÇÃO 30
2.6.1. Objetivos e Definição do Problema 32
2.6.2. Construção do Modelo Conceitual 32
2.6.3. Validação do Modelo Conceitual 32
2.6.4 Modelagem dos Dados de Entrada 32
2.6.5. Construção do Modelo Computacional 33
2.6.6. Verificação do Modelo Computacional 33
2.6.7. Validação do Modelo Operacional 33
2.6.8. Definição do Delineamento de Experimentos 35
2.6.9. Execução dos Experimentos 36
2.6.10. Análise Estatística 36
2.7. ESTRUTURA BÁSICA DA SIMULAÇÃO 36
2.8. TRATAMENTO DOS DADOS 37
2.8.1. Coleta e Análise dos Dados 37
2.8.2. Amostragem 38

2.8.3. Dimensionamento da amostra 38
2.8.4. Composição da Amostra 39
2.8.5. Outliers 40
2.8.6. Inferência Estatística 44
2.8.7. Delineamento de Experimentos (DOE) 55
3. MATERIAIS E MÉTODOS 58
3.1. DESCRIÇÃO DO MODELO CONCEITUAL 59
3.1.1. Variáveis de Entrada do Sistema 62
3.1.2. Variáveis de Saída do Sistema 67
3.2. TRATAMENTO DOS DADOS 68
3.3. CONSTRUÇÃO DO MODELO DE SIMULAÇÃO 69
3.3.1. Lógica do Modelo de Simulação 71
3.4. VALIDAÇÃO DO MODELO DE SIMULAÇÃO 73
3.5. DEFINIÇÃO DOS CENÁRIOS 73
3.6. REALIZAÇÃO DOS EXPERIMENTOS 74
3.6.1 Matriz de Planejamento 74
4. RESULTADOS E DISCUSSÃO 77
4.1. COLETA DE DADOS 77
4.1.1 Tempo de abertura das portas 77
4.2. OUTLIERS 78
4.2.1. Entrada de Passageiros 78
4.2.2. Abertura das Portas 80
4.3. ANÁLISE ESTATÍSTICA DOS DADOS 80
4.3.1. Entrada de Passageiros 81
4.3.2. Abertura de Portas 84
4.4. VALIDAÇÃO DO MODELO 87
4.4.2. Validação Face-to-Face 88
4.5. VALIDAÇÃO POR EVENTOS 91
4.6. VALIDAÇÃO ESTATÍSTICA 92
4.6.1 Parâmetro Entrada de Passageiros 92
4.6.2. Validação Preditiva 97
4.7. DELINEAMENTO DE EXPERIMENTOS 98
4.7.1. Análise de Cenários – Visão Geral 100

4.7.2. Análise de Cenários – Usando *ANOVA* 103
5. CONCLUSÃO 113
5.1. SUGESTÕES PARA TRABALHOS FUTUROS 114
6. REFERÊNCIAS BIBLIOGRÁFICAS 116
7. ANEXOS 122
8. APÊNDICES 126

1. INTRODUÇÃO

Os sistemas de transporte de passageiros em massa têm se tornado uma solução para ajudar no deslocamento de pessoas, especialmente nas grandes cidades e metrópoles urbanas. Estes sistemas ajudam a melhorar a qualidade de vida das pessoas que o utilizam, já que permitem chegar em menor tempo possível em um destino, além de reduzir os transtornos causados pelo trânsito e congestionamento nas cidades ademais das emissões de carbono (GIANNETTONI; SAVIO, 2004).

De acordo com os dados publicados no site oficial da Companhia do Metropolitano de São Paulo, o metrô de São Paulo transporta a cada dia em média 2,742 milhões de passageiros; em alguns horários do dia, as estações e os carros do metrô e do trem encontram-se superlotados, dificultando o acesso dos passageiros aos carros, causando acidentes e aumentando o tempo de transporte e espera. O metrô de uma cidade deve ser um meio completamente democrático, oferecendo as melhores condições para os passageiros e deve ser considerado uma alternativa viável para qualquer um que precise chegar a seu destino.

A eficiência do transporte é um requisito indispensável para garantir a mobilidade urbana, tornando-se crítico nas principais cidades e metrópoles devido a grande concentração da população. Um dos maiores desafios no planejamento do transporte público consiste em assegurar um sistema operacional e economicamente viável, que seja adequado e que esteja integrado ao entorno (DHINGRA; KRISHNA; MURALIDHAR, 2000).

Zhao e Zeng (2008), afirmam que, para um sistema de transporte ser uma alternativa viável para os passageiros, este deve fornecer aos usuarios um tempo razóavel de viagem e conveniência. Porém os orçamentos disponíveis para estes propósitos têm um significante efeto no nível e qualidade dos serviços. Segundo os mesmos autores, uma adequada configuração da rede pode ajudar tanto na redução nos custos de serviço, quanto ao aumento da qualidade do transporte.

De acordo com Marinov et al. (2014) dentro das últimas tendências na área de transportes está o crescente interesse pela modernização da infraestrutura ferroviária, considerada como uma consequência lógica e progressiva para cidades com um futuro mais sustentável. No entanto, os custos e tempos de implementação destas modernizações, embora viáveis, podem apresentar resultados somente a médio e longo prazo. Neste contexto, a simulação de eventos discretos é uma importante ferramenta para apoiar no planejamento e desenvolvimento de um sistema metroviário, operacional e economicamente eficiente e

sustentável, pois, a criação de um modelo computacional alimentado com dados estatísticamente coletados e validados, permitirá replicar a realidade com uma alta significância estatística e testar com um alto nível de confiança, com base em novas estratégias para adaptar o sistema a curto prazo ou com novas infraestruturas a médio ou longo prazo, tudo isso, sem impactar a segurança do sistema.

Os resultados desta simulação tornam-se mais potentes quando acompanhados de técnicas de planejamento de cenários como o Delineamento de Experimentos (DOE), cuja metodologia permite a geração de um número ótimo de experimentos e o estudo do comportamento das diferentes variáveis que compõem o sistema metroviário evitando técnicas de ensaio e erro.

1.1. JUSTIFICATIVA E CONTRIBUIÇÕES

O crescimento das cidades tem levado à necessidade de procurar novas estratégias para transportar os cidadãos de maneira cada vez mais eficiente e economicamente rentáveis para as empresas. Os sistemas sobre trilhos, por exemplo, são uma boa alternativa para transportar grandes volumes de pessoas entre pontos específicos das cidades.

Em 2011, o Metrô de São Paulo foi considerado o melhor sistema de transporte sobre trilhos da América Latina pelo *The Metro Awards*. Porém, no ano 2012, embora tenha ficado entre os finalistas, perdeu o seu lugar frente ao metrô da cidade de Santiago de Chile.

Segundo uma reportagem publicada pela Folha Online em abril de 2011, com dados extraídos da CoMet (Comunidade dos Metrôs), o metrô paulistano foi declarado como o mais lotado do mundo. Por sua vez, Cardoso et al. (2012) na sua pesquisa, comprovaram que a lotação em horário de pico chegou a atingir 11,4 passageiros por m², sendo que os valores recomendados internacionalmente são quatro pessoas por metro quadrado.

Tendo em vista estas informações, o presente estudo pretende analisar o complexo problema do planejamento das operações metroviárias, apoiando-se na simulação de eventos discretos para a criação de um modelo computacional que replique a realidade e o uso da metodologia de delineamento de experimentos, para uma definição objetiva de novos cenários simulados.

Este trabalho é relevante do ponto de vista operacional, já que permite definir o impacto da mudança dos parâmetros velocidade, headway, tempo de abertura das portas no compartamento da operação para diferentes níveis de demanda.

Do ponto de vista econômico, os custos envolvidos na modernização da infraestrutura de um sistema de transporte em massa podem ser muito altos, sendo assim, esta pesquisa propõe uma nova parametrização das variáveis envolvidas no processo, para que seja possível melhorar os indicadores do sistema sem a necessidade de novos investimentos.

Em uma perspectiva acadêmica, o transporte em massa nas grandes metrópoles urbanas é um problema complexo com múltiplas variáveis, o qual, embora amplamente abordado em vários estudos propõe um estudo a aplicação das metodologias de simulação de eventos discretos e DOE para a análise dos indicadores representativos do sistema, usando para o desenvolvimento uma importante componente de validação do sistema, mesmo que permite ter um modelo validado objetivamente, usando várias das metodologias de validação encontradas na literatura.

São poucas as abordagens existentes nas pesquisas que tratam desta união, mesmo porque o conhecimento sobre este assunto está longe de permitir o estabelecimento de conclusões definitivas que permitam guiar os pesquisadores na solução de problemas complexos, usando a metodologia de DOE para geração de cenários em problemas de transporte.

Nesse contexto, o modelo pode ser usado como referência para futuros trabalhos de modernização das linhas e como ponto de partida científica para analisar o comportamento e aplicabilidade das técnicas de simulação de eventos discretos e DOE nesses tipos de problemas.

1.2. OBJETIVOS

Os principais objetivos desta pequisa serão apresentados a seguir.

1.2.1. Objetivo Geral

O objetivo geral deste trabalho é investigar a partir de um modelo de simulação, o comportamento das variáveis envolvidas no transporte de passageiros, em um sistema de transporte em massa semelhante ao metrô de São Paulo, visando analisar os efeitos no tempo, das variáveis tempo de viagem dos passageiros, lotação dos vagões, tempo de espera e/ou tamanho das filas, com base na análise e no ajuste sistêmico dos indicadores associados à demanda no dia e no horário.

1.2.2. Objetivos Específicos

a) Realizar uma análise estatística das principais variáveis de entrada de um sistema de transporte em massa a fim de definir padrões, encontrar distribuições de probabilidade e usá-las em um modelo de simulação que represente com um alto nível de confiança o sistema real;
b) Analisar as filas e o tempo de espera dos passageiros;
c) Desenvolver um modelo de simulação da situação atual na linguagem de programação *ProModel*;
d) Usar a metodologia de delineamento de experimentos (DOE) a fim de definir estrategias que permitam melhorar os indicadores que determinam a eficiência de um sistema de transporte em massa (lotação, tempo de viagem, tempo de espera ou tamanho da fila) visando melhorar o serviço das operações.

1.3. DELIMITAÇÃO DO ESTUDO

O presente estudo foi restrito à análise das operações metroviárias para três linhas de metrô similares ao metrô São Paulo, porém os resultados podem ser aplicados em qualquer outro sistema metroviário.

O modelo de simulação foi desenvolvido visando reduzir o tempo de viagem dos passageiros e a lotação dos vagões, utilizando os recursos existentes e sem a necessidade de investimentos em novos equipamentos e infraestrutura. Desta forma, duas premissas foram definidas: não foram adicionadas novas estações à rede atual e não foram consideradas novas ferrovias.

1.4. REVISÃO DA LITERATURA

Ao longo do tempo, o problema de otimização do transporte público tem recebido grande atenção, principalmente no que se refere à redução do tempo médio de viagem dos passageiros, do tempo médio de espera nas estações e a otimização do fluxo de passageiros, buscando assim melhorar o conforto dos usuários e minimizar os custos operacionais gerais das empresas transportadoras.

As pesquisas mostram que as principais metodologias usadas nestes trabalhos foram a simulação de eventos discretos e a otimização por meio de diferentes técnicas heurísticas. Na

sequência serão apresentados alguns dos trabalhos desenvolvidos nesta área do conhecimento.

A primeira metodologia encontrada na literatura para a definição do sequenciamento de linhas de metrô foi prosposta por Cury, Gomide e Mendes (1980), o modelo representava o trem e o movimento dos passageiros por meio do somatório ponderado de funções lineares e não lineares, que simbolizam o tempo de espera do passageiro na estação, o conforto dentro dos vagões, o tempo de viagem e o número de trens em serviço. Considerando todas as restrições do sistema, o problema foi formulado com base na programação não linear e resolvido com o método iterativo de decomposição de vários níveis hierárquicos. Os autores restringem a análise para uma sequência de 15 trens, limitando-a a um pequeno domínio de dados, pois o número de trens por dia pode chegar até 400.

Para Higgings, Ferreira e Kozan (1996), a satisfação do cliente é uma variável composta de vários atributos que incluem o tempo total da viagem, a pontualidade e a confiabilidade na chegada dos trens. A confiabilidade é um item que considera vários fatores: o grau de liberdade relacionado à programação, o número e a posição dos trens, as prioridades de cada trem, o tempo de permanência nas estações e a velocidade dos trens. Os autores propõem um modelo de sequenciamento das operações procurando minimizar o tempo de viagem dos trens e maximizar a sua confiabilidade. Para isto apoiam-se em técnicas de otimização como: *Branch and Bound*, Algoritmos Genéticos, Busca Tabu e Heurística do *Nearest Neighbour* e a Variável Risco de Atraso, que é usada como função objetivo, (sendo que o risco de atraso constitui um dos parâmetros da função objetivo), e calculada com a distribuição de probabilidade de eventos imprevistos.

Assis e Melani (2000) usando o modelo de tráfego e todas as restrições consideradas em Cury, Gomide e Mendes (1980), propõem um novo modelo para geração de uma eficáz programação de trens por meio de um novo modelo de programação linear. Os autores ampliam o conjunto de restrições usadas por Cury, Gomide e Mendes (1980), devido as exigências de continuidade do fluxo nos terminais durante a operação comercial. Ambos propõem ainda um sequenciamento para um número *n* de trens, fazendo o modelo mais eficiente do que o proposto pelos predecessores da pesquisa, no entanto o sequenciamento considera períodos restritos e a demanda de passageiros é uma constante.

Li (2000) desenvolveu um modelo de simulação para analisar as operações que acontecem dentro de uma estação de trem, com o objetivo de entender o desenho das operações e o fluxo de passageiros para apoiar a tomada de decisões. Para isto, estudou os serviços prestados em uma estação de metrô: máquinas de venda automática de bilhetes,

pontos de aceitação do bilhete como catracas de saída e entrada, tempo de fila, e gargalos. Parâmetros como a entrada de passageiros foram simulados usando a distribuição de Poisson e as programações de chegada com base no histórico das modalidades pelas quais as pessoas chegam na estação, como por exemplo, caminhando, de taxi, de carro, de metrô/transferência interna, de ônibus, etc,. Nos parâmetros de entrada como *headway*, no número de trens por hora e de pessoas por trem, são usados valores constantes. As principais variáveis de saída para análise da eficiência são: número de pessoas que perdem o trem por demora nas filas, a utilização dos equipamentos e tempos de filas e serviço. O processo de validação é realizado usando alguma técnica objetiva, porém não é detalhada nem explicada e tampouco se conhece o nível de confiança do modelo. Os cenários são criados com técnicas de ensaio e erro e embora tenham apresentado um bom resultado, aumentam o risco de deixar de avaliar outros cenários que pudessem apresentar melhores resultados.

Assis e Milani (2004) abordaram o problema de sequenciamento das operações das linhas de metrô usando programação linear por meio da formulação de um modelo preditivo de controle. A aplicação desta metodologia foi feita em uma linha de metrô similar à linha norte-sul do metrô de São Paulo, e considera-se como restrições da operação o *trade-off* entre o custo operacional e a qualidade do serviço (conforto do passageiro). O conjunto de equações dinâmicas está reduzido em: equações de *headway* e equações de ingresso de passageiros. As primeras consideram o intervalo de saída entre um trem e outro, o tempo do trem na plataforma e o tempo de deslocamento do trem entre uma estação e outra. Os dois conjuntos de equações estão baseados na pesquisa de Cury, Gomide e Mendes (1980). Os resultados são apresentados desde o ponto de vista econômico, visão que traz aumento na lotação dos vagões e na redução do número de trens em serviço.

Para Hu et al. (2005), uma rede de transporte em massa é composta de quatro fatores principais: desenho da rede de transporte, sequenciamento das operações e determinação do *headway* (tempo entre a saída dos trens). Na sua pesquisa, os autores propõem um modelo para melhorar o fluxo dos passageiros sem transferência, otimizando o tempo de *headways* para todas as rotas da rede. O problema é abordado por meio de duas metodologias: Algoritmos Genéticos e Colônia de Formigas.

Zhao (2006) descreve um método de busca metaheurístico para otimizar tanto o layout da rede de trânsito, quanto o *headway*, usando como base a função de custo, este método propõe uma busca local usando um algoritmo para a busca do *Headway* ótimo.

Zhao e Zeng (2006) usando Algoritmos Genéticos e Simmulated Annealig propuseram um melhoramento na pesquisa de Zhao (2006) os autores usam uma demanda

estática de passageiros, além de assumida e não são considerados desenhos de horários de demanda.

Guan, Et.al (2006), apresentam um modelo otimização simultâneo da configuração das linhas de transporte e da alocação do passageiro nas redes gerais. O modelo é formulado usando programão linear binária inteira e resolvido usando o método *padrão branch and bound*. O modelo produz configurações sensíveis aos três parâmetros considerados como decisivos para o balanceamento da línea, estes são custos de operação, transferência de passageiros e distância da viagem. O estudo é aplicado no sistema metroviário da Hong Kong.

Posteriormente, Zhao e Zeng (2008), apresentam uma metodologia para otimização de rotas, headway e horários através das metodologias de Simmulated Annealig e o método de busca ISTG (método baseado no algoritmo greedy search (GS) desenvolvido por Zhao and Gan (2003) e a metodologia de Busca Tabu. A função objetivo representa cálculo do custo mínimo de uso da rede metroviária.

Yalçinkaya e Mira (2009), apresentam um modelo e uma solução para o parâmetro tempo de viagem dos passageiros usando a simulação de eventos discretos e a metodologia de otimização multi-resposta de Derringer–Suich. A pesquisa focou-se principalmente na busca de eficazes *headways* que permitam melhorar o tempo médio de viagem para atender a taxa de lotação máxima permitida. O modelo foi desenvolvido no software de simulação de eventos discretos *ARENA*, e a validação feita por meio de depuradores interativos, utilizando a substituição de variáveis contantes para aleatórias e da animação do sistema. Para o desenvolvimento do modelo, os autores usaram dados da demanda de passageiros e das falhas do trem presentes na documentação do metrô. A capacidade dos vagões foi assumida, a velocidade entre uma estação e outra considerada uma constante, e o *headway* é fixo para um mesmo intervalo de tempo. Os cenários foram desenvolvidos com base no planejamento fatorial de dois fatores completos e duas respostas são consideradas: tempo de viagem e taxa de ocupação. Os cenários são repricados 10 vezes para cada ponto e uma vez para o ponto central, sendo o *headway* o fator de variação. Finalmente uma otimização via multi-resposta de Derringer-Suich é aplicada, usando a ferramenta *Response Optimizer Tool* do *Minitab*. A variável principal de entrada é o *headway* e as variáveis de saída são o tempo de viagem e as taxas de lotação. A metodologia desenvolvida pelos autores apresenta um interessante procedimento para melhorar os indicadores de transporte metroviário, porém, a validação subjetiva do modelo de simulação pode invalidar os resultados pois não é apresentada nenhuma evidência da confiança do modelo.

Peng et.al (2009) realizaram um levantamento e determinação do tempo de deslocamento dos passageiros dentro das estações do metrô de Pequim, este trabalho tinha como objetivo realizar a coordenação entre as linhas de trem e reduzir o tempo de espera e transferência dos passageiros. O trabalho é centrado na elaboração de uma análise estatística do tempo de deslocamento e transferência dos passageiros. Para a análise, o método de amostragem foi concebido com base na análise de fatores que comprometem o tempo de deslocamento dentro das cinco estações de transferência ao longo da linha 5 do metrô de Pequim, em seguida, a distância equivalente a transferência é concebida considerando a distância real da passagem e a infraestrutura da estação. Finalmente, a análise de regressão de dados de 10 passagens de transferência demonstra que o valor esperado de deslocamento do passageiro é quádratico à distância equivalente à passagem.

Mussone e Calvo (2013) desenvolveram um modelo de otimização usando o Método Simplex, para calcular a capacidade ferroviária, isto é, o número máximo de trens que podem atravessar toda a estrada de ferro em um determinado período de tempo, sujeito a restrições de gestão (tais como a capacidade de junção, a capacidade de pista, a capacidade da linha e da probabilidade de interferência entre os trens). Para desenvolver o modelo a diferencia dos outros autores, não é baseado em horários e, além disso, pode indiretamente levar em conta as prioridades entre os trens e possíveis atrasos, é dar a porcentagem de ocupação do tempo em cada componente para cada categoria do sistema. Foi aplicada a um estudo de caso na linha de Bern-Thun está localizado no meio da rede de transportes Suíça ao longo da linha norte-sul ligando as cidades de Basel com Bern, o qual demonstrou a sua validade para aplicação para qualquer sistema ferroviário.

Nesse contexto, esta pesquisa esta motivada pelo fato principal, que, embora várias metodologias para solucionar o problema de sistema de transporte em massa tenham sido reportadas na literatura, ainda existe uma lacuna em relação à utilização da simulação de eventos discretos, acompanhada de delineamento de experimentos, como metodologia sistemática para resolver problemas de transporte. Além disto, os estudos em simulação encontrados na literatura, não fornecem evidências claras do uso de métodos de validação estatística que permitam dar um alto nível de confiança aos dados. Igualmente, para os autores pesquisados, o principal parâmetro de análise para melhoramento do sistema de transporte foi o *headway*, não foram encontradas evidências de análises, do impacto do tempo do trem na plataforma e da velocidade (dentro dos limites técnicos), nas variáveis de saída: tempo de viagem dos passageiros e lotação dos vagões.

1.5. TRABALHOS PUBLICADOS RESULTANTES DA DISSERTAÇÃO

- GUERRERO M. A.; LIBRANTZ, A. F. Reduction of the inventory in a packaging industry using discrete events simulation. In: 7th International Conference on Industrial Engineering and Industrial Management, 2013, Valladolid. **Proceedings of the ICIEOM 2013**, 2013. v. 1. p. 1-10.

- GUERRERO M. A.; BATISTELO R. A.; LIBRANTZ, A. F. Maximizing Profits with the Improvement in Product Composition Using Genetic Algorithms and K-Means Application to a Company of the Printing Industry. In: International Conference on Industrial Engineering and Operations Management, 2012, Guimarães. **Proceedings of the ICIEOM 2012**, 2012. v. 1. p. 1-11.

- GUERRERO M. A.; LIBRANTZ, A. F. Improvement of the supply chain for the sugar cane exportation process employing discrete events simulation techniques. In: **Acta Scientiarum. Technology**, 2013, v. 35, n. 4, p. 637-643.

- GUERRERO M. A.; LIBRANTZ, A. F. Discrete event simulation of the commodities' supply chain exportation. **Revisa Chilena de Ingeniería**, 2013.

1.6. ESTRUTURA DA DISSERTAÇÃO

Esta dissertação foi estruturada em cinco capítulos:

1. O primeiro capítulo apresenta a introdução, a justificativa, os objetivos, a delimitação do estudo, a revisão da literatura, o método de pesquisa, os trabalhos publicados resultantes da pesquisa, e a estrutura da dissertação
2. O segundo capítulo da dissertação descreve a fundamentação teórica apresenta conceitos de simulação, validação de modelos de simulação, técnicas de inferência estatística e delineamento de experimentos.
3. O terceiro capítulo da dissertação descreve os materiais e métodos aplicados no trabalho. Neste capítulo apresenta-se a descrição do modelo conceitual, as variáveis e parâmetros usados dentro da simulação, os métodos usados para a validação do modelo e as matrizes definidas para a criação dos cenários de simulação.

4. O quarto capítulo da dissertação apresenta os resultados da simulação e discussão. Os resultados são divididos de acordo com as carateristicas das linhas de metrô analisadas e os parâmetros investigados.
5. O quinto capítulo da dissertação apresenta a conclusão do trabalho.
6. O sexto capítulo da dissertação apresenta as referências bibliográficas.

Além destes capítulos, fazem parte deste trabalho o anexo A que apresenta as explicações de distribuições estatísticas, e os apêndices A, B, C, e D onde são apresentados os trabalhos publicados resultantes desta dissertação.

2. FUNDAMENTAÇÃO TEÓRICA

No presente capítulo será realizada uma definição dos principais conceitos trabalhados nesta pesquisa, com objetivo de fundamentar o modelo e todas as metodologias aqui utilizadas.

2.1. SISTEMAS DE TRANSPORTE EM MASSA

Um sistema de transporte em massa é composto de variáveis e componentes específicos do sistema, por tanto, para contextualizar nos termos principais usados na pesquisa, a continuação serão explicados os principais conceitos:

- Linha de Metrô, é a unidade menor de um sistema metroviârio, é um conjunto de trens, que se deslocam através de uma única rede metroviária com varios kilometros de comprimento. (MUSSONE; CALVO, 2013).
- Pátios, é o lugar onde ficam estacionados os trens, neste lugar pode acontecer a manutenção equipamento, e o câmbio de direção do trem.
- Trem, equipamento que realiza o transporte dos passageiros de uma estação para outra, um trem pode estar composto de vários carros, porém para o presente estudo é considerado o trem como a unidade mínima de medida, e a sua capacidade, é o resultado do somatorio da capacidade dos carros que o compõem.
- Estação, é o lugar onde acontece o embarque e desembarque dos passageiros do trem, é o ponto onde o trem estaciona e permite estas transferências.
- *Headway*, é o tempo entre a saída dos trens desde os pátios, é um dos principais fatores encontrados na literatura, que comprometem o tempo da viagem dos passageiros .(ASSIS; MILANI, 2004).

2.2. SIMULAÇÃO COMPUTACIONAL

A origem da Simulação Computacional está na segunda guerra mundial, quando os matemáticos Stan Ulam e John Von Neumann, precisavam resolver um problema matemático baseado em nêutron. Na sua pesquisa criaram um novo método chamado Simulação de

Montecarlo e desde então esta metodologia tem sido usada em infinitas áreas: econômica, empresarial, industrial e social.

O avanço no desenvolvimento da computação trouxe inovações igualmente importantes na área de tomada de decisões e no desenho e desenvolvimento de novos processos e produtos. A simulação computacional é uma das principais técnicas usadas nesta área do conhecimento.

2.2.1. Simulação de Eventos Discretos

Obaidat e Papadimitriou (2003) descrevem a simulação de eventos discretos como a imitação da operação de um sistema ou processo do mundo real, sobre o qual são realizadas experimetações através de um modelo construido baixo certas políticas de estudo, tudo com o objetivo de determinar oportunidades de melhora e definir caraterísticas de desenho de processos. Por sua parte, Frazelle (2002), define que a simulação de eventos discretos, permite valorar, analisar e escolher as melhores alternativas de desenho, pois está focada na programação orientada objetos, possue capacidade para apresentar os modelos operando em conjunto e representar a variabilidade real que existe nos processos.

2.3. SISTEMAS E MODELOS

Harrel et al. (2000), definem um sistema como um conjunto de múltiplos elementos inter-relacionados que trabalham conjuntamente para alcançar um objetivo comum. Law e Kelton (1982), por sua vez, definem sistema como um conjunto de entidades ou unidades que age e interage entre si, com a finalidade de desempenhar algum objetivo lógico.

Pode-se dizer que um exemplo de sistema é uma estação de metrô, pois é composta por um conjunto de entidades que interage entre si, como por exemplo os passageiros, os postos de venda de bilhetes e os trens que chegam na estação.

A representação desse sistema pode ser feita por meio de um modelo, que para Bertrand e Fransoo (2010), é uma representação da realidade no sentido em que a realidade completa não pode ser incluída.

Pereira (2000), explica que o objetivo dos modelos é o de estudar os sistemas, analisar novas estratégias e estudar os resultados dessas modificações sem que haja necessidade de alterá-lo previamente.

Sendo assim se conclui que um modelo de simulação é um modelo matemático estudado com base no uso da simulação (LAW; KELTON, 2000; BANKS et al., 2005).

Para Shannon (1998), a simulação é o processo de projetar o sistema em um modelo computacional, a fim de realizar testes e experimentos permitindo assim compreender o seu comportamento diante de diferentes estratégias de operação.

Segundo Law (2006) o uso de modelos de simulação, permite substituir os experimentos que são utilizados no sistema real, aqueles que realmente existem ou estão em fase de projeto. Para muitos destes sistemas, a implementação de experimentos no sistema real pode ser econômicamente inviável ou impossível de ser executado.

Carson (2003) afirma que um Modelo de Simulação é um modelo descritivo de um processo ou sistema, o qual inclui parâmetros para representar diferentes configurações. Ainda segundo o autor, um modelo de simulação pode ser usado para experimentar, avaliar e comparar diferentes cenários alternativos do sistema.

Desta forma a simulação computacional pode ser definida como uma ferramenta de análise de sistemas complexos, a qual, por meio da representação funcional de um sistema real no computador, permite replicá-lo com alta precisão, permitindo implementar mudanças sem comprometer a integridade do sistema real, respondendo a questões do tipo “o que aconteceria se”. Esta flexibilidade permite definir experimentos que seriam impossíveis de testar no sistema real, trazendo importantes economias para as empresas e a possibilidade de compreender melhor o comportamento das variáveis. O uso da simulação se faz pertinente quando existem múltiplas variáveis afetando os processos e se faz necessário entender as interações entre os parâmetros e a significância estatística.

Existe uma caracterização importante para os modelos de simulação computacional, que podem ser classificadas como simulação discreta e contínua. Os modelos contínuos são aqueles nos quais as relações entre as variáveis relevantes, definem-se por meio de equações diferenciais, já que estas permitem conhecer o comportamento da variável em um tempo contínuo. Nos modelos discretos, as variáveis variam discretamente em pontos específicos do tempo simulado. Para Garcia, et al. (2006), a simulação de eventos discreto pode ser definida como o conjunto de relações lógicas, matemáticas e probabilísticas que integra o comportamento de um sistema quando acontece um evento determinado.

Complementando o estudo anterior, Law e Kelton (2000) afirmam que a simulação de eventos discretos consiste na criação de um modelo que evolui ao longo do tempo onde as variáveis simuladas mudam em pontos específicos do tempo.

Os modelos também são classificados em determinísticos e estocásticos. Os modelos determinísticos não tratam variáveis randômicas e as características operacionais envolvem relações exatas. Já em um modelo estocástico, pelo menos uma variável do sistema se comporta de forma probabilística.

2.4. SOFTWARES DE SIMULAÇÃO

O objetivo da simulação consiste em comprender, analisar e melhorar as condições relevantes do sistema e o uso de softwares facilita este processo, minimizando o tempo de tomada de decisão das empresas.

Shannon (1998) descreve que o objetivo dos pacotes ou softwares de simulação é reduzir os erros, facilitar o processo de programação e incrementar a flexibilidade do modelo, reduzindo a distância entre o modelo conceitualizado e a forma executável. Na atualidade existem inúmeros softwares de simulação, por isso Shannon (1998) os divide em duas categorias: a) Linguagens de propósitos gerais e b) Linguagens com próposito exclusivo de simulação.

Na primeira categoria estão os pacotes que podem resolver quase todos os problemas de simulação discreta (ARENA®, AweSim®, GPSS/H™, Simscript II.5®, Extend™ etc) e na segunda categoria estão os pacotes utilizados na simulação de sistemas de manufatura e problemas de manuseio de material (SimFactory, ProModel®, AutoMod™, Taylor II®, and Witness®, BPSimulator™, processModel™, SIMPROCESS® e Extend+BPR).

Para o desenvolvimento desta pesquisa a variante da simulação que será utilizada é a Simulação de Eventos Discretos e o software foi ProModel®.

A simulação de eventos discretos usando o ProModel® permite interações com outras ferramentas computacionais, como excel, etc, e sua integração com o Windows o torna um sistema amigável e simples de ser utilizado (HASAN; AL-HUSSEIN, 2010).

2.5. VANTAGENS E DESVANTAGENS DA SIMULAÇÃO

A simulação apresenta inúmeras vantagens quando comparada com modelos de análisis número rico. Dentre as principais encontra-se a facilidade de entendimento e a interação visual com o modelo. Ademais pode-se citar outras vantagens com base nos estudos de Shannon (1998):

- Permite obter uma visão geral da operação do sistema e avaliar os impactos das possíveis mudanças no desempenho;
- Melhora e facilita a compreensão do sistema e dos seus elementos;
- Facilita a experimentação de políticas e novas configurações no sistema;
- Permite avaliar potenciais economias e reduzir os custos de implementações errôneas;
- Permite controlar o tempo, pois é possível operar o sistema para vários meses ou anos, em poucos segundos, permitindo assim uma rápida visão no horizonte de tempo.

Ainda segundo Shannon (1998), as desvantagens da simulação são:

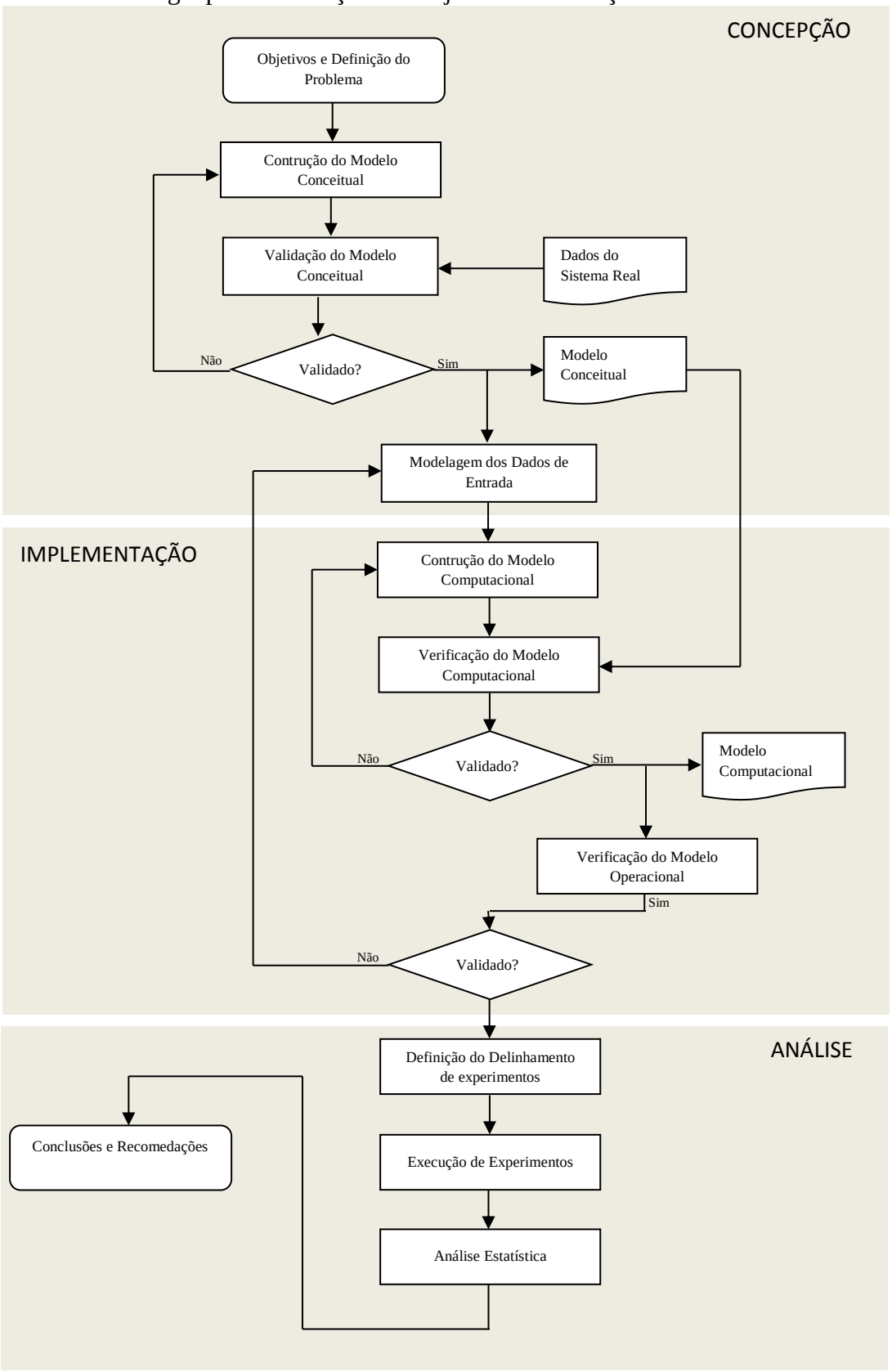
- A simulação é uma arte que requer um especialista altamente treinado e a utilidade do modelo depende da qualidade e as habilidades do modelador;
- A análise dos dados de entrada pode consumir muito tempo.

Finalmente, é importante destacar que a simulação por si mesma não entrega o melhor resultado pois depende da experiência do modelador e do conhecimento do processo para estabelecer corretamente os possíveis cenários de melhora, chegando em resultados subótimos e provavelmente limitando a análise, sem avaliar todas as possibilidades no espaço de busca.

2.6. PROCESSO DE SIMULAÇÃO

Montevechi et al. (2007) propõe uma metodologia para elaboração de projetos de simulação, que contém as etapas a serem seguidas num projeto de simulação proposta por Chwif (1999) e os passos para o delineamento de experimentos usando DOE, propostos por Montgomery (2008). Esta metodologia permite realizar uma análise mais aprofundada do sistema e ampliar as vantagens das duas propostas, isto, por meio de um modelo contextualizado e definido detalhadamente desde o momento da concepção, até a definição dos cenários, conforme o fluxo apresentado na figura 1. Cada uma das etapas da figura 1 é descrita na sequência.

Figura 1 - Metodologia para elaboração de Projetos de Simulação



Fonte: MONTEVECHI et al., 2007.

2.6.1. Objetivos e Definição do Problema

O objetivo desta fase é determinar o objeto de estudo baseado na poblemática observada no sistema, no qual são definidas as variáveis que influenciam o sistema. Para Banks et al. (2005) nesta etapa deve-se avaliar se a simulação é a metodología adequada para responder as questões definidas nos objetivos.

2.6.2. Construção do Modelo Conceitual

Para Law, Kelton (1991) e Shannon (1998), a construção do modelo conceitual é a fase mais importante do processo de simulação, pois vai garantir a correta definição dos parâmetros do sistema. Aqui, o sistema real objeto da pesquisa é resumido usando métodos gráficos, que são selecionados de acordo com as caraterísticas do sistema e os objetivos do estudo (LEAL et al., 2011).

De acordo com Harrel et al. (2000), o modelo conceitual é o resultado dos dados e informações coletadas, este modelo permite definir e entender o funcionamento do sistema real.

2.6.3. Validação do Modelo Conceitual

É realizada por meio da comparação entre o processo mapeado e o sistema real. O processo deve ser apresentado a especialistas e conhecedores do sistema.

2.6.4 Modelagem dos Dados de Entrada

É a etapa crítica do processo, na qual os dados devem ser coletados, analisados estatísticamente e ajustados numa distribuição de probabilidade. Esta etapa é considerada crítica porque a confiança e significância estatística dos resultados do modelo resultam da precisão dos dados introduzidos no sistema. Para uma análise estatística dos dados e o ajuste na distribuição de probabilidade, o pesquisador pode apoiar-se no uso dos pacotes estatísticos que acompanham o software de simulação, como o caso de *StatFit* para o Software *ProModel*. Outras ferramentas podem ser usadas: Excel, Minitab, SPSS.

2.6.5. Construção do Modelo Computacional

É a programação das rotinas do sistema em uma linguagem de simulação apropriada. Os componentes necessários para o desenvolvimento de um modelo de simulação também são expostos na seção 2.7.

2.6.6. Verificação do Modelo Computacional

É a fase na qual é avaliada a lógica de programação do modelo computacional, cujo objetivo é garantir que se seguem os passos lógicos descritos no modelo conceitual.

2.6.7. Validação do Modelo Operacional

É a confirmação de que o modelo trabalha apropriadamente e representa a realidade. De acordo com Banks et al. (2005), o processo de validação pode ser executado por meio de testes subjetivos e objetivos. Os testes subjetivos envolvem o conhecimento e juízo das pessoas que conhecem o modelo e estão focados nas entradas e saídas do sistema, no entanto, estes testes não fornecem cálculos que confirmem a precisão e confiança do modelo. Os testes objetivos por sua vez, resolvem estas falhas buscando a validação utilizando testes estatísticos.

Embora os métodos objetivos forneçam maior precisão nos resultados, os métodos de validação subjetivos são continuamente usados durante a fase de tradução do modelo.

Banks et al. (2005), apresenta os principais métodos de validação subjetiva. São eles:

- Animação: possível graças aos modernos softwares de simulação que permitem acompanhar os movimentos do sistema, durante os diferentes experimentos feitos com o modelo. É possivel ter os resultados de indicadores em tempo real, valores de tamanho da fila, tempo de espera, destino e tempo de viagem que podem ser validadas graficamente enquando o modelo está sendo executado;
- Validação por eventos: Os eventos e saídas do modelo são comparados com o sistema real para encontrar similaridades. Permite mensurar para cada evento as saídas do sistema (tamanho de filas, tempo de espera, etc.);

- Face-to-Face: Gestores e conhecedores do processo analisam a performance do sistema;
- Validação Interna: Várias réplicas do modelo são feitas para determinar a variabilidade do modelo. Uma alta variabilidade pode levar a suspeitar da confiança do modelo;
- Análise de Sensibilidade: As entradas no sistema são modificadas para entender o comportamento do modelo frente estas mudanças;
- Validação Preditiva: O modelo é usado para predizer comportamentos do sistema, os resultados são comparados com o sistema real;
- Teste de *Turing*: Resultados do modelo e do sistema real são apresentados para pessoas que conhecem o processo, para que possam definir quais os resultados que fazem referência ao sistema real e quais ao modelo.

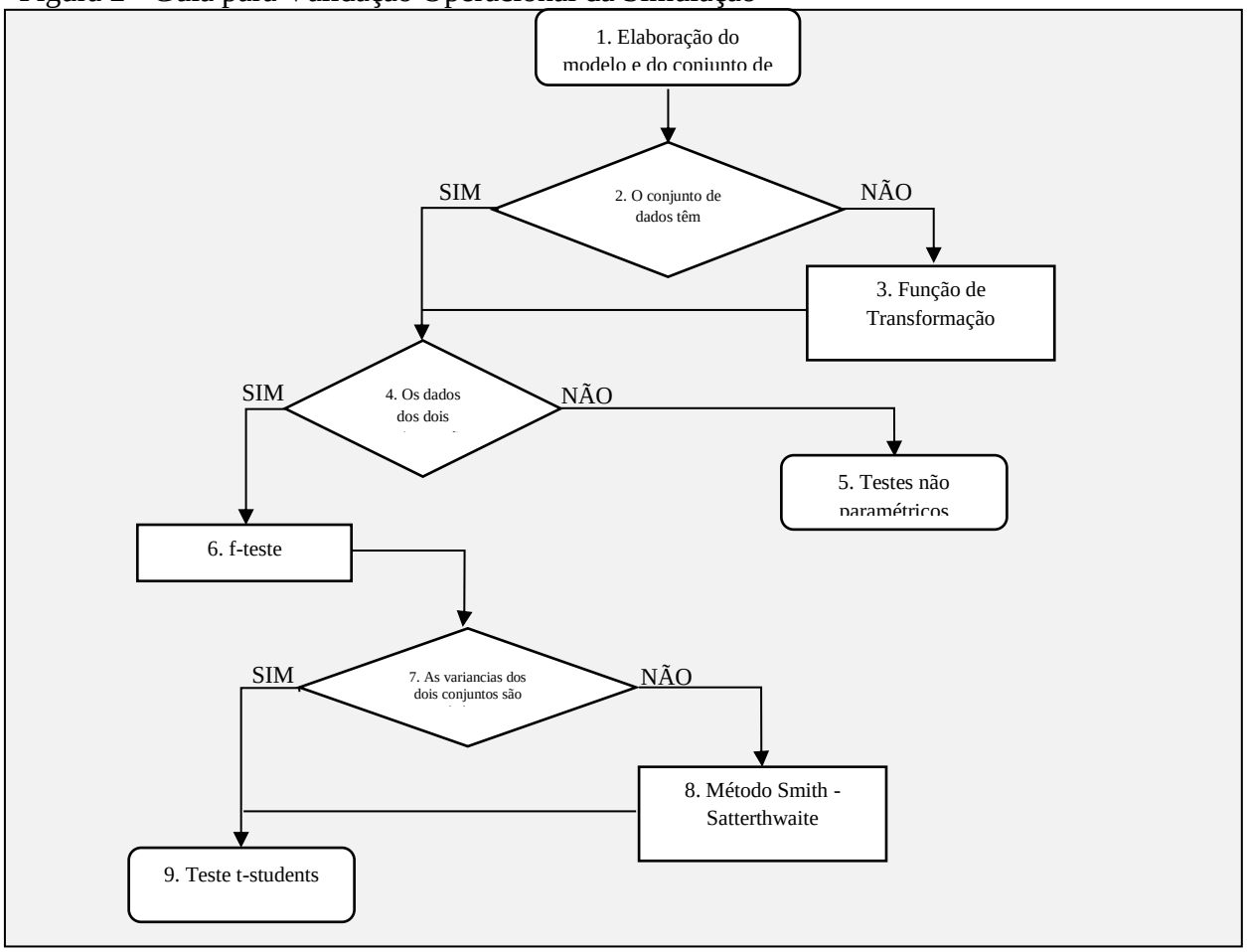
Balci (2003) descreve que o processo de validação é uma arte que requer criatividade e não pode ser considerada como uma variável binária que demonstre que o modelo é completamente correto ou incorreto.

O processo de validação tem sido pouco analisado na literatura e na maioria das vezes, os modelos são validados com métodos subjetivos. (BEKKER; VIVIERS, 2008; GUERRERO; LIBRANTZ, 2013; KOUSKOURAS; GEORGIOU, 2007; LONGO, 2010; NAZZAL et al., 2006). Chegando assim a testes estatísticos como uso de média, desvio padrão e o uso de intervalos de confiança (LAM; LAU, 2004; MAHFOUZ et al., 2010; POTTER et al. 2007).

A necessidade do uso de dados amostrais, suposições e restrições, faz com que a simulação não represente 100% a realidade, portanto, uma boa validação deve procurar que os dados alcancem um erro aceitável para que não comprometa o sistema no momento da implementação.

Leal et al. (2011) estabelecem uma metodologia de validação da simulação utilizando testes paramétricos e não paramétricos e seleção feita, considerando a normalidade nos dados. Estes testes permitem validar a simulação quando os testes demostram que os resultados do modelo apresentam médias e variações similares aos valores reais do sistema. Os autores propõem, por meio de um fluxograma os passos a seguir para uma validação operacional dos modelos de simulação, conforme ilustrado na figura 2.

Figura 2 - Guia para Validação Operacional da Simulação



Fonte: LEAL et al., 2011.

2.6.8. Definição do Delineamento de Experimentos

Nesta fase são testados novos cenários, rotinas ou experimentos. Em muitos estudos, existem evidências do uso de diversas metodologias usadas para a geração de novos cenários, entre elas encontra-se:

- Ensaio e erro.
- Softwares especializados como *SimRunner*, (pacote que acompanha *ProModel*) que permitem otimizar os resultados de uma variável específica. O *SimRunner* trabalha com Algoritmos Genéticos para obter a melhor solução da função objetivo estabelecida pelo programador dentro do espaço de busca.
- Delineamento de experimentos (DOE), técnica que permite a redução da quantidade de experimentos e a determinação dos valores que mais afetam os resultados.

Para modelos estocásticos, devem ser executadas várias replicações a fim de chegar no nível de confiança esperado, para modelos determinísticos, apenas uma replicação é suficiente (GONZAGA, 2006).

2.6.9. Execução dos Experimentos

Os valores definidos com DOE são testados no modelo de simulação, as variáveis resultantes devem ser armazenadas em uma tabela para posterior comparação.

2.6.10. Análise Estatística

São analisados os principais efeitos de cada variável testada na simulação.

2.7. ESTRUTURA BÁSICA DA SIMULAÇÃO

Para uma contextualização clara da simulação, as fronteiras entre os elementos devem ser claramente delimitadas. Carson (2003) define que, os componentes necessários para criar um modelo de simulação classificam-se em Entidades, Recursos, Locais, Redes de Transporte, Variavéis e Atributos.

Entities

São elementos que sofrem algum tipo de modificação, movimento ou processamento durante a simulação. A chegada das entidades no sistema pode ser definida como um evento determinístico ou estocástico em função das caraterísticas do sistema.

Resources

Podem ser pessoas, ferramentas, veículos ou outro objeto que pode ser usado para: transportar a entidade de um local de trabalho a outro ou executar alguma operação em uma entidade em algum local de trabalho.

Location

São Locais fixos no sistema, são áreas onde são executadas tarefas de processamento, armazenamento ou simplesmente localizações para tomada de decisões. A estrutura de *locations* considera a definição de capacidade, regras de roteamento e paradas períodicas em função das atividades desenvolvidas no local.

Path Network

São as redes pelas quais são feitas o deslocamento dos recursos e dos materiais.

Variables

São condições cujos valores modificam-se por meio de equações matemáticas ou relações lógicas. Podem ser contínuas (horário de chegada dos passageiros no trem) ou discretas (número de passageiros que chegam em um horário específico)

Attributes

São caraterísticas atribuídas a uma entidade ou local que permitem definir regras de prioridade e estabelecer rotas.

2.8. TRATAMENTO DOS DADOS

Nesta seção são explicadas as tecnicas de coleta de dados usadas nesta pesquisa, e cuja adequada implementação levaram para um satisfatória definição dos dados que foram usados durante o processo de simulação.

2.8.1. Coleta e Análise dos Dados

Uma das principais etapas nos projetos de simulação é o processo de coleta de dados, pois estes devem ser compatíveis e refletir o comportamento do sistema real. Para Newman et

al. (1998), por exemplo, o ajuste nos dados depende do nível de detalhamento que se requer no modelo.

Segundo Costa (2000), a coleta de dados pode comprometer a qualidade da resposta se esta for insuficiente. Por outro lado, a coleta de dados que não interessa ao problema em questão pode encarecer os projetos de simulação, pois este processo pode ser dispendioso. O nível de detalhamento, também pode comprometer a eficiência do modelo, pois o excesso de detalhe pode levar a simular eventos secundários que não são críticos nem relevantes para atingir o objetivo do projeto.

2.8.2. Amostragem

Geralmente, as pesquisas realizadas na área de modelagem e simulação de sistemas precisam de um estudo dos elementos que o compõem.

Nem sempre é possivel analisar todos estes elementos, fazendo-se necessário o uso de alguma técnica de amostragem, pois, o tamanho e a composição da amostra vão garantir a significância estatística dos dados e a confiança dos resultados da simulação.

Fonseca e Martins (1996) definem que, limitações como o tempo, o custo e as vantagens das técnicas estatísticas de inferência justificam o uso de planos amostrais.

2.8.3. Dimensionamento da amostra

A quantidade de dados coletados garante a significância estatística de um modelo e o número de dados coletados deve ser suficiente e permitir que se atinja o nível de precisão que se espera do modelo. O tamanho da amostra pode ser calculado considerando como informação prévia o tamanho da população. A fórmula usada para definir o tamanho da amostra apresenta-se a continuação (FONSECA; MARTINS, 1996)

$$n_o = \frac{1}{\varepsilon_o^2} \quad (1)$$

$$n = \frac{N \cdot n_o}{(N + n_o)} \quad (2)$$

Na qual:

ε_0 = erro amostral esperado

N = tamanho da população

2.8.4. Composição da Amostra

É importante estabelecer uma metodologia de coleta de dados, já que a sua consistência vai garantir que os resultados espelhem a realidade do sistema, sendo aconselhável estabelecer um processo padronizado para este fim. Fonseca e Martins (1996), classificam os métodos de composição da amostra em Métodos Probabilísticos e Métodos Não Probabilísticos ou Intencionais.

Método Probabilístico

Garante que cada elemento da população com tamanho N tenha a mesma probabilidade 1/N de ser selecionado. Este método garante a aplicação das técnicas estatísticas de inferência. Fonseca e Martins (1996) descrevem que a metodologia de amostragem pode ser:

- Amostragem Aleatória Simples: É o processo mais frequentemente usado, todos os elementos da população são rotulados e são realizados sorteios sucessivos até completar o número n de elementos da amostra.
- Amostragem Sistemática: É uma variação da amostragem aleatória simples, porém os dados estão organizados com algum critério. Calcula-se o intervalo da amostragem N/n, os elementos resultantes deste intervalo irão compor a amostra.
- Amostragem Estratificada: Para populações heterogêneas, nas quais podem ser definidos estratos ou subpopulações. Após a determinação dos estratos seleciona-se uma amostra aleatórica de cada subpopulação.
- Amostragem por Conglomerados: Usada para populações onde se faz extremamente difícil a identificação dos elementos, mas é relativamente simples a definição de subgrupos. Para estes casos, amostras aleatórias dos subgrupos podem ser coletadas e uma contagem completa deve ser feita para os conglomerados sorteados.

Métodos Não Probabilísticos

São amostras formadas por uma seleção deliberada dos elementos. Esta amostragem não garante a significância estatística nem a representatividade da população. Os tipos de amostragem não probabilísticos são:

- Amostragem Acidental: Trata-se de uma amostra formada por elementos que vão aparecendo até completar o número de elementos da amostra.
- Amostragem Intencional: Tendo um critério definido, um grupo de elementos é escolhido intencionalmente para compor a amostra.
- Amostragem por Quotas: Usada principalmente para levantamentos de mercado ou consultas eleitorais.

2.8.5. Outliers

Para Chawsheen e Latif (2006), os *outliers* são um conjunto de dados de valores excepcionais encontrados na maioria das bases de dados.

Iranzo (2007) por sua vez, descreve que um *outlier* representa anomalías nas observações, quando não são usados dados da incidência ou do padrão dinâmico do seu resultado, que são revelados na análise dos dados.

Existem vários motivos para que os *outliers* existam em uma população: podem ser resultado de erros na entrada dos dados, problemas dos sistemas, ou podem pertencer a outra população. Para observações experimentais que se distribuem normalmente (Anexo A), existem principalmente quatro metodologias para a detecção de *outliers*: (1) Test de Rosner, (2) Test de Dixon, (3) Regra de Boxplot, (4) Test de Grubbs.

Teste de Rosner

O teste de Rosner para detectar k *outliers*, pode ser usado quando o número de observações é maior ou igual a 25. Este teste ajuda a identificar *outliers* que são simultaneamente pequenos ou grandes. Ordenam-se os dados do menor para o maior e calcula-se a média e o desvio padrão. O procedimento supõe eliminar do conjunto de dados a observação *x*, mesmo que esteja mais longe da média. O teste calcula-se como segue:

$$R_{i+1} = \frac{|x^{(i)} - \bar{x}^{(i)}|}{s^{(i)}} \quad (3)$$

Na qual:

$i =$ Domínio da observação

$R_{i+1} =$ Teste estatístico Rosner para determinar se o valor mais extremo i+1 é *outlier.*

$s^{(i)} =$ Desvio padrão das n observações (dados) após o valor ter sido apagado.

$x^{(i)} =$ O ponto mais distante, após a observação mais extrema (i+1) ter sido eliminada.

O valor de R é comparado com um valor crítico. A Hipótese Nula afirma que os dados seguem uma distribuição normal, comprovada a continuação: se R é menor do que o valor crítico, a hipótese nula ou de normalidade não pode ser rejeitada e, por tanto, não existem *outliers*. Caso contrário, se R é maior do que o valor crítico, a hipótese nula é rejeitada e se aceita a presença de k *outliers*. O teste também pode ser usado para dados distribuídos com lognormal, quando os algoritimos dos dados são usados para realizar os cálculos.

Teste de Dixon

Usado geralmente para detectar um pequeno número de *outliers*, pode ser usado quando o tamanho da amostra está entre 3 e 25 observações. Os dados são organizados em forma ascendente. Considerando o tamanho da amostra, calcula-se o valor estatístico τ para o maior e o menor valor, como é apresentado na Tabela 1.

Tabela 1 - Intervalo de confiança considerando o número de observações

Observações	Valor Superior Suspeito	Valor Inferior Suspeito
3 a 7	$\tau = \frac{x_n - x_{n-1}}{x_n - x_1}$	$\tau = \frac{x_2 - x_1}{x_n - x_1}$
8 a 10	$\tau = \frac{x_n - x_{n-1}}{x_n - x_2}$	$\tau = \frac{x_2 - x_1}{x_{n-1} - x_1}$
11 a 13	$\tau = \frac{x_n - x_{n-2}}{x_n - x_2}$	$\tau = \frac{x_3 - x_1}{x_{n-1} - x_1}$
14 a 20-30	$\tau = \frac{x_n - x_{n-2}}{x_n - x_3}$	$\tau = \frac{x_{n3} - x_1}{x_{n-2} - x_1}$

Fonte: (FALLON; SPADA, 1997)

O valor estatístico τ é comparado com o valor crítico para o nível de confiança (α) escolhido. Se τ é maior que o valor crítico, a hipótese nula é rejeitada e a conclusão é que o valor mais distante é um *outlier*.

Perez (2007) afirma que este teste pode resultar repetitivo e o poder do teste se reduz à medida que o número de repetições aumenta.

Regra de Boxplot

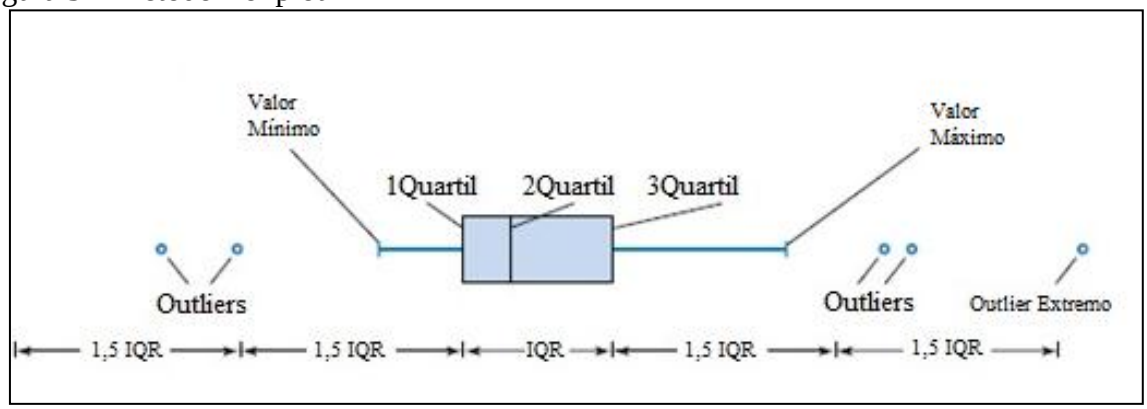
Ferramenta gráfica utilizada para visualizar a distribuição de dados amplamente usada em diversas áreas. Este gráfico permite obter informações sobre a localização, propagação, assimetria, bem como as caudas dos dados.

A box plot apresenta os três quartis, o mínimo e o máximo dentro de uma caixa retangular, alinhada horizontal e verticalmente.

A caixa encerra o intervalo interquartil com a borda esquerda (ou inferior) no primeiro quartil, q1, e com a borda direita (ou superior) no terceiro quartil, q3.

O intervalo interquantil está incluso dentro de um rectángulo (box) e os intervalos de confiança de 5% e 95% são indicados com barras fora da caixa. Aqueles valores que estão fora dos intervalos de confiança, são considerados *outliers,* como apresentado na figura 3.

Figura 3 – Método Boxplot



Fonte: MONTGOMERY; RUNGER (2002, adaptado).

A maioria dos dados espera encontrar no intervalo interquartil (IQR) ou entre os dois limites externos (L). Valores situados fora do limite $\pm$1,5L ou 1,5(IQR) que são chamados *Outliers* leves (*mild outliers*) e aqueles localizados fora do limite $\pm$3L ou 3(IQR) são chamados *Outliers* extremos (*Extreme outliers*).

Os cálculos são realizados em excel seguindo a sequência (IRANZO, 2007).

1. Calcula-se a mediana, o quartil inferior (Q1) e o quartil superior (Q3);

2. Subtrai-se o quartil superior do quartil inferior = (L)
3. Os valores que estiverem fora do intervalo de *Q1-1,5L* e *Q3+1,5L* serão considerados *outliers.*

Os testes para detecção de *outliers* para observações normalmente distribuídas são fáceis de usar e robustos, já aqueles usados para dados que não seguem uma distribuição normal são mais difíceis de aplicar e não tão robustos. Em muitas situações os dados podem ser transformados para se aproximarem em uma distruição normal para assim usar uma das técnicas apresentadas anteriormente (IRANZO, 2007).

Teste de Grubbs

Recomendado pela US EPA (Environmental Protection Agency). A EPA recomenda coletar os dados ambientais em algorítmos os quais geralmente têm uma distribuição Log-Normal. Os dados são ordenados ascendentemente e se calcula a média e o desvio padrão. Aqueles pontos mais altos ou baixos podem ser testados como *outliers*.

Teste estatístico para o menor valor

$$\tau = \frac{\bar{X} - X_1}{s} \quad (4)$$

Teste estatístico para o maior valor

$$\tau = \frac{X_n - \bar{X}}{s} \quad (5)$$

No qual:
s = desvio padrão da amostra
$\bar{x}$ = média da amostra
x = valor testado

O valor τ teste obtido deve ser comparado com o valor crítico τ, considerando o tamanho da amostra e o α esperado. Se τ teste for maior que o τ crítico, a hipótese nula é rejeitada e o dado é considerado como *outlier*.

2.8.6. Inferência Estatística

2.8.6.1. Métodos Descritivos para determinar Normalidade nos dados

Wilk e Gnanadesikan (1964) consideram que os gráficos se tornaram uma importante ferramenta para a análise estatística usada desde as fases iniciais da investigação até a divulgação e fechamento dos resultados.

Para Pereira (2010) o seu uso se torna mais eficiente a medida que os gráficos forem mais simples e autoexplicativos.

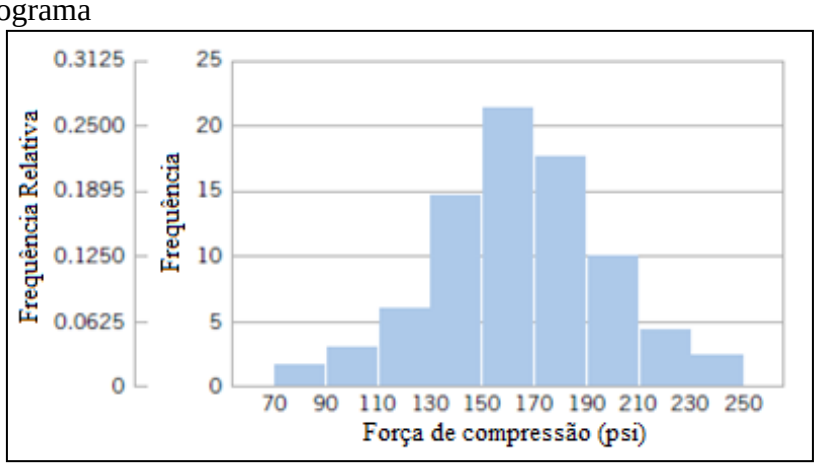
Para análise de normalidade algumas das ferramentas gráficas mais amplamente usadas são os histogramas e os diagramas de probabilidade normal.

2.8.6.2. Histograma

É uma das sete ferramentas da qualidade, conhecida também como distribuição de frequências. Os dados devem ser divididos em intervalos uniformes chamados classes ou intervalos, o histograma é construído a partir de uma série de retângulos cuja base representa a classe e a altura representa a frequência.

O histograma é uma importante ferramenta, pois graficamente permite avaliar a distribuição dos dados e a sua simetria, dando indícios que permitem determinar se um conjunto de dados tem distribuição normal ou não. A imagem do histograma pode ser observada na figura 4.

Figura 4 - Histograma



Fonte: MONTGOMERY; RUNGER (2002, adaptado).

2.8.6.3. Diagrama de Probabilidade Normal

A função de probabilidade está representada por:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty < z < \infty \quad (6)$$

E a função de distribuição acumulada é dada pela expressão:

$$\varphi(z) = P(X \leq z) = \int_{-\infty}^{z} \varphi\,(x)dx$$

(7)

para todo $-\infty < z < \infty$

O quartil p_j da distribuição normal define-se como segue (MEEKER; ESCOBAR, 1998)

$$y_j = \mu + \sigma \cdot \varphi^{-1}\,(p_j) \quad (8)$$

No qual:

$\varphi^{-1}\,(p_j) = z_j$ é o quantil p_j da distribuição normal padrão.

O gráfico é construído com a seguinte equação:

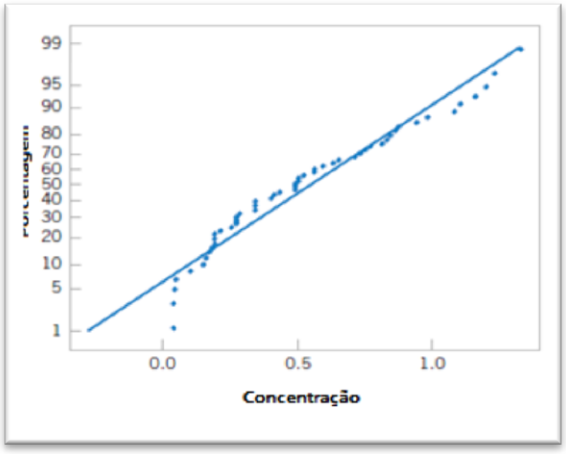
$$F(y) = \varphi\left(\frac{y-\mu}{\sigma}\right) \quad (9)$$

O gráfico de probabilidade normal tem como objetivo principal avaliar se os valores provenientes das obsevações *y1*, *y2*, *y3* ,.... yn de uma variável aleatória Y, são provenientes de uma distribuição normal $\varphi(z)$. Existem dois tipos principais de gráficos normais: o gráfico dos quantis e o gráfico dos percentis.

O gráfico dos quantis é o mais utilizado, pois permite uma comparação gráfica dos quantis de uma distribuição com os quantis de outra. Desvios de linearidade podem indicar que os dados não estão representados por uma distribuição normal (MEEKER; ESCOBAR, 1998)

A figura 5 apresenta um exemplo da visualização do gráfico de probabilidade para distribuição normal.

Figura 5 - Diagrama de probabilidade para distribuição normal



Fonte: MONTGOMERY; RUNGER (2002, adaptado).

2.8.6.4. Inferência Estatística: Teste de Hipóteses

Uma hipótese estatística é uma suposição quanto ao valor que pode tomar um parâmetro populacional no que se refere a distribuição de probabilidade, na qual um grupo de dados pode estar ajustado. O teste de hipóteses é uma regra de decisão que permite aceitar ou rejeitar uma hipótese, com base nos elementos amostrais e em um nível de confiança previamente estabelecido. O teste de hipótese tem dois fatores de avaliação:

- H_0 conhecida como hipótese nula ou de igualdade.
- H_1 conhecida como hipótese alternativa ou de desigualdade.

Montgomery & Runger (2002) afirmam que o teste de hipótese é um dos aspectos mais úteis de inferência estatística, uma vez que muitos tipos de problemas de tomada de decisão, testes ou experiências no mundo da engenharia podem ser formulados como problemas de teste de hipóteses.

Existem dois possíveis erros quando é realizado um teste de hipótese:

- Erro Tipo I: A H_0 é rejeitada sendo verdadeira, designada como α. Também conhecido como nível de significância ou α-erro.
- Erro Tipo II: A H_0 é aceita sendo falsa, designada como β.

O parâmetro *p-value* tem sido amplamente adotado na prática, que representa o menor nível de significância que levaria à rejeição da hipótese nula H0 com os dados fornecidos.

O critério usado para aceitar a H0 usando o valor de *p-value* é:

Se $p - value > \alpha$ então a H_0 é aceita, do contrário é rejeitada.

Os testes de hipóteses podem ser classificados em dois grandes grupos dependendo das caraterísticas da amostra: (1) Testes Paramétricos e (2) Testes Não Paramétricos.

Testes Paramétricos

Estes testes estatísticos são mais robustos e requerem menor número de dados para chegar a uma conclusão com elevada significância estatística. No entanto, para serem aplicados os dados precisam cumprir com três parâmetros: (a) os dados devem ser normalmente distribuídos, (b) amostras obtidas aleatoriamente e (c) existe um parâmetro de interesse que buscamos estimar.

A continuação são apresentados alguns dos testes paramétricos mais amplamente utilizados:

a. Teste F

O teste de análise de variâncias permite definir se existem evidências de igualdade das variâncias entre dois conjuntos de dados.

Para Chung (2004), somente uma das versões do teste F, é útil para práticas em simulação. Esta versão está representada pela seguinte equação:

$$F_{teste} = \frac{s_M^2}{s_m^2} \qquad (10)$$

Na qual,

s_M^2 maior variância entre os dois conjuntos de dados.

s_m^2 menor variância entre os dois conjuntos de dados.

A definição das hipóteses é feita conforme apresentação abaixo:

H_0 = a variância dos dados é similar.
H_1= não existe evidência estatística das duas variáveis terem variâncias similares.

Para a validação das hipóteses, o valor resultante de F_{teste} precisa ser comparado com o $F_{crítico}$, obtido da tabela *F-Snedecor* para o nível de significância desejado.

A aceitação é dada por: $F_{teste} < F_{crítico}$, aceita-se a H_0 e rejeita-se a H_1, concluindo-se que existe evidência estatística da igualdade entre as variâncias.

b. Teste Z

Permite realizar inferência estatística para comparação de médias. Para ser aplicado:

- A amostra deve ser maior que 30 (n>30)
- O desvio padrão populacional deve ser conhecido (σ conhecido)

Para aplicação do teste é necessário realizar os seguintes passos:

1. Definição de Hipóteses:

Ho: $\bar{X}$ = μo
H1: $\bar{X} \neq$ μo

2. Definir o valor de α (nível de significância)
3. Aplicar a equação:

$$Z_{teste} = \frac{\bar{X} - \mu o}{\sigma / \sqrt{n}} \quad (11)$$

Na qual:
$\bar{X}$ = média amostral
μo = valor a testar
σ = desvio padrão populacional
n = tamanho da amostra

4. O valor resultante de Z_{teste} deve ser comparado com o $Z_{crítico}$, que é obtido na tabela de distribuição normal para o nível de significância definido (α/2 para um teste bi-caudal).
Para un teste bi-caudal, se $Z_{teste} > -Z_{crítico}$ e $Z_{teste} < Z_{crítico}$ aceita-se a H_0 e rejeita-se a H_1, concluindo-se que existe evidência estatística da igualdade entre as médias.
A aceitação também pode realizar-se considerando o críterio de *p-value*, descrito no item denominado inferência estatística.

c. Teste T Student

É provável que seja o teste estatístico mais amplamente usado. Foi desenvolvido por um estatístico na cervejaria Guinness, permite realizar inferência estatística para comparação de médias. Para ser aplicado:

- As amostras devem ser pequenas (n<30). Para n grande pode ser aplicado Teste T-Students ou Teste Z.
- O desvio padrão populacional deve ser desconhecido (σ desconhecido)
- As Variâncias devem ser similares entre os dois conjuntos de dados (analisado com Teste F)

Para aplicação do teste é necessário realizar os seguintes passos:

1. Definição de Hipóteses:

H_0: $\mu = \mu_0$
H_1:$\mu \neq \mu_0$

2. Definir o valor de α (nível de significância)
3. Aplicar a equação:

$$t_{teste} = \frac{\bar{X} - \mu o}{\frac{s}{\sqrt{n}}} \quad (12)$$

No qual:

$\bar{X}$ = média amostral

μo= valor a testar

s = desvio padrão da amostra

n = tamanho da amostra

O valor resultante de t_{teste} deve ser comparado com o $t_{crítico}$, que é obtido na tabela de distribuição normal para o nível de significância desejado (α/2 para um teste bi-caudal) e n-1 graus de liberdade.

Em um teste bi-caudal, se $t_{teste} > -t$ e $t_{teste} < t_{crítico}$ aceita-se a Ho e rejeita-se a H1, concluindo-se que existe evidência estatística da igualdade entre as médias. A aceitação também pode realizar-se considerando o críterio de *p-value*, descrito no item (inferência estatística).

O *teste t* original sem nenhuma correção nos graus de liberdade poderia ser aplicado sem qualquer problema se as variâncias das populações forem iguais. Entretanto, se essa desigualdade ocorrer pode comprometer os resultados e aumentar os riscos de se cometer o erro tipo I e II.

Por tanto, quando a desigualdade entre as variâncias é comprovada, o método *Smith-Satterthwaite* permite ajustar os graus de liberdade para um novo valor $t_{crítico}$. (LEAL et al., 2011). Este cálculo é realizado a partir da seguinte equação:

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}} \qquad (13)$$

Na qual:

df = graus de liberdade

s_1^2 = variância do primeiro conjunto de dados

s_2^2 = variância do segundo conjunto de dados

n_1 = número de elementos do primeiro conjunto de dados

n_2 = número de elementos do segundo conjunto de dados

d. ANOVA - Análise de Variância

Método estatístico desenvolvido por Fisher que permite verificar quais fatores produzem mudanças em alguma variável com base em um teste de igualdade de médias. Este método permite comprovar a igualdade de várias médias, usando a análise de variabilidade. O procedimento basicamente consiste em desdobrar a variância total em vários componentes e decidir sobre as hipóteses previamente definidas á igualdade, baseando-se no valor das mesmas e no nível de confiança definido (FONSECA; MARTINS, 1996).

Exemplo, supondo que os parâmetros A, B, C e D são modificados no processo de transporte e que em cada experimento são coletados variáveis de saída de interesse; o objetivo é saber se estes parâmetros modificaram o resultado final da variável e se as médias das variáveis de saída são iguais.

Cada parâmetro é considerado um grupo e examinado de maneira independente. Assim, podem-se obter as médias do grupo $x_g = \overline{x_\iota}$ e uma média geral $x_m = \overline{\overline{x}}$.

Usando as médias do grupo é possível determinar a variância amostral dentro de cada grupo: s_{dg}^2 ; com a média geral, é possível determinar a variância total amostral das observações: s_{total}^2. E usando ambas as variâncias calcula-se a variância entre os grupos $s_{eg}^2 = var(x_g - x_m)$. Logo, a variância fica descomposta em dois termos: o primeiro que mede a variância dentro dos grupos e o segundo que mede a variância entre os grupos:

$$s_{total}^2 = s_{dg}^2 + s_{eg}^2 \qquad (14)$$

Na qual:

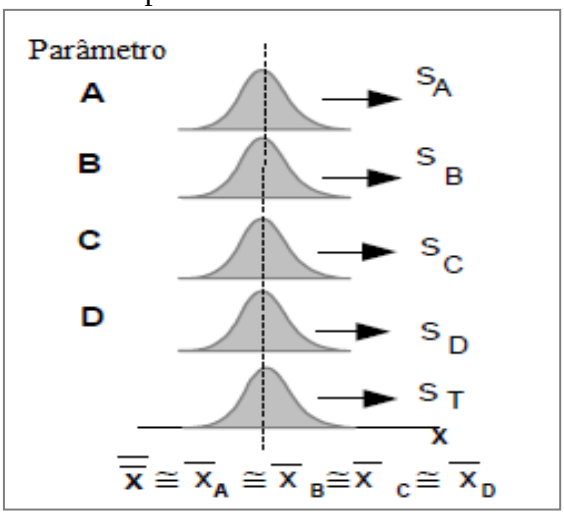
s_{total}^2 é a variância total

s_{dg}^2 é a viariância dentro dos grupos

s_{eg}^2 é a viariância entre os grupos

Assim como pode-se visualizar na figura 6, se não existir diferença entres os grupos de valores médios amostrais, os grupos serão muito similares, e por tanto, a variância entre os grupos deverá ser próxima de zero, permitindo aceitar a hipótese de igualdade das médias.

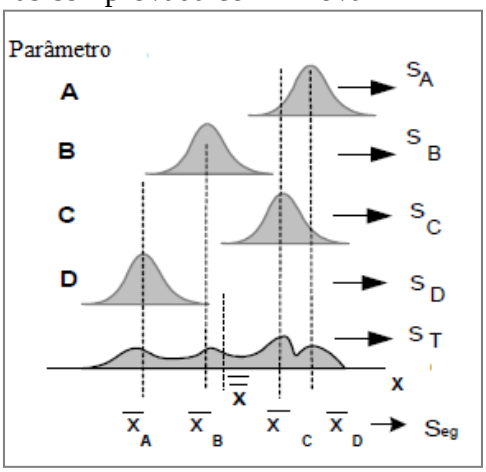
Figura 6 - Igualdade de Médias comprovado com Anova



Fonte: VILLAMIL; GARCIA (2003, adaptado)

Quando a variância entre os grupos s_{eg}^2 é suficientemente grande em relação à variância dentro dos grupos s_{dg}^2, pode-se rejeitar a hipótese de igualdade, pois existem diferenças entre as médias, que é resultado da troca dos parâmetros dentro do processo, diferença que é apresentada na figura 7.

Figura 7 - Diferença de Médias comprovada com Anova



Fonte: VILLAMIL; GARCIA (2003, adaptado)

A relação entre duas médias de variâncias amostrais s_1^2 e s_2^2, obtidas de uma mesma população, normalmente distribuída, determina uma variável aleatória que segue a distribuição F de Fisher.

$$F = \frac{s_1^2}{s_2^2} \tag{15}$$

Logo, a variância pode relacionar-se com a variável χ_v, onde v são os graus de liberdade da prova, traduzindo a equação 15 na seguinte expressão:

$$F = \frac{s_1^2}{s_2^2} = \frac{\frac{x_1^2}{v_1}}{\frac{x_2^2}{v_2}} \tag{16}$$

A distribuição da variável $F(v_1, v_2)$, será a função dos graus de liberdade v_1 e v_2 usados para definir ambas as variâncias amostrais.

Sendo conhecida a função de probabilidade da variável F, pode-se calcular o intervalo de confiança para F, intervalo que deve mostrar a relação entre as variâncias, permitindo assim aceitar a hipótese de que ambas as variâncias amostrais são estatísticamente iguais, correspondendo à mesma população.

Simplifica-se o cálculo por meio de softwares estatísticos como Matlab, SPSS e o Excel.

Testes Não Paramétricos

Dentre os numerosos testes não paramétricos que existem, serão apresentados alguns que permitem comprovar o ajuste de uma amostra em relação a uma função de distribuição conhecida.

a. Teste Kolmogorov-Smirnov

Considerado um dos principais testes de aderência baseado na função de distribuição empírica de uma variável aleatória. (JUSTEL et al, 1997)

Usado para determinar o ajuste entre duas distribuições de probabilidade subjacentes. Para realizar o teste ***Kolmogorov-Smirnov (K-S)*** inicialmente se deve definir a H_0 e a H_1:

H_0: Os dados seguem uma distribuição M.

H_1: Os dados não seguem uma distribuição M.

Para o caso de variáveis de análise única o teste pode ser aplicado com a equação 17:

$$D_n = Sup_{x \in R} |F_n(x) - F(x)| \quad (17)$$

Onde F_n representa a função de distribuição empírica de uma amostra.
Assim D_n é a maior diferença absoluta observada entre a frequência acumulada.
$F(x)$ é a frequência acumulada teórica $F_n(x)$. Se os valores são similares o D_n será pequeno.

Os critérios para a tomada de decisão entre as duas hipóteses são:
Se $D_n \leq D_\alpha$ aceita-se a H_0, caso contrário rejeita-se.

Onde D_α depende da distribuição que se deseja testar e na literatura encontra-se tabulado.

O critério de aceitação também pode realizar-se com o *p-value* associado à significância estatística (α) esperada para o teste.

Se $p - value \geq \alpha$ aceita-se a H_0
Se $p - value < \alpha$ rejeita-se a H_0

Pacotes estatísticos que acompanham os softwares de Simulação como *Stat Fit* de *ProModel*, realizam estes cálculos automaticamente e permitem analisar o ajuste pela comparação do *p-value* com o valor do α.

b. Teste Anderson-Darling

Teste estatístico que permite definir se uma amostra de dados segue uma determinada distribuição de probabilidade (SCHOLZ E STEPHENS, 2012).

A estatística é dada por:

$$AD = \frac{m \int_{-\infty}^{\infty} \{F_m(x) - F_o(x)\}^2}{F_o(x)(1 - F_o(x))} dF_0(x) \quad (18)$$

Esta equação permite testar a hipótese de uma amostra aleatória $x_1, x_2, \ldots \ldots x_m$, com distribuição empírica $F_m(x)$, que vem de uma população contínua com função de distribuição $F_0(x)$.

Programas estatísticos como *Stat Fit*, realizam os cálculos de maneira automática e permitem comprovar o ajuste com base na comparação do *p-value* com o valor de α, assim como foi explicado nos testes anteriores.

2.8.7. Delineamento de Experimentos (DOE)

Segundo Villamil e Garcia (2003) a metodologia DOE busca definir o menor número de ensaios ou experimentos e os valores que deve tomar cada uma das variáveis, com a finalidade de determinar os fatores que influenciam um ou vários parâmetros objetivo.

Mathews (2005) define o delineamento de experimentos como uma técnica formal estruturada para estudar situações que impliquem na variação da resposta em função de uma ou diversas variáveis ou fatores independentes e as interações resultantes entre elas.

De acordo com Montgomery (2008) o delineamento de experimentos é uma metodologia que permite planejar, executar e analisar estatísticamente experimentos e dados resultantes a fim de encontrar respostas objetivas e válidas.

Rekab e Shaikh (2005) por sua vez, definem o delineamento de experimentos como o planejamento, desenho e execução de testes, pelas quais são realizadas trocas sistemáticas nas variáveis de entrada, com o objetivo de encontrar mudanças nas respostas, ou variáveis de saída, permitindo assim melhorar a eficiência e eficácia no desenho.

O Planejamento fatorial é uma das técnicas mais usadas para o delineamento de experimentos, por ser uma das mais simples de se aplicar, comumente utilizada quando existe um grande número de variáveis ou fatores possíveis a serem analisados. Esta técnica permite a análise conjunta dos efeitos sobre a resposta dos fatores isolados e das interações entre eles, de forma simultânea, tornando-a bastante recomendada na otimização de experimentos. (PEIXOTO, 2012). Nesta metodologia os fatores são variados conjuntamente, assim para cada réplica todas as combinações possíveis são avaliadas (MONTGOMERY; RUNGER, 2003). Esta abordagem permite a análise dos parâmetros em todos os níveis, permitindo conhecer o resultado da sua interação dentro da delimitação do espaço de busca.

Embora permita reduzir o número de experimentos testados, o número de experimentos aumenta exponencialmente com o aumento do número de fatores.

O planejamento fatorial é uma metodologia amplamente usada na engenharia para processos de fabricação, sistemas de gestão da qualidade e de desenvolvimento de produtos. Ademais, traz importantes resultados como melhoramento da produção, redução de

desperdícios, controle da variabilidade dos processos e redução dos custos de produção e de desenvolvimento de produtos.

2.8.7.1. Conceitos básicos de experimentação

Na continuação são apresentados alguns termos fundamentais para a aplicação e análise de experimentos (CARDOZA, 2002; OLIVEIRA, 1999; MONTGOMERY, 2008).

- **Variáveis de Resposta:** Variáveis dependentes que sofrem alguma mudança nos experimentos, quando estímulos são introduzidos propositalmente no sistema, que pode possuir uma ou diversas variáveis de resposta y.
- **Fatores de Controle:** Parâmetros de entrada que compõem um modelo. Podem ser quantitativos, quando assumem valores numéricos, ou qualitativos quando representam conceitos que não podem ser quantificados. Igualmente podem classificar-se como controláveis ou incontroláveis, dependendo do nível de aprofundamento procurado nos projetos. Law e Kelton (2000) explicam que os fatores de controle definidos estão relacionados diretamente aos objetivos do estudo.

- **Efeito:** Definido como a mudança na resposta produzida pela variação do nível do fator (MONTGOMERY, 2008).
- **Interação:** Montgomery e Runger (2003) definem que existe interação quando a diferença na resposta entre os níveis de um fator não é a mesma em todos os níveis de outros fatores.
- **Fatores de ruído:** Fatores que podem ser ou não conhecidos e que podem influenciar no resultado final das variáveis de resposta. É importante evitar que os efeitos dos fatores de controle sejam misturados ou mascarados com os efeitos dos fatores de ruído, por tanto, cuidados especiais devem ser tomados no momento da realização dos experimentos.
- **Níveis dos Fatores:** domínio nos quais, se pode encontrar os fatores de controle, que são identificados com os valores -1 (nível baixo), 0 (nível médio), 1 (nível alto). Pode-se considerar os valores extremos do intervalo quando estes trabalharem com variáveis quantitativas.

- **Tratamentos:** Cada uma das corridas do experimento é conhecida como um tratamento.
- **Matriz de experimentos:** É o plano de trabalho dos experimentos, onde contém os fatores de controle, os níveis dos fatores e o tratamento do experimento.
- **Aleatorização:** Realizada basicamente para controlar os efeitos dos fatores de ruído e para atender aos requisitos dos métodos estatísticos. Neste momento é definida uma ordem de elaboração dos tratamentos da matriz. Quando os tratamentos são avaliados utilizando um modelo de simulação esta aleatorização não terá um efeito significativo, pois os fatores de ruído já podem ser parte do modelo computacional.
- **Repetição:** Processo de repetir cada um dos experimentos para encontrar uma estimativa do erro experimental, garantindo assim que as diferenças entre os dados possuam significância estatística.

2.8.7.2. Processo para conduzir experimentos

Para desenhar um experimento é necessário saber claramente o objetivo da avaliação para poder determinar os fatores e seus níveis.

Cardoza (2002) descreve os seguintes passos para conduzir os experimentos: (1) Definição dos objetivos do experimento, (2) Parâmetros do experimento, (3) Seleção dos fatores de controle e das variáveis de resposta, (4) Seleção da matriz experimental, (5) Análise de dados, (6) Interpretação dos resultados, (7) Elaboração dos relatórios.

O desenvolvimento e entendimento de cada um dos passos anteriormente descritos, acompanhado de uma clara descrição dos objetivos do experimento, permitirá encontrar efetivamente os parâmetros e fatores críticos do sistema, ajudando desta forma na melhora da qualidade da análise. Assim mesmo permite desenhar as condições ideais do processo, cumprindo com as expectativas e os limites técnicos e físicos, através da redução do número de experimentos a serem simulados.

3. MATERIAIS E MÉTODOS

Desde o ponto de vista metodológico, segundo a classificação feita por Marconi & Lakatos (2010), o método científico no qual se classifica esta pesquisa é *Método Hipotético-Dedutivo*; devido a sua natureza exploratória. A pesquisa requer um processo de conhecimento prévio do problema e da identificação de lacunas, ou seja, a resposta as perguntas estabelecidas dependerão da geração de hipóteses e do falseamento das mesmas com base na experimentação de novos cenários no modelo desenhado.

Para desenvolver um modelo de simulação é necessário entender em detalhe o sistema, de forma que o trabalho pode ser caraterizado inicialmente, como uma *Pesquisa Exploratória*, fase na qual são analisados todos os fatores e variáveis significativas do processo.

Segundo a classificação feita por Nakano (2012), a pesquisa aqui proposta carateriza-se como uma *Pesquisa Empírica de Caráter Quantitativa*, classificada assim, devido à necessidade de usar artifícios matemáticos para análise de amostras e da comparação de fenômenos por meio de técnicas estatísticas avançadas.

De acordo com a classificação proposta por Bertrand e Fransoo (2002), este trabalho pode ser classificado como uma Pesquisa Quantitativa baseada em modelos, de caráter empírico, descritivo, pois neste trabalho foram analisados e testados modelos de simulação para encontrar as relações entre as diferentes variáveis do sistema, estes modelos buscaram descrever adequadamente relações causais dos parâmetros do processo, levando a um bom entendimento do mesmo. O modelo deve ser planejado para testar a validade de modelos teóricos quantitativos e suas soluções.

Desta forma, o trabalho pode ser caraterizado como uma *Simulação*, pois permite analisar os fatores críticos do sistema e compreender as variáveis e fatores significativos com base na experimentação de mudanças em um modelo de simulação, e como *Modelagem* devido à necessidade de gerar modelos matemáticos para solução aos objetivos aqui definidos.

Os sistemas serão modelados usando o software de simulação de eventos discretos *ProModel*, seguindo as fases da simulação propostas por Montevechi et al. (2007), apresentadas na figura 1.

O trabalho começa com uma *Pesquisa bibliográfica* e para identificar o estado da arte para o tema proposto, foram realizadas consultas em livros, artigos, teses e internet, com referências a autores que publicaram estudos sobre o assunto e áreas afins. A pesquisa de

artigos nos sistemas de busca foi principalmente en inglês, usando como palavras-chave *Simulation, Subway, railway, Design of Experiments, ProModel, Arena, Optimization, Headway, Waiting Time, Travel Time Optimization, Mixed integer linear programming, Average passenger travel time, Response surface methodology*.

Também foram coletados dados do site oficial do metropolitano de São Paulo e de pesquisas de lotação dos trens apresentados pela mídia, com o propósito de projetar dois cenários e as análises de sensibilidade do modelo.

A coleta de dados também foi realizada por meio de reuniões com os envolvidos no processo, utilizando como instrumentos papel e caneta. Nestas reuniões foram solicitados os históricos das variáveis e foram realizadas perguntas pontuais para esclarecer informações básicas do funcionamento do metrô e das operações envolvidas no sequenciamento dos trens.

Baseando-se em uma *Pesquisa Documental,* foram coletados os indicadores operacionais do sistema metroviário de São Paulo, alguns dos quais foram entregues pela Coordenadoria de Estratégia Operacional do Metropolitano para o período de um ano e extraídos diretamente do sistema de informação da empresa; para aquelas informações consideradas estratégicas e confidenciais, a empresa só entregou valores determinísticos, os quais foram usados para o desenho do modelo de simulação e considerados no cálculo da confiança do mesmo.

Igualmente foram realizadas algumas coletas de dados no sistema real, como tempo de viagem, tempo de transporte, tempo de abertura de portas, (Variáveis de entrada e saída do modelo) com o propósito de alimentar e validar o modelo.

3.1. DESCRIÇÃO DO MODELO CONCEITUAL

O trabalho foi desenvolvido com base em 3 linhas de metrô semelhantes as linhas 1, 2 e 3 do metrô de São Paulo.

Estas linhas foram diferenciadas com os nomes de Linha Verde, Azul, e Vermelha e apresentam comportamentos típicos e diferenciados na demanda de passageiros.

A linha azul possui 23 estações com duas plataformas cada uma, diferenciadas pelo sentido do trajeto. As estações da linha azul, receberam como identificador um número inteiro entre 1 e 23. Segundo sua posição na rede a demanda de passageiros nesta linha pode enquadrar-se como demanda média. As estações 8 e 9 são pontos de transferência com a Linha Verde, e a estação 13 com a Linha Vermelha. Estas estações receberam estes mesmos números respectivamente para cada linha.

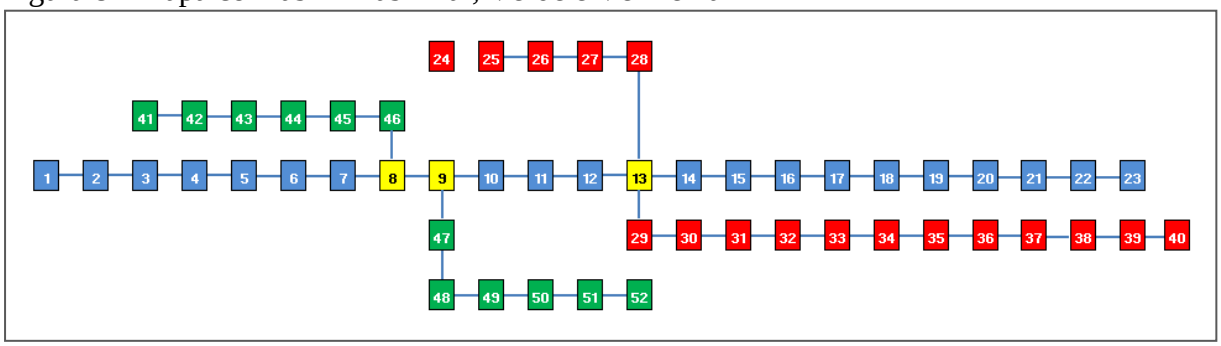
Já a linha vermelha possui 19 estações com duas plataformas, diferenciadas também pelo sentido do trajeto. As estações da linha receberam como identificador um número inteiro consecutivo entre 24 e 40, e segundo sua posição na rede, a demanda de passageiros nesta linha pode enquadrar-se como demanda alta.

A linha verde possui 14 estações com duas plataformas cada, diferenciadas pelo sentido do trajeto. As estações da linha receberam como identificador um número inteiro entre 41 e 52, e segundo sua posição na rede, a demanda de passageiros nesta linha pode enquadrar-se como demanda baixa.

A classificação da demanda nos níveis anteriormente descritos, é apresentada na Tabela 32.

Graficamente, a rede é apresentada na figura 8, diferenciada pela cor. É apresentada cada uma das linhas, com os postos de interseção.

Figura 8 - Mapa com as Linhas Azul, Verde e Vermelha

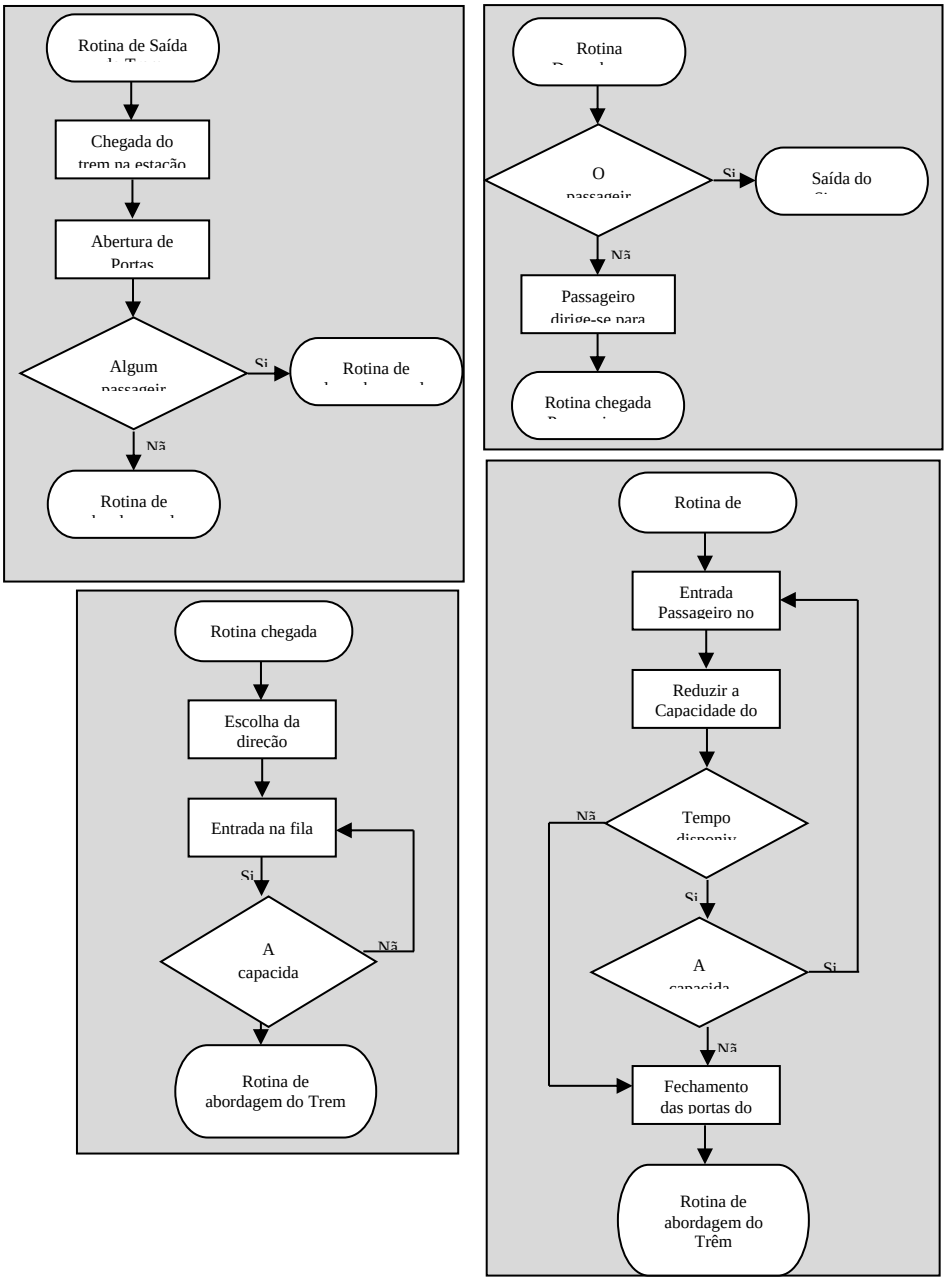


Para cada uma das estações o fluxo lógico descrito na figura 10 é composto pelas seguintes rotinas:

- ***Rotina de chegada dos passageiros na estação:*** o passageiro entra em uma estação E_i, com o conhecimento do destino final, e fica na plataforma que corresponde ao fluxo lógico para chegar ao seu destino. Aguarda até o trem chegar e se houver espaço no vagão ele ingressa, caso contrário aguarda o próximo trem que tenha espaço disponível.
- ***Rotina de abordagem do trem:*** O trem fica na plataforma em um tempo t_p, que será chamado dentro das variáveis definidas posteriormente como Tempo de abertura de Portas.

- ***Rotina de desembarque do trem*:** Quando chega na estação seguinte, se for o destino final o passageiro desce do trem, caso contrário continua no vagão até chegar ao destino final definido previamente. Para aqueles passageiros que vão de uma linha a outra, a rota é definida considerando que o passageiro precisa descer nas estações de baldiação 8, 9 ou 13 e dirige-se aos que possuem o destino final, repetindo a mesma lógica de ingresso no vagão descrita anteriormente.
- ***Rotina de Saída do Trem*:** O trem por sua parte sai a cada tempo t_h dos pátios *(Headway)*, dirige-se sequencialmente para cada uma das estações, com uma velocidade específica para cada linha, levando um tempo t_p para abertura de portas, diferenciado para cada estação. O trem continua o trajeto até chegar a estação final de cada linha. O tempo que o trem fica em cada uma das estações, permite o ingresso (*Rotina de abordagem do Trem)* ou saída dos passageiros (*Rotina Desembarque do trem*). Terminando o trajeto, o trem chega até os pátios de estacionamento, onde fica para manutenção ou repete o procedimento anterior na direção contrária. A seguir, na figura 9, o Modelo Conceitual do Processo de Transporte Metroviário.

Figura 9 - Modelo Conceitual Processo de Transporte Metroviário

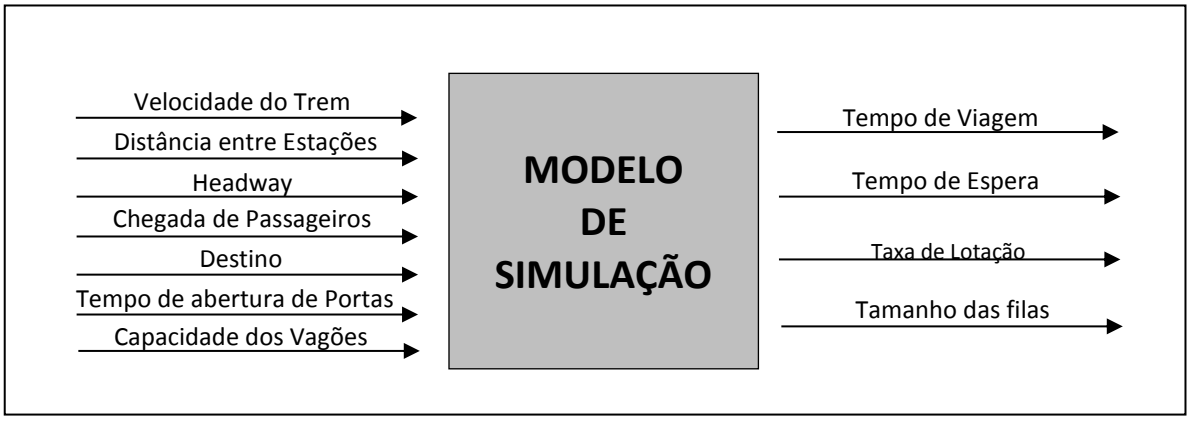


3.1.1. Variáveis de Entrada do Sistema

Foram analisadas as operações de três linhas de metrô com caraterísticas similares às linhas 1, 2 e 3 do metrô de São Paulo. As principais variáveis de entrada foram: velocidade do trem, distância entre as estações, intervalo mínimo de tempo entre trens (headway), chegada de passageiros/unidade de tempo, probabilidades destino dos passageiros para cada estação, tempo de abertura das portas em cada estação e capacidade dos vagões. As variáveis de saída foram: tempo médio de espera, tempo médio de viagem e taxa de utilização. A figura

10 apresenta um modelo conceitual das variáveis de entrada e saída para o modelo de simulação proposto no processo de transporte metroviário.

Figura 10 - Modelo conceitual das variáveis de entrada e saída para um modelo de simulação de um sistema metroviário



Na sequência será explicada cada uma das variáveis de entrada e saída.

- **Velocidade do Trem**

Assim como foi definido por Assis e Milani (2004) e Yalçinkaya e Bayhan (2009), este parâmetro foi definida como uma constante; para cada linha foi calculado a velocidade conforme o seu desempenho no sistema real.

Linha Azul = 622 m/min
Linha Vermelha = 700 m/min
Linha Verde = 614 m/min

- **Distância entre estações**

A distância entre as estações é um parâmetro fixo, os valores correspondem aos dados entregues pela empresa. Estes valores são apresentados na tabela 2.

Tabela 1 - Distância em metros entre as estações das Linhas analisadas

LINHA AZUL		LINHA VERMELHA		LINHA VERDE	
Trecho	**Distância (m)**	**Trecho**	**Distância (m)**	**Trecho**	**Distância (m)**
1 - 2	1.183	24 - 25	1.544	41 - 42	1.281
2 - 3	1.096	25 - 26	926	42 - 43	1.697
3 - 4	790	26 - 27	772	43 - 44	981
4 - 5	896	27 - 28	641	44 - 45	1.121
5 - 6	1.312	28 - 13	688	45 - 46	1.119
6 - 7	1.138	13 - 29	727	46 - 8	1.526
7 - 8	973	29 - 30	1.044	8 - 9	706
8 - 9	702	30 - 31	873	9 - 47	1.015
9 - 10	683	31 - 32	1.825	47 - 48	954
10 - 11	843	32 - 33	1.427	48 - 49	982
11 - 12	825	33 - 34	1.280	49 - 50	1.063
12 - 13	544	34 - 35	2.280	50 - 51	769
13 - 14	726	35 - 36	1.183	51 - 52	1.489
14 - 15	853	36 - 37	1.534		
15 - 16	658	37 - 38	1.722		
16 - 17	691	38 - 39	2.109		
17 - 18	1.245	39 - 40	1.461		
18 - 19	760				
19 - 20	740				
20 - 21	1.474				
21 - 22	1.057				
22 - 23	992				

Fonte: Companhia do Metropolitano do Estado de São Paulo

- **Número de passageiros que entram no sistema / unidade de tempo**

Foram coletados os valores correspondentes à entrada de usuários em cada uma das estações por hora.

A amostra corresponde ao ingresso de passageiros em cada uma das 52 estações analisadas, durante o período de 163 dias úteis (entre segunda e sexta-feira). Há 1 dado para cada dia, dentro do período estimado pela empresa como vale ou seja, o período de menor fluxo de passageiros (entre às 10h e às 16h).

Esse indicador é resultado dos dados coletados pelos pontos de ingresso e pagamento da entrada nas estações. A empresa entregou estas informações na unidade passageiros/hora, para fins de simulação, o indicador foi transformado em passageiro/segundo, e a frequência de chegada dos passageiros no sistema foi definida de 10 segundos, isto para uniformizar as entradas no sistema, este suposto foi definido, pois esta unidade de medida permite que o número de passageiros entrando no sistema seja maior que zero. Cabe destacar que esta

suposição não comprometerá os indicadores resultantes e seu uso foi necessário para reproduzir condições do sistema real.

A tabela 3, apresenta o formato de entrega dos dados de demanda de passageiros para algumas das estações.

Tabela 2 - Demanda de Passageiros

DATA	1	2	3	4	5	6
03/set	4.397	1.672	985	1.822	1.042	2.847
04/set	4.515	1.600	1.008	1.790	1.108	2.765
05/set	4.121	1.663	895	1.748	1.138	2.744
10/set	5.255	1.676	1.042	1.871	1.061	2.752
11/set	4.754	1.802	1.021	1.840	1.072	2.803
12/set	4.234	1.706	803	1.781	1.140	2.758
13/set	4.244	1.596	1.008	1.724	1.107	2.554
14/set	4.188	1.663	1.022	1.829	1.064	2.636
17/set	4.681	1.588	947	1.712	1.005	2.789

Fonte: Companhia do Metropolitano do Estado de São Paulo

- **Destino dos passageiros**

Uma matriz de probabilidade origem – destino foi disponibilizada pela empresa para este fim. Um exemplo dos dados é apresentado na tabela 4 e cada uma das probabilidades foi considerada no modelo de simulação.

Tabela 3 - Matriz origem destino dos passageiros

		Destino					
		1	2	3	4	5	6
Destino	1		0,0276	0,0676	0,0477	0,0903	0,0585
	2	0,0222		0,0278	0,0276	0,0771	0,0253
	3	0,0222	0,0177		0,0138	0,0220	0,0237
	4	0,0270	0,0611	0,0358		0,0198	0,0285
	5	0,0338	0,0640	0,0258	0,0163		0,0269
	6	0,0569	0,1054	0,0676	0,0842	0,0771	
	7	0,0338	0,0345	0,0378	0,0389	0,0419	0,0206
	8	0,0636	0,0808	0,0358	0,0641	0,0573	0,0411
	9	0,0415	0,0424	0,0219	0,0239	0,0132	0,0332

Fonte: Companhia do Metropolitano do Estado de São Paulo

- **Capacidade dos Vagões (Número de $\mathbf{passageiros/m^2}$)**

A empresa informou a capacidade, em número de pessoas por trem, para uma ocupação de 4 $passageiros/m^2$

Linha Azul = 1115 $passageiros/m^2$
Linha Vermelha = 1192 $passageiros/m^2$
Linha Verde = 1100 $passageiros/m^2$

Embora a ocupação "ótima" seja conhecida (capacidade que usa 4 passageiros por metro quadrado, segundo o que os padrões internacionais sugerem), a capacidade dos trens para o cenário atual (horários de vale) foi definida em 2200 pessoas/trem para assim reproduzir condições observadas de superlotação e não restringir o acesso de passageiros além dos 1100.

Cardoso et al. (2012) mostrou na sua pesquisa, que atualmente a composição do Metrô de São Paulo para os horários de pico encontra-se com 11,4 $passageiros/m^2$. Por tanto, para o cenário que será definido dentro desta pesquisa para os horários de pico, a capacidade dos vagões foi definida como 3300 pessoas/trem para as três linhas analisadas.

- ***Headway***

Este valor foi considerado uma constante que só difere entre uma linha e outra.

Os autores Assis e Milani (2004) e Yalçinka e Bayhan (2009), igualmente definem este indicador como um valor fixo para cada execução da simulação. Os valores de entrada no sistema correspondem aos dados entregues pela companhia.

Linha Azul = 131 segundos.
Linha Vermelha = 119 segundos.
Linha Verde = 161 segundos.

- **Tempo de abertura das portas**

Esta variável representa o tempo em que o trem está disponível para embarque ou descida dos passageiros. Para este indicador, a coleta de dados foi realizada durante um mês, dentro dos horários considerados pela empresa como períodos demanda vale (entre às 10h00 e às 16h00). O levantamento foi realizado diretamente nas estações do metrô, usando como ferramenta um cronômetro digital.

3.1.2. Variáveis de Saída do Sistema

- Tempo médio de espera para embarcar: Tempo de espera do passageiro na plataforma. Este valor é analisado como o valor médio por linha e estação.
- Tempo Médio da viagem: Tempo transcurrido desde a entrada do passageiro desde a estação de origem até a chegada no destino.
- Tempo da viagem em todas as combinações de estações: Foi analisado o tempo da viagem para todos os trajetos para assim definir as rotas críticas.
- Taxa de utilização do espaço físico do trem (número de pessoas): Lotação média do trem ao final do exercício.

3.1.3. Parâmetros da Simulação

- **Tempo de Simulação**

O tempo de simulação foi definido da seguinte forma:

Uma (1) hora de *warmup* do sistema, para assim permitir que todos os trens estejam nas suas posições, pois iniciar a simulação sem *warmup* faria com que os resultados do tempo de espera dos passageiros não fossem reais, já que os passageiros estariam aguardando até os trens ocuparem as posições reais no tempo.

Duas (2) horas de simulação, este período permite retratar o movimento de um alto número de passageiros, além de permitir que a entrada das entidades terminem seu ciclo de deslocamento

- **Número de Réplicas**

O número de réplicas foi definido como 10, parâmetro usado na literatura. (LEAL et al., 2010; YALÇINKA; BAYHAN, 2009). Para estes autores, este número de réplicas, garante a significância estatística do modelo; alguns softwares de simulação como Arena por exemplo, permitem o cálculo desta significância com base no número de réplicas, mas o *ProModel*, não fornece as informações necessárias para o cálculo.

3.2. TRATAMENTO DOS DADOS

Tanto os dados de entrada de passageiros como o tempo de abertura de portas, precisaram ser analisados passando primeiro por uma eliminação de *outliers*. Para isto optou-se pela regra de boxplot para deteção de *outliers*, por ter uma abordagem objetiva e quantitativa acompanhada da detecção sem supervisão (CHAWSHEEN; LATIF, 2006). O que é útil quando se possui um número *n* elevado de dados.

Iglewicz e Hoaglin (1993), afirmam que este método entrega resultados bastante razoáveis quando a distribuição de dados é simétrica e com forma de monte como é o caso das distribuições normal, lognormal, weibull e beta.

Para Iranzo (2007), a eliminação de *outliers* só pode realizar-se quando os erros são legítimos e não podem ser corrigidos. No entanto, já que a eliminação destes *outliers* não mudou o comportamento e nem o desempenho do sistema, optou-se pela eliminação, para obter um melhor ajuste dos dados. Os dados resultantes foram analisados nos softwares *StatFit*, responsáveis pela entrega das distribuições probabilísticas para o comportamento de cada estação.

Com o propósito de reduzir a quantidade de dados ingressados no modelo e permitir um melhor entendimento e classificação das estações, optou-se por buscar um agrupamento dos dados de tempo de abertura de portas (por estação). Assim, após analisada a normalidade dos dados e comparadas as médias por meio da metodologia ANOVA, disponível no software Matlab, comprovou-se que a média do tempo de abertura das portas nas diversas estações era diferente. Por tanto, os dados foram classificados em grupos usando a metodologia de Turkey, o que permitiu reduzir o número de cálculos e variáveis de entrada no sistema, bem como encontrar padrões de comportamento para as estações.

O tempo de abertura de portas para os grupos resultantes foi analisado usando o software *Statfit*, obtendo assim uma distribuição probabilística para o tempo de abertura de portas de cada grupo.

As distribuições de probabilidade encontradas foram comprovadas com base nos testes Kolmogorov-Smirnov e Anderson-Darling, disponíveis no mesmo software e foi usado o *p-value* como parâmetro de aceitação das hipóteses para um nível de significância estatística de 95%.

3.3. CONSTRUÇÃO DO MODELO DE SIMULAÇÃO

O modelo de simulação foi construído no software de simulação *ProModel*, em um computador com sistema operacional Windows 7.

Para construir um modelo de simulação que reproduzisse as ações que acontecem em um sistema de transporte metroviário e que correspondam ao fluxo definido na figura 10, foi necessário criar os seguintes parâmetros dentro da estrutura básica do modelo:

- **Parâmetros comuns a todas as estações**

Comum a todas as estações foram criados as *locations*:

***Locations* Comuns:**

BL_PatiosNorte

BL_PatiosSul

Estações nas quais chegam os trens para começar o processo de transporte, ou onde finalizam a sequência após terem passado por todas as estações e assim serem estacionados.

Chegadas:

Os trens chegam a cada tempo t_h (*Headway*), com uma frequência nos locais, chamada pátios em cada uma das linhas.

***Path Network*:**

Por meio do comando *Path Networks* foi criada uma rota para cada uma das linhas, unida com nós e interfaces para cada um dos locais das estações de metrô.

***Entities* Comuns:**

Metro_Linha1_Inicio

Metro_Linha1_Blue (um para cada linha de metrô)

Metro_Linha1_Transito1

Passageiros

Passageiros/transito

- **Parâmetros particulares para cada uma das estações**

Particularmente dentro de cada uma das estações foram criados os seguintes parâmetros de simulação:

***Entities*:**

Ordem_Subir_2

***Location*:**

BL_Vagao2

BL_Vagao2_Aux

BL_Desagrupar2

BL_Entrada2

BL_Estacao_2

BL_Estacao_2_SAI

***Variables*:**

BL_NoPassageiros_LA_E2_Var

BL_Numero_Passageiros_E2_Var

BL_Passageiros_Viajando_ate_E2

BL_Hora_Entrada

BL_Tempo_Espera_E2

BL_Passageiros_Transito_E2

BL_Passageiros_aprovados_embarcar_E2

Numero_Passageiros_E2_continuam_viagem

BL_Espaco_Disponivel _E2

BL_Capacidade_Trem

BL_Tempo_Abertura_Portas

***Attributes*:**

Estacao_Saida

NoPassageiros_LA_E2

Hora_Entrada_Pas_E2

H_Ent_2

Destino

3.3.1. Lógica do Modelo de Simulação

Uma vez configuradas cada uma das *location*, *entities* e *path networks* com as informações apresentadas no item de **Variáveis de Entrada do Sistema,** foi realizada a programação das rotinas do sistema no software de simulação, com base na seguinte lógica:

Os trens chegam ao BL_PatiosNorte, com uma frequência fixa para cada uma das linhas definida como ***Headway*.**

Por outro lado, há um grupo de passageiros que chega na *location auxiliar* BL_E2. O passageiro chega com uma frequência de 10 segundos, um número definido por meio de uma distribuição estatística calculada para a chegada em cada uma das estações (**Número de passageiros que entram no sistema/ unidade de tempo)**

O passageiro chega carregando um atributo principal chamado Hora_Entrada_Pas_E2.

Imediatamente, os passageiros passam para a *location* BL_Estacao_2b ou BL_Estacao_2, com base nas tabelas de probabilidade apresentadas na tabela 3. Aqui, é adicionado um novo atributo ao passageiro chamado Destino.

Por sua parte o trem que vem desde os pátios, passa pela Estação 1 e embarcam os passageiros que ali aguardam, ao chegar na Estação 2, detêm-se na *location BL_Vagon2*. Aqui o total de passageiros é descarregado e passa para uma *location auxiliar BL_Desagrupar2* com o nome de Passageiros/ trânsito, onde imediatamente, é ativada a subrutina BL_SubE2(), (nesta subrutina uma entidade chamada Orden_Subir_2, entidade que vai originar uma série de cálculos relacionados com a ocupação e a capacidade disponível no trem).

A partir deste momento, um contador de tempo é ativado para medir o tempo de abertura das portas, e paralelamente, é desenvolvida um rotina para calcular o número de passageiros que pode entrar no trem baseado na capacidade disponível (usando a *entitie* Orden_Subir_2 na *location* BL_Estacao_2_SAI), que é calculado utilizando a seguinte fórmula:

$$BL_Espaco_Disponivel\ _E2 = BL_Capacidade_Trem - BL_Passageiros_Transito_E2 \quad (19)$$

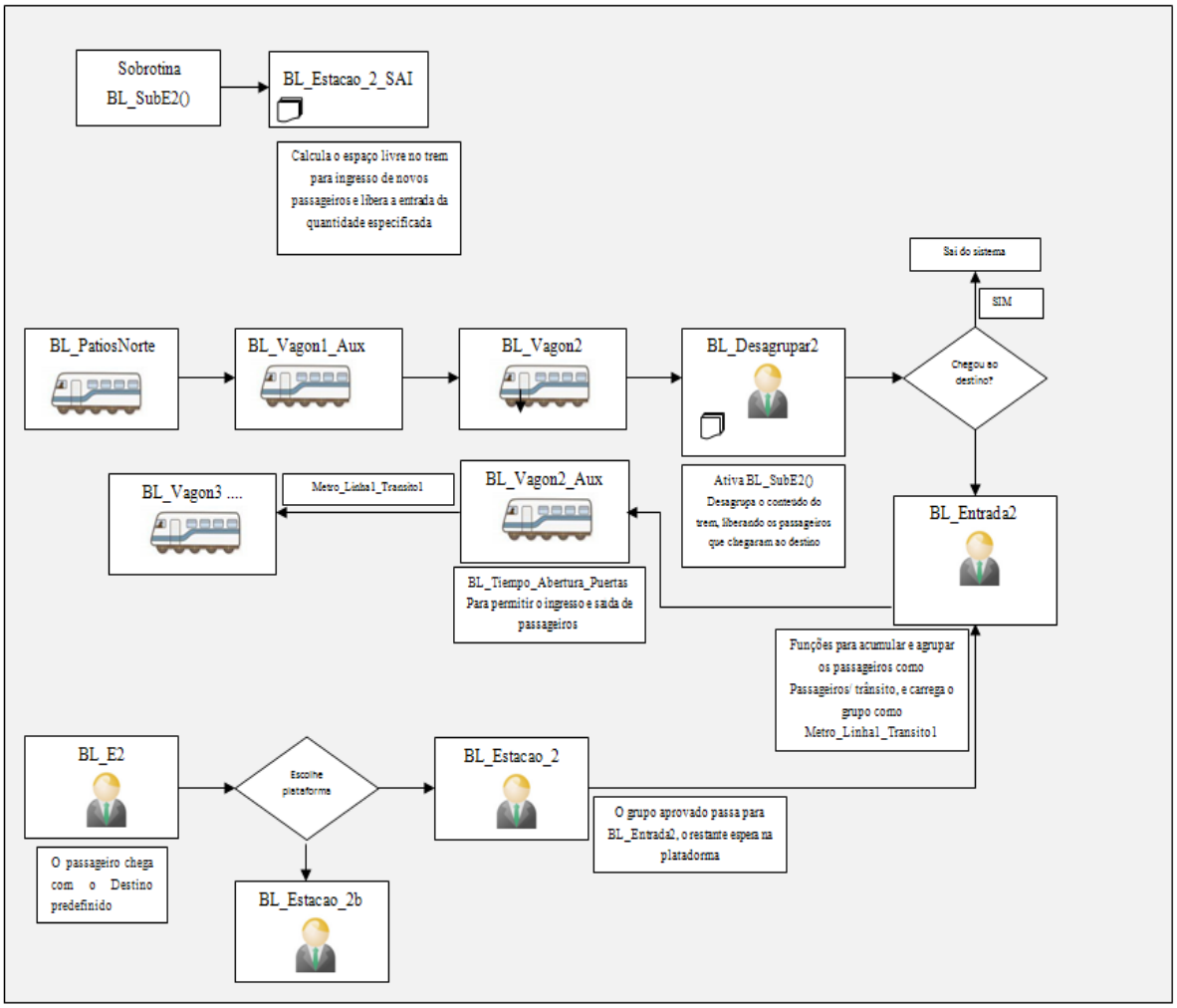
na qual, *BL_Capacidade_Trem*, é a capacidade fixa do trem definida inicialmente dentro dos parâmetros de configuração, e *BL_Passageiros_Transito_E2*, são os passageiros que encontram-se naquele momento no trem e que ainda não chegaram no destino final.

Os passageiros que podem entrar no sistema passam para a *location* BL_Entrada2 com o nome de *Passageiros*.

O contador de tempo de abertura das portas detêm-se, liberando uma rotina que faz com que os passageiros sejam agrupados com o nome de *Passageiros_trânsito*, e inmediatamente são carregados no trem na *location BL_Vagao2_Aux.*

Neste momento, o trem sai com o nome de Metro_Linha1_BLUE, e dirige-se para a *location BL_Vagao3*, onde repete a mesma rotina apresentada anteriormente, até chegar na última *location* e liberar o trem nos pátios. Esta rotina é apresentada graficamente na figura 11.

Figura 11 - Lógica da Programação em ProModel da sequência de atividades de um sistema metroviário



De um modo geral, o procedimento apresentado na figura 12, se repete tantas vezes quanto o número de estações existentes em cada uma das linhas. Em total o modelo resultou composto de:

Locations: 707
Entities: 30
Path Network: 3

Chegadas no sistema: 58 linhas, incluindo a chegada de passageiros em cada uma das estações e a chegada dos trens nos pátios.

Atributos: 15
Variáveis: 4572
Subrotinas: 104
Linhas no processo: 1197

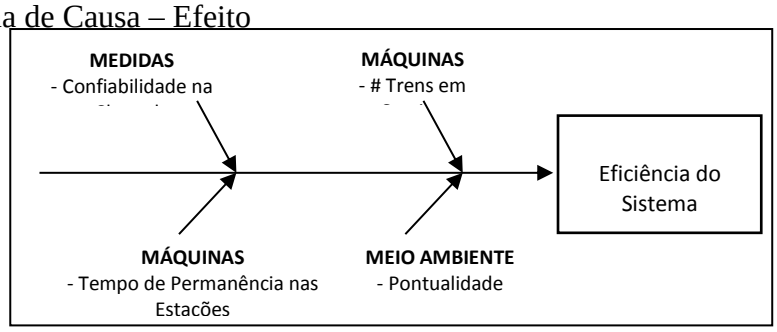
3.4. VALIDAÇÃO DO MODELO DE SIMULAÇÃO

O modelo computacional foi validado usando as técnicas de animação *face-to-face*, validação por eventos, preditiva apresentada por Banks et al. (2005) e o processo de validação objetiva proposta por Leal et al. (2011).

3.5. DEFINIÇÃO DOS CENÁRIOS

Conforme o procedimento descrito na seção de Delineamento de experimentos, foi utilizado o diagrama de Causa – Efeito proposto por Ishikawa (1976), para representar os fatores de controle no processo de transporte metroviário encontrados na literatura e escolher os que serão usados como base para esta pesquisa baseado nas caraterísticas do modelo de simulação aqui desenvolvido.

Figura 12 - Diagrama de Causa – Efeito



Com base na figura 12, nas caraterísticas do modelo de simulação e nas informações disponíveis, definiu-se que os fatores de controle usados para o delineamento de experimentos seriam: (1) Tempo de Permanência nas Estações (Com base no tempo de abertura das portas), (2) *Headway*, (3) Velocidade.

Estes fatores foram selecionados porque foram classificados como os parâmetros que respondem mais rapidamente às mudanças que ocorrem no sistema metroviário.

Seleção das variáveis de resposta

Dando resposta ao objetivo desta pesquisa, as variáveis de resposta foram: (a) Tempo médio de espera, (b) Lotação média, (c) Tempo de viagem por linha desde a primeira até a última (d) Tamanho de médio da fila.

3.6. REALIZAÇÃO DOS EXPERIMENTOS

Esta seção apresenta os principais passos realizados para a definição e cálculo dos experimentos que foram simulados através do processo de simulação.

3.6.1 Matriz de Planejamento

Definidas as etapas anteriores, o passo seguinte foi construir a matriz de planejamento dos experimentos. Foram criadas 3 matrizes, uma para cada linha e considerados 3 fatores de controle, cada um com 3 níveis. Decidiu-se trabalhar com planejamento fatorial rotacional.
O número de experimentos responde à seguinte fórmula:

$$Número\ de\ Experimentos = 2^k + ensaios\ ponto\ central + 2k \qquad (20)$$

Na qual:

k = 3 (número de fatores)
Ensaios ponto central = 3

Assim, resultou um total de 17 combinações para cada linha. A tabela 5 apresenta a matriz experimental com as combinações dos níveis dos fatores que foram testados para cada linha.

Decidiu-se trabalhar com matrizes independentes para cada linha, pois, caso fosse feita uma matriz geral, seriam necessárias 533 combinações entre os fatores (k=9) o qual, computacionalmente não era viável, pois cada experimento levou aproximadamente 30 minutos para ser simulado.

Tabela 4 - Matriz de experimentação, experimento fatorial rotacional 2^k

	Velocidade Comercial (m/min)	**Headway (segundos)**	**Tempo abertura das portas (segundos)**
N° Teste	**A**	**B**	**C**
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1
9	0	0	0
10	0	0	0
11	0	0	0
12	-alfa	0	0
13	Alfa	0	0
14	0	-alfa	0
15	0	alfa	0
16	0	0	-alfa
17	0	0	Alfa

Na tabela 6, são apresentados os valores definidos como -1, 0 e 1 para cada fator. Para todos os fatores, o nível 0 é o valor atual do indicador, usado no desenvolvimento da simulação.

No que se refere à velocidade, os valores -1 e 1 foram calculados aplicando respectivamente, uma diminuição e um aumento de 20% na velocidade atual, (para manter o sistema dentro dos limites técnicos)

Quanto ao *Headway*, o valor mínimo foi definido com a ajuda do pessoal responsável pelos indicadores dentro da companhia, e o valor máximo foi calculado, mantendo a mesma distância que se tem entre o valor mínimo e o ponto central (0).

No Tempos de Abertura de portas, para o valor 0, foi considerada a média fornecida pela companhia. O valor de 1, foi calculado somando 5 segundos ao valor médio, e o valor -1 , foi obtido restando 5 segundos na mesma média.

Tabela 5 - Valores definidos para os fatores de controle

	Velocidade Comercial (m/min)			**Headway (segundos)**			**Tempo abertura das portas (Segundos)**		
	-1	**0**	**1**	**-1**	**0**	**1**	**-1**	**0**	**1**
Linha Azul	497,6	622,0	746,4	95	131	167	15	20	25
Linha Verde	491,2	614,0	736,8	95	161	227	15	20	25
Linha Vermelha	560,0	700,0	840,0	95	119	143	15	20	25

A normalização dos fatores foi feita seguindo o procedimento:

- O cálculo dos valores α foi realizado seguindo a fórmula:

$$\alpha = 2^{k/4} \tag{21}$$

- O valor codificado e que foi usado no experimento foi calculado com a fórmula (22):

$$x_i = \frac{X_i - X_O}{\Delta X} \tag{22}$$

Na qual:

X_i = é o valor normalizado
X_0 = ponto médio (0)
x_i = alfa (α)
ΔX = variação do fator em relação ao ponto médio.

Os experimentos resultantes foram simulados com 10 réplicas cada um e os resultados foram tabulados e comparados usando ANOVA, por meio da comparação dos valores de $F_{Crítico}$ e F_{teste} e validado com apoio de gráficos de superfície de resposta, ambas metodologias desenvolvidas usando o Excel.

Os resultados foram analisados por Linha, para cada uma das variáveis de saída definidas previamente e apresentados no capítulo seguinte.

4. RESULTADOS E DISCUSSÃO

Neste capítulo recorrendo aos dados obtidos da operação de um sistema de transporte em massa, são apresentados os resultados da execução da metodologua apresentada no capítulo anterior.

4.1. COLETA DE DADOS

O único dado que precisou ser coletado foi o tempo de abertura das portas, pois os outros indicadores foram analisados com informações entregues pela companhia e descritos na seção 2.2.

4.1.1 Tempo de abertura das portas

Aplicando a equação de tamanho de amostra definido na seção 2.7.3, os cálculos foram realizados considerando os seguintes parâmetros:

ε erro amostral esperado = 5%

N (tamanho da população) foi definido como o número trens que passam em uma estação durante uma hora. Para isto, foi necessário usar um *Headway* (tempo entre a saída dos trens) de 130 segundos, com base na seguinte equação:

$$Número\ de\ Trens\ hora = \frac{3600\ seg}{130\ seg}$$

(23)

$$Número\ de\ Trens\ hora\ (N) = 28\ Trens/Hora$$

(24)

Posteriormente, foi calculado o tamanho da amostra e os intervalo de amostragem:

$$n_o = \frac{1}{0{,}05^2} = 400$$

(25)

$$n = \frac{28 \cdot 400}{(28+400)} = 26{,}17$$

(26)

A composição da amostra foi realizada usando o método probabilístico de amostragem estratificada descrito no Capítulo 1.

$$Intervalo\ de\ amostragem = \frac{28}{26} = 1$$

(27)

A coleta de dados foi realizada durante um mês, dentro dos horários considerados pela empresa como períodos com demanda vale (entre as 10h00 e as 16h00). O levantamento foi realizado diretamente nas estações do metrô, usando como ferramenta um cronômetro digital.

Para cada estação foram coletados dados consecutivos dos trens que entravam nas plataformas (pois o intervalo de amostragem é 1). Devido a dificuldade no levantamento deste indicador só foi possível coletar 24 dados de abertura de portas para cada estação.

4.2. OUTLIERS

Antes de analisar os dados e ajustá-los em distribuições de probabilidade, foi necessário realizar uma correção dos *outliers*. Para isto foi usado o teste Boxplot explicado anteriormente e escolhido pelas vantagens definidas na metodologia.

4.2.1. Entrada de Passageiros

Na sequência, é apresentado um exemplo da eliminação de *outliers* na Estação 1. Os resultados em excel sugeriram a eliminação de 7 dados, comprovado na Tabela 7.

Tabela 6 - Cálculos em Excel para detecção de *outliers*

	Estação 1
Média	12,08
DesvioPadrão	0,81
Mediana	11,99
Q1	11,58
Q3	12,45
L	0,88
Q1-1,5L	10,26
Q1+1,5L	13,77

Este processo foi replicado para todas as estações resultando em um número de dados de entrada (coluna Dados Resultantes) para cada estação apresentado na Tabela 8

Tabela 7. Entrada de Passageiros: número de dados resultantes depois da eliminação de outliers

Estação	**# Outliers**	**Dados Resultantes**
1	7	156
2	2	161
3	9	154
4	1	162
5	3	160
6	7	156
7	3	160
8	5	158
9	5	158
10	1	162
11	1	162
12	4	159
13	1	162
14	9	154
15	0	163
16	8	155
17	8	155
18	8	155
19	4	159
20	5	158
21	3	160
22	9	154
23	5	158
24	6	157
25	10	153
26	7	156
27	11	152
28	5	158
29	9	154
30	4	159
31	5	158
32	11	152
33	3	160
34	2	161
35	10	153
36	5	158
37	10	153

38	8	155
39	5	158
40	4	159
41	5	158
42	2	161
43	2	161
44	8	155
45	4	159
46	4	159
47	4	159
48	3	160
49	1	162
50	7	156
51	3	160
52	6	157

4.2.2. Abertura das Portas

O processo começa com a eliminação de *outliers* seguindo a regra de BoxPlot. Assim, são eliminados aqueles dados que podem alterar os resultados do processo, ocasionados por erros na coleta de dados ou por condições extremas que poderiam ocasionar uma mudança nos resultados. Poderiam estar entre os *outliers*, dados resultantes de operações com chuva, objetos nos trilhos, problemas mecânicos e outros, os quais não são objeto do presente estudo.

Como exemplo do cálculo de deteção de outliers, para a estação 1, não foi necessário eliminar qualquer *outlier*, os cálculos para a detecção de *outliers*, são apresentados na Tabela 9.

Tabela 8 - Cálculos em Excel para detecção de *outliers*

	Estação 1
Média	88,75
DesvioPadrão	4,84
Mediana	88,13
Q1	85,47
Q3	91,45
L	5,98
Q1-1,5L	76,5
Q1+1,5L	100,42

4.3. ANÁLISE ESTATÍSTICA DOS DADOS

Na presente seção, é explicado o tratamento estatístico realizado para cada um dos parâmetros que alimentaram o modelo de simulação.

4.3.1. Entrada de Passageiros

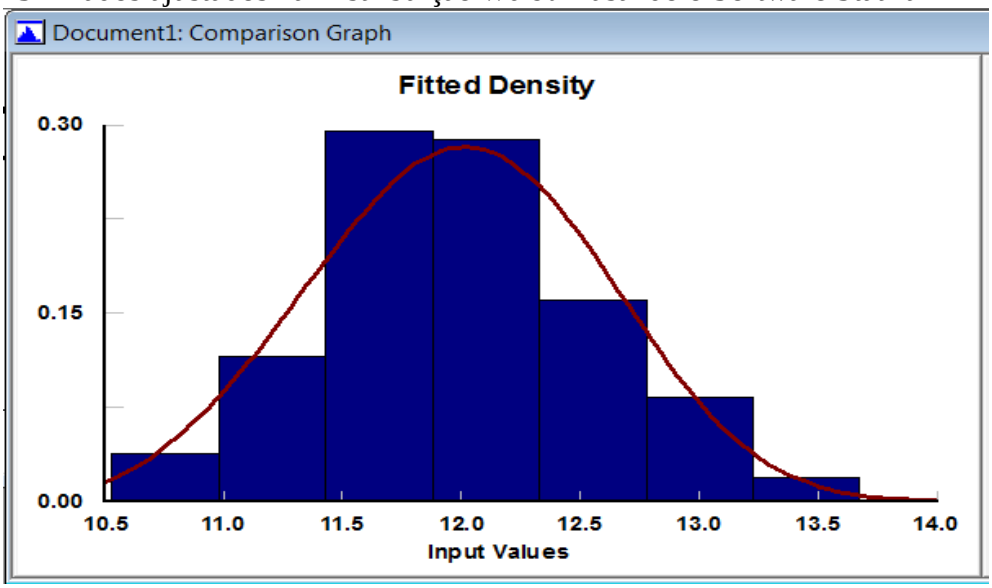
Para que o modelo seja probabilístico foi necessário estabelecer as distribuições de probabilidade nas quais os dados podem ser ajustados.

Para isto, foi usado o Software *StatFit*, um pacote de análise estatístico que acompanha o *ProModel*. Este software permite ajustar os dados numa distribuição específica, além de permitir análisar os gráficos, resultados e realizar os testes de aderência Kolmogorov Smirnov e Anderson Darling.

Análise da Estação 1

Os dados foram inseridos no software *StatFit*, que calcula os parâmetros e ajusta os dados nas distribuições estatísticas disponíveis no sistema, além de sugerir a distribuição na qual melhor se ajustam os dados analisados. Além disto, pode-se obter um gráfico com a distribuição selecionada, para a estação analisada e o software sugiriu a Distribuição Weibull, cujo gráfico de densidade é apresentado na Figura 13.

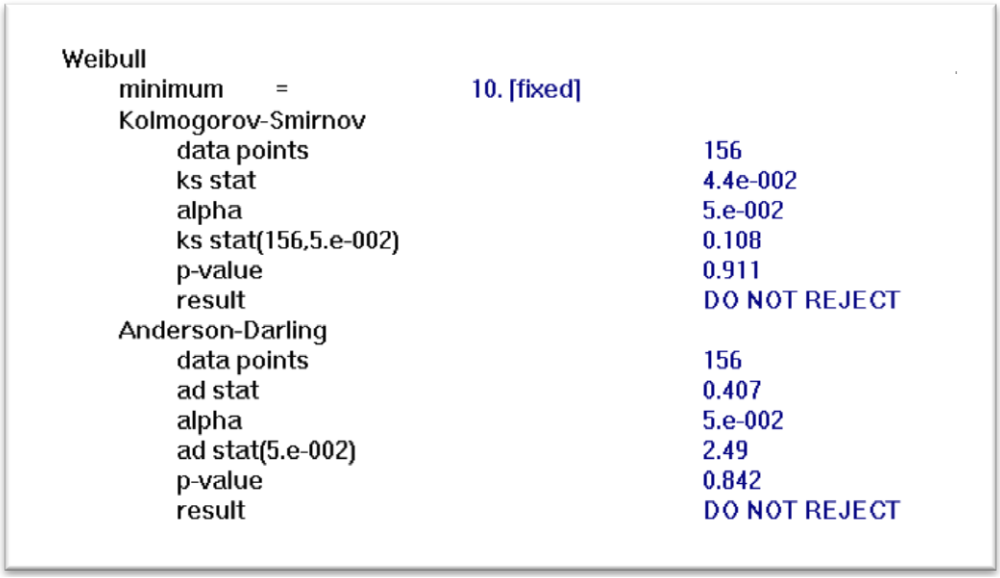
Figura 13 - Dados ajustados na Distribuição *Weibull* usando o Software Statfit



Os testes aplicados para definir o melhore ajuste foram Kolmogorov Smirnov e Anderson Darling e o ajuste conferido com o resultado do *p-value*, com uma confiança nos dados de 95%, conforme se pode observar na figura 14.

No teste Kolmogorov-Smirnov o *p-value* resultante foi 0,911, que por ser maior que o α=0,05 permite concluir que a hipótese de ajuste dos dados na distribuição Weibull é aceita. Isto pode ser conferido com o teste Anderson Darling, com um resultado no *p-value* de 0,842.

Figura 14 - Resultado dos Testes Estatísticos obtidos com o Software Statfit



Os resultados do ajuste estatístico para as 52 estações são apresentados na Tabela 10.

Tabela 9 - Distribuição Estatística da entrada de Passageiros nas Estações do Metrô

Estação	TESTE (P-Value)		Desvio Padrão	Média	DISTRIBUIÇÃO
	Kolmogorov - Smirnov	Anderson - Darling			StatFit
1	0,72	0,49	0,60	11,99	W(10;3,29;2,23)
2	0,72	0,75	0,21	4,55	W(4;3,02;0,617)
3	0,41	0,38	0,15	2,74	W(2;4,71;0,801)
4	0,98	1,00	0,20	4,91	B(4;6,11;10,9;14,4)
5	0,33	0,63	0,15	2,99	W(2;7,26;1,05)
6	0,23	0,28	0,37	7,63	W(6;4,41;1,77)
7	0,58	0,84	0,30	3,70	W(2;6,21;1,81)
8	0,46	0,72	0,24	5,26	W(4;5,72;1,35)
9	0,71	0,48	0,30	4,37	W(3;4,53;1,5)
10	0,64	0,82	0,36	3,40	B(2;4,01;4,11;1,81)
11	0,20	0,21	1,35	5,30	T(2;8,51;5,33)
12	0,99	1,00	0,25	3,40	B(2;4,63;12,9;11,4)
13	0,87	0,96	0,66	8,80	B(6;10,5;5,92;3,62)
14	0,47	0,52	0,50	7,38	W(5;4,72;2,58)
15	0,82	0,97	1,34	17,56	B(14;21,4;3,35;3,7)
16	0,42	0,36	0,16	1,70	W(1;4,43;0,757)
17	0,20	0,17	0,25	3,84	W(3;3,1;0,954)
18	0,80	0,84	0,74	7,68	B(6;22;3,98;32,6)
19	0,88	0,85	0,15	1,75	LogNormal(1;-0,296;0,211)
20	0,63	0,88	0,31	8,62	LogNormal(7;0,471;0,204)
21	0,58	0,80	0,11	2,18	LogNormal(1;0,162;0,097)
22	0,11	0,11	0,11	2,31	B(1;2,69;21,8;6,56)
23	0,47	0,42	0,33	8,51	W(7;4,7;1,63)
24	0,98	0,95	1,91	22,84	W(18;2,59;5,53)
25	0,54	0,69	0,24	4,30	W(3;5,75;1,39)
26	0,25	0,14	0,36	3,83	W(2;5,61;1,97)
27	0,63	0,48	0,96	16,86	W(13;3,82;4,28)
28	0,76	0,85	0,45	6,01	W(4;4,65;2,18)
29	0,44	0,22	0,15	2,28	LogNormal(1;0,233;0,149)
30	0,64	0,47	1,23	14,08	W(11;2,64;3,48)
31	0,80	0,57	0,41	4,15	W(2;5,66;2,31)
32	0,65	0,72	0,22	5,67	B(2;6,6;18,3;10,4)
33	0,69	0,50	0,34	9,29	W(7;6,93;2,42)
34	0,31	0,42	0,41	7,33	W(5;5,93;2,48)
35	0,15	0,17	0,19	5,26	B(4;6,38;11,9;10,8)
36	0,45	0,24	0,19	4,48	LogNormal(3;0,389;0,162)
37	0,12	0,06	0,16	3,44	B(2;4,26;14,7;8,69)
38	0,42	0,63	0,14	3,06	B(2;3,57;14,1;6,89)
39	0,13	0,07	0,38	9,31	W(7;8,55;2,31)
40	0,07	0,02	0,55	12,54	B(9;15,6;10,8;9,35)
41	0,25	0,25	0,24	3,45	LogNormal(2;0,364;0,179)
42	0,58	0,82	0,49	5,13	B(4;6,28;2,11;2,22)
43	1,00	1,00	0,18	3,68	B(3;4,29;6,34;5,74)
44	0,51	0,59	0,12	2,06	LogNormal(1;0,0444;0,126)
45	0,20	0,28	0,09	1,48	LogNormal(1;-0,755;0,207)
46	0,65	0,88	0,05	0,70	LogNormal(0;-0,351;0,0807)
47	0,97	0,98	0,31	4,85	B(3;5,93;10,9;6,39)
48	0,49	0,91	0,25	3,69	B(2;4,26;10,4;3,65)
49	0,06	0,10	2,25	15,46	B(12;17,8;1,66;0,922)
50	0,36	0,48	0,34	4,12	B(2;4,83;7,57;2,7)
51	0,89	0,90	0,13	1,36	W(1;3,02;0,404)
52	0,59	0,60	0,13	2,90	LogNormal(2;-0,116;0,17)

B=Distribuição Beta, W=Distribuição Weibull

4.3.2. Abertura de Portas

Como a coleta de dados para abertura de portas foi feita em todas as 52 estações e nas duas plataformas (direções) disponíveis para cada estação, decidiu-se agrupar os dados para facilitar tanto o processo de programação em *ProModel* bem como para reduzir o número de variáveis, para uma futura tomada de decisões.

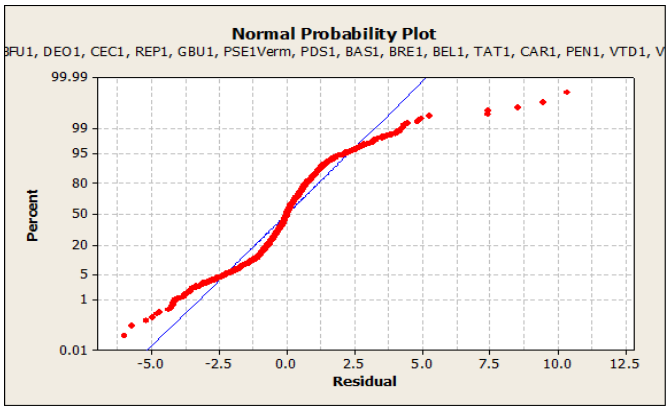
Já que contaria-se com um grande número de dados para cada grupo gerado, decidiu-se separar a amostra em dois grupos, o primeiro para análise estatística com 62% dos dados (Amostra de análise), e o segundo para validação do modelo de simulação com 38% dos dados restantes (Amostra de validação). Embora o processo de análise estatístico e definição do comportamento de cada uma das estações seja fundamental para a precisão do modelo, também se considerou importante ter uma amostra adequada para a validação, que igualmente representa o comportamento da operação.

Na literatura não se encontraram indícios de metodologias para este tipo de análise em simulação, porém, Gibb (1996) e Kröse e Van der Smagt (1996), recomendam dividir as amostras para treinamento e prova das redes neurais em uma proporção 70/30.

Para tratamento estatístico da amostra de análise decidiu-se usar o Teste Anova, para comparação das médias dos dados, e o teste de Tukey para apoiar o processo de agrupamento. Cabe ressaltar que estes testes foram realizados usando o software Minitab versão 16.

Antes de começar com o teste Anova, foi indispensável conferir a normalidade dos dados com apoio do software Minitab. Na figura 15, apresentam-se os resultados da análise onde apresenta-se evidências da normalidade dos dados.

Figura 15 - Análise de Normalidade dos dados

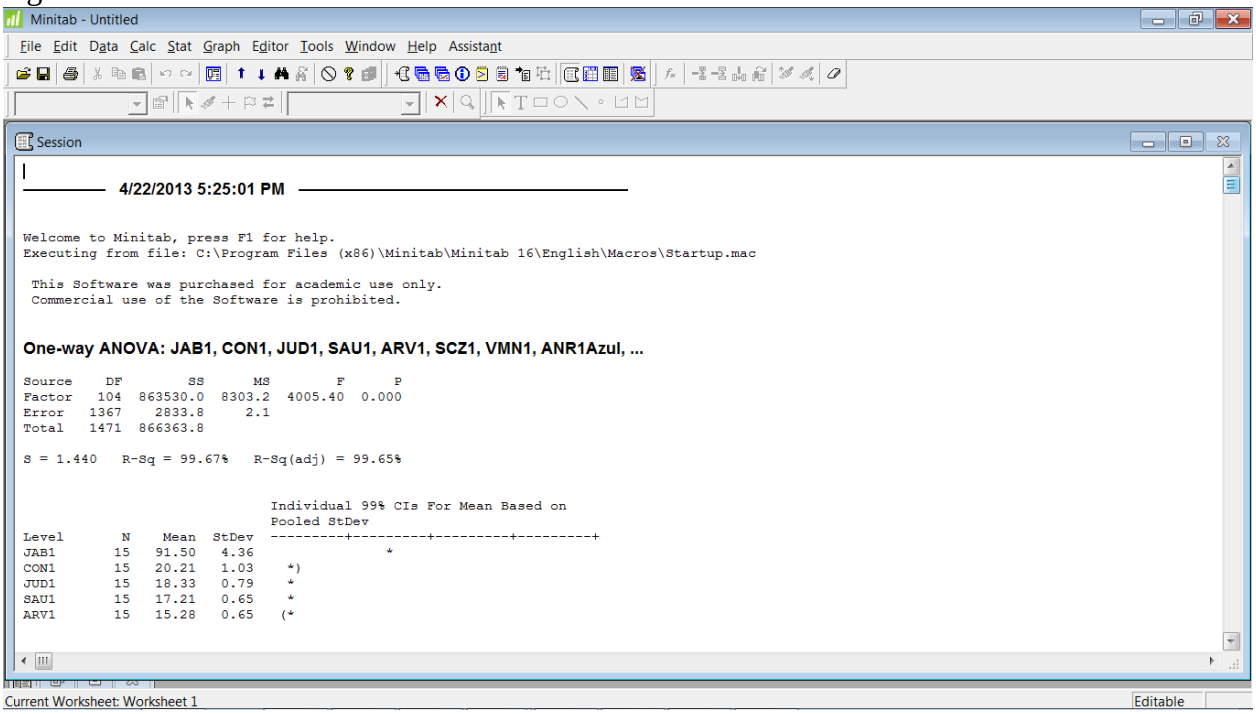


Para realizar o teste Anova no software Minitab, foi escolhida a opção ANOVA One-Way (Unstacked), na qual o teste é realizado com 99% de nível de confiança. As hipóteses definidas para o teste ANOVA são:

H_0= o tempo médio de abertura das portas entre as estações é igual.

H_1= nem todas as médias são iguais.

Figura 16 - Resultados do Teste *Anova* em *Minitab*



Os resultados apresentados na figura 16 podem ser interpretados da seguinte maneira:

Para uma confiança de 99% e α=1%, conclui-se que ao menos uma das estações tem média diferente das outras:

$$p - value\ (0{,}000) < \propto (0{,}01)$$

Assim é rejeitada a H_0 e aceita a H_1.

Constatando-se a desigualdade entre os tempos de abertura das portas, procurou-se um agrupamento dos dados usando o teste de Tukey com um nível de confiança de 95%.

O Teste de Tukey apresenta um agrupamento, o qual permitiu dividir os dados em 17 grupos assim como apresentado na Tabela 11.

A nomenclatura apresentada para no número das estações tem o seguinte significado:

A=Estação de transferência da Linha Azul.

V=Estação de transferência da Linha Verde.

R=Estação de transferência da Linha Vermelha.

a e b = Sentido do trem

Tabela 10 - Agrupamento das Estações por tempo de abertura das portas usando teste de *Tukey*

Grupo																				
1	4	28	50	44	24	24b	32b	45	22b	5										
2	13A	31	9bA	3	15	27b	48b													
3	16b	36b	12b	10b	31b	33	14	34	39b	8bA	11b	35b	47							
4	18	45b	33b	18b																
5	44b	17b	13bA																	
6	13R	8bV	46b	17																
7	43b	49	28b	41	9V	51	47b	9bV	48	2	13bR									
8	43	30	42b	46	8V	49b	51b													
9	26b	19	12	16	6	14b	35	10	6b	19b	29b	38	7	9A	11	5b	20	38b	25b	8A	20b
10	37b	25	26	29	36															
11	30b	34b	22	21	21b	27	39	4b	7b											
12	37	52	42	50b	32															
13	15b	3b																		
14	40																			
15	23																			
16	1																			
17	2b																			

Os dados foram agrupados e os cálculos das distribuições estatísticas feitos com ajuda do software *StatFit*, conforme apresentado na tabela 12. Para a validação da distribuição usou-se os testes Kolmogorov-Smirnov e Anderson Darling, validando-se a hipótese de ajuste para um $\propto = 0{,}05$, onde $p - value > \propto$

Para o Grupo 8, embora o teste Kolmogorov-Smirnov tenha rejeitado a H_0, a distribuição foi aceita pois o teste Anderson Darling aceitou a H_0.

Tabela 11 - Distribuição Estatística Abertura de Portas nas Estações

	Distribuição	# dados	Teste	
			Kolmogorov-Smirnov	Anderson Darling
1	**N(16.14,1.13)**	143	0,27	0,269
2	**N(18.38,0.95)**	101	0,58	0,59
3	**N(14,33 ; 0,96)**	185	0,05	0,0329
4	**N(32,74 ; 1,59)**	55	0,10	0,382
5	**N(30,99 ; 1,64)**	44	0,66	0,706
6	**N(29,34 ; 2,54)**	57	0,60	0,777
7	**N(20,79 ; 1,58)**	151	0,15	0,193
8	**N(26,11 ; 1,94)**	87	0,03	0,151
9	**N(11,86 ; 0,78)**	290	0,13	0,203
10	**N(9,85 ; 0,74)**	71	0,16	0,53
11	**N(13,03 ; 0,77)**	126	0,63	0,663
12	**N(22,77 ; 2,08)**	73	0,18	0,607
13	**N(43,45 ; 1,62)**	30	0,97	0,912
14	**L(222,70 ; 4,96)**	15	0,32	0,391
15	**N(96,21 ; 2,53)**	15	0,77	0,725
16	**N(91,49 ; 4,20)**	15	0,20	0,446
17	**N(45,71 ; 0,97)**	14	0,13	0,318

4.4. VALIDAÇÃO DO MODELO

O processo de validação foi realizado usando várias das metodologias apresentadas na fundamentação teórica na seção 2.6.7, as quais são descritas na sequência.

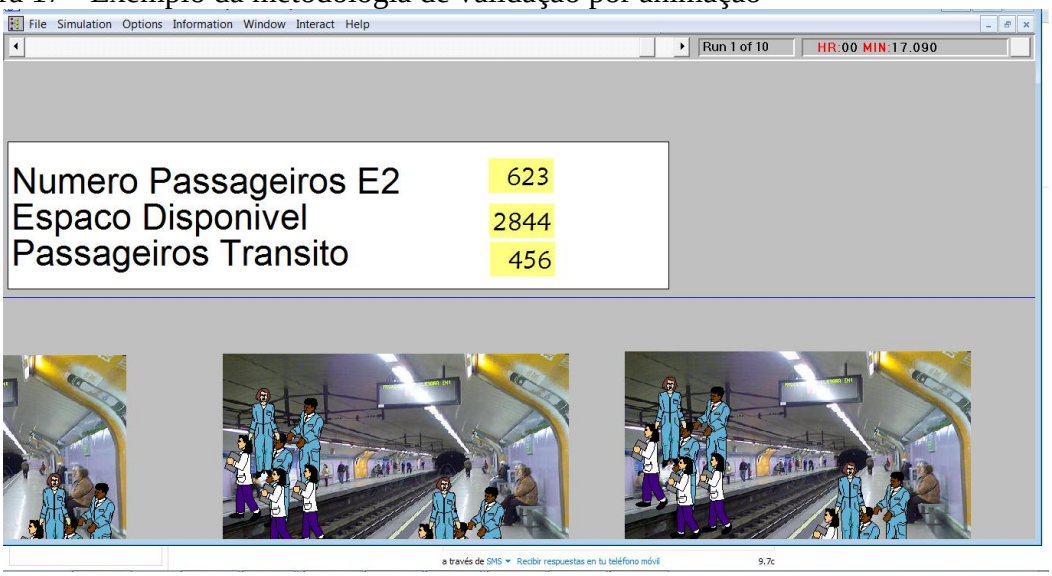
4.4.1. Animação

Uma das vantagens do uso do *ProModel* é a capacidade de validar o modelo por meio da animação durante o processo de programação. Os aspectos mais importantes validados com a simulação foram:

- Por meio de variáveis, foi apresentada na tela uma análise em tempo real das entradas e saídas no sistema, para garantir que todos os passageiros fossem transportados para seu destino, as variáveis de interese foram acompanhadas em tempo real, conforme apresentado no exemplo da figura 17.
- Movimento dos trens garantindo que todos os trens passassem e parassem em cada uma das estações.
- Tempo do trem em cada estação: com base nas pausas no sistema, foi possível avaliar a hora de entrada e saída do trem em cada uma das estações.

- Evitar bloqueios no sistema que poderiam acontecer por falhas na programação ou pelo alto tempo de espera do trem em alguma das estações.

Figura 17 - Exemplo da metodologia de validação por animação



4.4.2. Validação Face-to-Face

Erro! Indicador não definido.

Para a validação Face-to-Face, basicamente foram analisadas duas situações específicas, como são o total de passageiros transportados e o nível de ocupação dos trens, já que metodologia não sugere análise estatística, o processo focaliza-se na obtenção de dados coerentes com a realidade do sistema.

4.4.2.1. Análise das entradas e saídas do sistema

Em duas horas de simulação, no total foram transportados 194.650 passageiros, entre as estações azul, vermelha e verde, como apresentado na tabela 13.

Tabela 12 - Entradas e saídas do sistema

	Média	**Desvio Padrão**
Total Passageiros Transportados (2h)	194.650	6.902,19
Tempo Médio no sistema	17,96	0,2
Tempo Médio em movimento	11,49	0,14
Tempo Médio em espera	1,79	0,04
Tempo Médio em operação (entrando nos vagões)	2,05	0,04
Tempo Médio bloqueado	2,61	0,07

4.4.2.2. Ocupação dos vagões

A taxa de ocupação foi calculada considerando-se a capacidade dos trens para cada linha, com dados fornecido pela companhia e calculados considerando o tamanho do vagão e a capacidade para uma taxa de 4 pessoas por metro quadrado.
Os dados de capacidade dos trens são:

Linha Azul: 1.115 pessoas/trem
Linha Vermelha: 1.192 pessoas/trem
Linha Verde: 1.111 pessoas/trem

Tal como apresentado nas tabelas 14 e 16, a simulação aponta que as Linhas Azul e verde alcançam níveis de ocupação de 37% e 26%, enquanto a linha vermelha pode atingir 85% de ocupação como demonstra a tabela 15. Embora não tenha sido possível validar numericamente estes dados através do método de Análise das entradas e saídas do sistema, a experiência dos usuários reflete o comportamento das linhas, em especial a problemática da linha vermelha, a qual tem uma alta taxa de ocupação mesmo nas horas de vale.

Tabela 13 - Taxa de Ocupação dos trens da Linha Azul em horário vale

LINHA AZUL				
Estação	**Lotação Minima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
12	19	466	415	37%
11	40	457	408	37%
10	26	414	367	33%
15	87	450	352	32%
7	27	381	349	31%
9	21	393	347	31%
16	3	440	345	31%
17	13	413	330	30%
8	28	375	330	30%
6	66	358	330	30%
18	21	393	322	29%
14	46	424	319	29%
13	22	445	317	28%
19	4	382	311	28%
20	33	322	271	24%
21	5	316	267	24%
22	5	311	264	24%
5	19	299	275	25%
4	43	279	259	23%
1	139	307	229	21%
3	20	230	216	19%
2	45	207	197	18%
23	-	-	-	0%

Tabela 14 - Taxa de Ocupação dos trens da Linha Vermelha

LINHA VERMELHA				
Estação	**Lotação Mínima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
28	94	1.057	1.009	85%
27	325	994	946	79%
24	544	1.184	869	73%
36	42	908	853	72%
35	55	908	852	71%
34	80	904	849	71%
33	116	892	832	70%
37	18	882	827	69%
30	174	882	824	69%
38	25	874	820	69%
29	27	879	821	69%
31	49	874	819	69%
32	57	873	819	69%
13	216	874	810	68%
39	70	835	781	65%
26	66	732	698	59%
25	84	682	648	54%
40	-	-	-	0%

Tabela 15 - Taxa de Ocupação dos trens da Linha Verde

ESTAÇÃO VERDE				
Estação	**Lotação Mínima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
49	140	332	287	26%
50	24	321	277	25%
51	4	312	269	24%
48	33	296	243	22%
47	53	293	230	21%
9	13	275	200	18%
46	14	207	192	17%
45	25	193	179	16%
8	47	242	176	16%
44	25	164	152	14%
43	40	138	127	11%
42	42	98	91	8%
41	46	99	73	7%
52	-	-	-	0%

4.4.2.3. Tempo de espera por linha

O tempo de espera médio por linha, éfoi consistente com o tempo entre saída dos trens (*headway*) e a lotação dos trens, conforme apresenta a tabela 17.

Tabela 16 - Tempo Médio de Espera e Tamanho da Médio da Fila por Linha

	Tempo Médio de Espera (Min)	Tamanho Médio da Fila por Linha
Linha Azul	1,32	30,81
Linha Vermelha	2,32	100,60

Linha Verde	1,52	27,59

4.5. VALIDAÇÃO POR EVENTOS

Para a validação do tempo da viagem decidiu-se confrontar o tempo da volta (ida e volta) das três linhas analisadas, realizado através da comparação entre o tempo de volta da simulação vs o tempo de volta real, e entre o tempo de volta da simulação vs o tempo teórico.

Para esta validação usou-se a amostra de validação da variável tempo de abertura de porta (amostra de 38% dos dados) e o tempo de deslocamento do trem para as três estações. Para esta última informação contava-se com um único dado real devido à dificuldade de coleta.

Para a validação foi feita um somatório do tempo médio de abertura de portas para cada linha e este valor somado ao tempo de deslocamento. Foi usado o resultado da primera réplica da simulação, tal como apresentado na tabela 18.

Tabela 17 - Cálculo do Tempo da Viagem Real

	Linha Azul	Linha Vermelha	Linha Verde
Tempo de Deslocamento (min)	35,51	31,27	22,49
Tempo Médio Abertura de Portas(min)	8,4	8,4	4,9
Tempo da Viagem Real (min)	44,0	39,7	27,4

O resultado foi comparado com o indicador *Tempo de Viagem* obtido na simulação, chegando-se a um erro de 6,0% na linha azul, de 8,1% na linha vermelha e de 1,8% na linha verde, como pode-se observar na tabela 19.

Tabela 18 - Erro no tempo da volta Real vs Simulação

	Linha Azul	Linha Vermelha	Linha Verde
Real	44,0	39,7	27,4
Simulação	46,6	36,4	27,9
Erro (%)	**6,0%**	**8,1%**	**1,8%**

Foi feita uma segunda validação neste indicador, relacionando o valor simulado com o tempo teórico fornecido pela companhia.

Para os cálculos foram usados os dados teóricos de velocidade para cada linha e o tempo de abertura de portas de 20 segundos (fixo para todas as estações), conforme cálculo apresentado na tabela 20.

Tabela 19 - Cálculo do Tempo de Meia Volta Teórico

	Linha Azul	Linha Vermelha	Linha Verde
Distância (Km)	20,18	22,04	14,70
Velocidade (Km/h)	31,00	42,00	35,00
Tempo (min)	39,06	31,48	25,21
Abertura de Portas (min)	7,7	6,0	4,67
Tempo de Meia Volta (min)	**46,73**	**37,48**	**29,87**

O resultado foi comparado com o indicador *Tempo de Viagem* obtido na simulação, chegando em um erro de 12,9% para a linha azul, de 6,3% para a linha vermelha e de 6,9% na linha verde, como pode-se observar na tabela 21.

Tabela 20 - Erro no tempo da Meia volta Teórico vs Simulação

	Linha Azul	Linha Vermelha	Linha Verde
Teórico (min)	46,7	37,5	29,9
Simulação (min)	46,6	36,4	27,9
Erro (%)	**0,3%**	**2,8%**	**6,7%**

4.6. VALIDAÇÃO ESTATÍSTICA

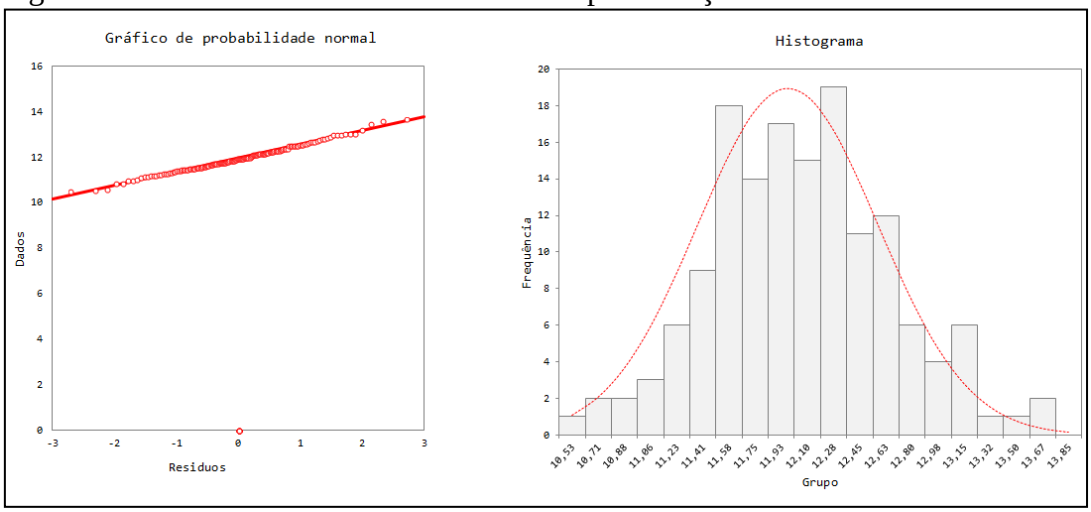
Foram analisados os dados de saída do modelo seguindo a metodologia proposta por Leal et al. (2011), com o objetivo de garantir que as distribuições estatísticas calculadas para cada parâmetro realmente estivessem sendo projetadas na simulação.

4.6.1 Parâmetro Entrada de Passageiros

Foi realizada inicialmente uma prova de normalidade para decidir qual teste estatístico seria usado.

A normalidade foi analisada usando o método empírico de Gráfico de Probabilidade Normal e o histograma para analisar a dispersão dos dados e seu ajuste na curva normal. Na figura 19, é apresentada a análise realizada para o número de passageiros que ingressa na Estação 1. Por meio deste teste validando-se-á a Normalidade dos dados.

Figura 18 - Gráfico de Probabilidade Normal para estação 1



Este procedimento foi realizado com as 52 estações, confirmando a normalidade em todos os casos. Após a análise de normalidade, continuando com os passos propostos por Leal (2011) apresentado na figura 2, foi realizado o teste F para análise das variâncias. Assim, foram comparadas as variâncias dos dados do sistema real com as dos resultados da simulação. O teste foi realizado seguindo a equação 10. O resultado de F teste foi comparado com o dado obtido na tabela de Distribuição F de Snedecor para α=5%. O teste confirmou que só para algumas estações a H_0 nula é válida, concluindo que os dados simulados têm a mesma variância que os dados reais. Os cálculos são apresentados na tabela 22.

Tabela 21 - Estações onde a H_0 é aceita = Os dados simulados e os dados reais têm variâncias similares

Estação	F teste	Validação das Hipóteses
5	1,92	Aceita Ho
8	1,58	Aceita Ho
9	1,26	Aceita Ho
13	1,97	Aceita Ho
20	2,18	Aceita Ho
21	1,10	Aceita Ho
22	1,49	Aceita Ho
32	2,03	Aceita Ho
33	2,70	Aceita Ho

35	1,34	Aceita Ho
36	2,12	Aceita Ho
37	1,68	Aceita Ho
38	1,06	Aceita Ho
39	1,32	Aceita Ho
43	2,59	Aceita Ho
44	2,68	Aceita Ho
45	2,41	Aceita Ho
46	2,29	Aceita Ho
51	1,81	Aceita Ho

Por outro lado, houve 24 estações nas quais a H_0 não foi rejeitada, concluindo assim que a variância dos dados reais é diferente dos dados simulados, como pode ser observado na tabela 23.

Tabela 22 - Estações onde a H_0 foi rejeitada e confirmada a H_1= Não existe nenhuma evidência da igualdade entre as variâncias

Estação	F teste	Validação das Hipóteses
1	115,25	Rejeita Ho
2	19,47	Rejeita Ho
3	5,64	Rejeita Ho
4	4,60	Rejeita Ho
6	4,39	Rejeita Ho
7	7,19	Rejeita Ho
10	18,49	Rejeita Ho
11	76,39	Rejeita Ho
12	2,95	Rejeita Ho
14	4,11	Rejeita Ho
15	14,67	Rejeita Ho
16	5,48	Rejeita Ho
17	4,31	Rejeita Ho
18	11,77	Rejeita Ho
19	3,55	Rejeita Ho
23	179,59	Rejeita Ho
24	230,51	Rejeita Ho
25	8,27	Rejeita Ho
26	6,96	Rejeita Ho
27	14,03	Rejeita Ho
28	8,35	Rejeita Ho
29	3,14	Rejeita Ho
30	9,87	Rejeita Ho
31	6,57	Rejeita Ho
34	5,00	Rejeita Ho
40	98,82	Rejeita Ho
41	198,09	Rejeita Ho
42	338,03	Rejeita Ho
47	11,36	Rejeita Ho
48	7,14	Rejeita Ho
49	59,63	Rejeita Ho
50	4,03	Rejeita Ho
52	17,36	Rejeita Ho

O teste *t-student* pode ser aplicado para o primeiro conjunto de dados (tabela 22), pois foi confirmada a igualdade entre as variâncias.

No segundo grupo (tabela 23), o teste *t-student* não pôde ser aplicado diretamente, pois não cumpre com o requisito de igualdade entre as variâncias. Por tanto, seguindo o fluxo definido por Leal et al (2011), o método Smith-Satterthwaite foi aplicado para ajustar os graus de liberdade para obter o valor crítico de t.

São definidas duas hipóteses para cada estação:

- $\boldsymbol{H_0}$**:** a média do número de usuários que ingressa em cada estação, obtida com base nos dados providenciados pela empresa, é igual a média do número de passageiros que ingressa em cada estação obtida com o *ProModel*.
- $\boldsymbol{H_1}$: não existem indícios de que as médias sejam semelhantes.

Para os dois grupos, o cálculo de t foi realizado com a ferramenta excel e replicado para todas as estações conforme a equação 12.

Para o primeiro grupo, o $\boldsymbol{t_{critico}}$ foi procurado nas tabelas para n-1 (9) graus de liberdade e 95% de confiança.

Para o primeiro grupo da tabela 22, no qual o teste de igualdade de variância foi confirmado, tem-se:

$$\boldsymbol{t_{critico\,\alpha/2} = 2,262}$$

Para o segundo grupo de dados, $\boldsymbol{t_{critico\,\alpha/2}}$ tomou diferentes valores, de acordo com o número de graus de liberdade calculado com o método Smith-Satterthwaite, na tabela 24 são apresentandos os resultados do teste de hipóteses para as 52 estações analisadas.

Este teste permitiu concluir que, para todas as estações, os resultados da Simulação têm médias semelhantes com os dados reais providenciados pela empresa.

Confirmando a hipótese nula e validando os resultados da simulação.

Tabela 23 - Teste de Hipóteses, comparação das médias com dados reais e simulados usando Teste t de *student*

	Entrada de Passageiros						
	Dados Reais		Dados Simulados				
Estação	Média	Desvio Padrão	Média	t teste	Graus de Liberdade	t critico	Resultado
1	4.317	216	4.337	1,2	154	2,0	Aceita Ho
2	1.637	74	1.619	-1,5	41	2,0	Aceita Ho
3	988	55	968	-1,5	16	2,1	Aceita Ho
4	1.767	74	1.757	-0,5	15	2,1	Aceita Ho
5	1.077	55	1.049	-1,7	11	2,3	Aceita Ho
6	2.747	132	2.763	0,5	15	2,1	Aceita Ho
7	1.332	109	1.317	-0,6	19	2,1	Aceita Ho
8	1.895	87	1.918	0,8	10	2,3	Aceita Ho
9	1.574	109	1.545	-0,8	10	2,3	Aceita Ho
10	1.225	129	1.201	-1,2	39	2,0	Aceita Ho
11	1.907	487	1.897	-0,2	131	2,0	Aceita Ho
12	1.224	91	1.203	-0,8	13	2,2	Aceita Ho
13	3.168	237	3.126	-0,5	10	2,3	Aceita Ho
14	2.658	181	2.627	-0,6	14	2,1	Aceita Ho
15	6.320	483	6.413	1,1	32	2,0	Aceita Ho
16	613	56	602	-0,8	16	2,1	Aceita Ho
17	1.384	90	1.335	-2,0	14	2,1	Aceita Ho
18	2.763	267	2.767	0,1	26	2,1	Aceita Ho
19	630	53	627	-0,2	13	2,2	Aceita Ho
20	3.104	112	3.106	0,1	12	2,3	Aceita Ho
21	785	40	782	-0,2	10	2,3	Aceita Ho
22	830	39	832	0,2	11	2,3	Aceita Ho
23	3.063	120	3.065	0,2	165	2,0	Aceita Ho
24	8.224	687	8.127	-1,8	168	2,0	Aceita Ho
25	1.547	86	1.532	-0,8	20	2,1	Aceita Ho
26	1.379	131	1.339	-1,3	18	2,1	Aceita Ho
27	6.071	345	6.036	-0,6	30	2,0	Aceita Ho
28	2.163	163	2.183	0,5	21	2,1	Aceita Ho
29	819	54	819	0,0	13	2,2	Aceita Ho
30	5.069	442	5.186	1,3	23	2,1	Aceita Ho
31	1.493	148	1.471	-0,7	18	2,1	Aceita Ho
32	2.040	78	1.991	-2,1	11	2,3	Aceita Ho
33	3.344	121	3.331	-0,4	12	2,3	Aceita Ho
34	2.639	147	2.619	-0,5	15	2,1	Aceita Ho
35	1.895	69	1.862	-1,5	11	2,3	Aceita Ho
36	1.614	70	1.616	0,1	12	2,3	Aceita Ho
37	1.237	59	1.249	0,7	11	2,3	Aceita Ho
38	1.103	50	1.099	-0,2	10	2,3	Aceita Ho
39	3.350	138	3.313	-0,9	11	2,3	Aceita Ho
40	4.513	198	4.535	1,4	147	2,0	Aceita Ho
41	1.241	86	1.235	-1,0	167	2,0	Aceita Ho
42	1.845	178	1.828	-1,2	168	2,0	Aceita Ho
43	1.324	64	1.283	-2,2	12	2,3	Aceita Ho
44	741	44	768	2,2	12	2,3	Aceita Ho
45	534	32	549	1,6	12	2,3	Aceita Ho
46	253	19	254	0,2	12	2,3	Aceita Ho
47	1.745	111	1.735	-0,5	26	2,1	Aceita Ho
48	1.328	89	1.293	-1,7	19	2,1	Aceita Ho
49	5.564	810	5.446	-1,6	113	2,0	Aceita Ho
50	1.484	124	1.452	-1,0	14	2,1	Aceita Ho
51	490	45	500	0,8	11	2,3	Aceita Ho
52	1.044	49	1.051	0,9	37	2,0	Aceita Ho

4.6.2. Validação Preditiva

Como as informaçãos do fluxo de passageiros nas horas de pico não foram disponiblizadas pela companhia, foram calculados os indicadores de tempo de viagem (do passageiro), tempo de viagem do trem, tempo de espera médio e lotação dos trens, assumindo uma demanda nas horas de pico igual a 3 vezes o valor dos horários de vale; este valor de 3, permitiria atingir os níveis de lotação de 12 $pessoas/m^2$.

Assim, estes dados permitiram realizar uma análise de sensibilidade, no qual os dados de entrada do modelo foram chamados de **Sistema Atual Hora Pico** (SAHP), onde foram geradas e analisadas as novas estratégias de planejamento das operações (tabela 25, 26, e 27)

Tabela 24 - Taxa de Ocupação SAHP dos trens da Linha Azul

LINHA AZUL				
Estação	**Lotação Minima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
12	72	1.477	1.332	119%
11	133	1.452	1.307	117%
10	94	1.300	1.175	105%
7	96	1.220	1.127	101%
9	75	1.228	1.105	99%
15	285	1.366	1.098	98%
16	17	1.342	1.083	97%
6	213	1.144	1.058	95%
8	93	1.163	1.046	94%
17	52	1.264	1.042	93%
18	79	1.197	1.017	91%
19	25	1.179	1.011	91%
14	156	1.301	1.003	90%
13	73	1.357	995	89%
20	118	1.040	905	81%
5	72	955	889	80%
21	20	1.012	887	80%
22	23	998	878	79%
4	145	891	835	75%
1	428	944	709	64%
3	76	732	694	62%
2	146	648	617	55%
23	-	-	-	0%

Tabela 25 - Taxa de Ocupação SAHP dos trens Linha Vermelha

LINHA VERMELHA				
Estação	**Lotação Mínima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
24	1.632	3.300	3.298	277%
25	276	1.936	3.212	269%
27	1.003	2.932	2.841	238%
28	304	3.141	2.780	233%
36	148	3.337	2.775	233%
26	228	2.125	2.728	229%
35	186	3.347	2.725	229%
38	90	3.326	2.712	228%
37	68	3.209	2.704	227%
34	270	3.160	2.663	223%
39	233	3.342	2.622	220%
33	372	2.724	2.593	218%
32	196	2.680	2.555	214%
30	552	2.668	2.551	214%
31	167	2.665	2.547	214%
29	96	2.656	2.544	213%
13	737	2.621	2.503	210%
40	-	-	-	

Tabela 26 - Taxa de Ocupação SAHP dos trens da Linha Verde

ESTAÇÃO VERDE				
Estação	**Lotação Mínima**	**Lotação Máxima**	**Lotação Média**	**Taxa de Lotação**
49	439	1.020	901	81%
50	84	983	872	79%
51	16	958	851	77%
48	120	920	763	69%
47	177	893	714	64%
9	58	829	611	55%
46	28	575	543	49%
8	170	696	525	47%
45	64	549	518	47%
44	66	477	451	41%
43	141	413	389	35%
42	98	272	257	23%
41	151	330	242	22%
52	-	-	-	0%

4.7. DELINEAMENTO DE EXPERIMENTOS

Os cenários foram criados usando as equações 21 e 22 apresentadas na metodologia resultando nos valores normalizados apresentados nas tabelas 29, 29 e 30 para cada uma das linhas.

Tabela 27 - Normalização dos dados da Linha Azul

	Xo (Pont. Medio)	Delta X	Xi(codificado) Inferior	Xi(codificado) Superior
Velocidade Média (m/min)	622,0	124,4	412,8	831,2
Headway (segundos)	131	36	70,5	191,5
Tempo abertura das portas	20	5	11,6	28,4

Tabela 28 - Normalização dos dados da Linha Verde

	Xo (Pont. Medio)	Delta X	Xi(codificado) Inferior	Xi(codificado) Superior
Velocidade Média (m/min)	614,0	122,8	407,5	820,5
Headway (segundos)	161	66	50,0	272,0
Tempo abertura das portas	20	5	11,6	28,4

Tabela 29 - Normalização dos dados da Linha Vermelha

	Xo (Pont. Medio)	Delta X	Xi(codificado) Inferior	Xi(codificado) Superior
Velocidade Média (m/min)	700,0	140,0	464,5	935,5
Headway (segundos)	119	24	78,6	159,4
Tempo abertura das portas	20	5	11,6	28,4

A matriz consolidada com os experimentos para cada uma das linhas é apresentada na tabela 31.

Cada experimento corresponde a um cenário que é analisado individualmente na sequência.

Tabela 30 - Matriz de experimentação com valores, experimento fatorial rotacional 2^k

	Azul			Verde			Vermelha		
1	497,6	95,0	15,0	491,2	95,0	15,0	560,0	95,0	15,0
2	746,4	95,0	15,0	736,8	95,0	15,0	840,0	95,0	15,0
3	497,6	167,0	15,0	491,2	227,0	15,0	560,0	143,0	15,0
4	746,4	167,0	15,0	736,8	227,0	15,0	840,0	143,0	15,0
5	497,6	95,0	25,0	491,2	95,0	25,0	560,0	95,0	25,0
6	746,4	95,0	25,0	736,8	95,0	25,0	840,0	95,0	25,0
7	497,6	167,0	25,0	491,2	227,0	25,0	560,0	143,0	25,0
8	746,4	167,0	25,0	736,8	227,0	25,0	840,0	143,0	25,0
9	622,0	131,0	20,0	614,0	161,0	20,0	700,0	119,0	20,0
10	622,0	131,0	20,0	614,0	161,0	20,0	700,0	119,0	20,0
11	622,0	131,0	20,0	614,0	161,0	20,0	700,0	119,0	20,0
12	412,8	131,0	20,0	407,5	161,0	20,0	464,5	119,0	20,0
13	831,2	131,0	20,0	820,5	161,0	20,0	935,5	119,0	20,0
14	622,0	70,5	20,0	614,0	50,0	20,0	700,0	78,6	20,0
15	622,0	191,5	20,0	614,0	272,0	20,0	700,0	159,4	20,0
16	622,0	131,0	11,6	614,0	161,0	11,6	700,0	119,0	11,6
17	622,0	131,0	28,4	614,0	161,0	28,4	700,0	119,0	28,4

4.7.1. Análise de Cenários – Visão Geral

Conforme apresentado na tabela 32, as linhas foram classificadas de acordo com a sua demanda, em Linha de Demanda Média (Azul), Baixa (Verde) e Alta (Vermelha)

Tabela 31 - Classificação da Demanda de Passageiros

	Média Passageiros / hora	**Total**
Linha Vermelha (Demanda Alta)	2853	45%
Linha Azul (Demanda Média)	2044	32%
Linha Verde (Demanda Baixa)	1466	23%

Os resultados dos experimentos são apresentados nas tabelas 33 para a linha Azul, 34 para a linha Vermelha, e 35 para a linha Verde.

Os resultados apresentados na tabela 33 mostram que para uma linha com *Demanda Média (Azul),* os resultados são melhores quando comparados com o resultado atual. Estes resultados aparecem coloridos para serem diferenciados dos outros cenários simulados.

- Para o fator de decisão *Tempo de Espera*, os melhores cenários são: o Experimentos 2 e o Experimento 6, a decisão na mudança destes fatores permitiria uma redução no tempo de espera de até 10%.
- Para o fator de decisão *Lotação*, é recomendável implementar os resultados obtidos com os cenários Experimento 2 e Experimento 6, pois o primeiro permite uma redução de 15,3% no nível de ocupação do trem, enquanto o segundo uma redução de 15% no mesmo indicador.
- Já para o Tempo de Viagem, o modelo comprovou que os melhores cenários são Experimento 4 e Experimento 13, pois permitem uma redução do tempo de viagem de 26,6% e 29,8%, respectivamente.
- Considerando o fator de decisão *tamanho da fila*, os cenários Experimento 2 e Experimento 6 trariam uma redução de 14,95% e 14,98%, respectivamente.

Tabela 32 - Resultado dos Cenários Simulados na Linha Azul

Experimento	LINHA AZUL						
	Fatores				Respostas		
	Velocidade Comercial (m/min)	Headway (segundos)	Tempo abertura das portas	Tempo de Espera	Lotação Média	Tempo Viagem 1-23 Médio	Tamanho da fila Médio
Ex1	497,6	95,0	15,0	1,5	1.184,0	65,9	115,8
Ex2	746,4	95,0	15,0	**1,0**	**803,7**	51,4	**78,2**
Ex3	497,6	167,0	15,0	1,5	1.270,9	54,5	123,1
Ex4	746,4	167,0	15,0	1,5	1.268,7	**36,2**	122,7
Ex5	497,6	95,0	25,0	1,4	1.189,8	68,7	115,6
Ex6	746,4	95,0	25,0	**1,0**	**806,1**	54,2	**78,1**
Ex7	497,6	167,0	25,0	1,5	1.275,5	58,0	123,2
Ex8	746,4	167,0	25,0	1,5	1.267,8	39,9	122,9
Ex9	622,0	131,0	20,0	1,2	998,0	49,1	99,9
Ex10	622,0	131,0	20,0	1,2	954,6	49,1	95,6
Ex11	622,0	131,0	20,0	1,2	954,6	49,1	95,6
Ex12	412,8	131,0	20,0	1,7	1.389,6	73,8	133,8
Ex13	831,2	131,0	20,0	1,1	951,8	**34,6**	91,7
Ex14	622,0	70,5	20,0	1,1	915,5	59,4	90,3
Ex15	622,0	191,5	20,0	1,6	1.394,8	44,0	134,6
Ex16	622,0	131,0	11,6	1,1	955,6	46,4	92,8
Ex17	622,0	131,0	28,4	1,1	954,5	51,9	91,7
Ex18	622,0	95,0	20,0	1,1	915,5	59,4	90,3
Ex19	622,0	167,0	20,0	1,4	1.213,8	43,6	90,3
Atual				1,1	948,3	49,3	91,9

Na tabela 33, são apresentados os resultados dos cenários simulados em uma linha com *Demanda Alta (Vermelha).* Os melhores resultados (comparados com o sistema atual) aparecem coloridos para serem diferenciados dos outros cenários simulados.

- Se o fator de decisão fosse *Tempo de Espera*, os melhores cenários seriam Experimento 2 e Experimento 4, a decisão na mudança destes fatores permitiria uma redução no tempo de espera de até 7,9%.
- Se o fator de decisão fosse a *Lotação*, é recomendável implementar os resultados obtidos nos cenários Experimento 8 e Experimento 16, pois o primeiro, permite uma redução de 7,2% no nível de ocupação do trem, enquanto o segundo uma redução de 9,5% no mesmo indicador.
- Já para o Tempo de Viagem, o modelo comprovou que os melhores cenários seriam Experimento 2, e Experimento 4, os quais permitiriam reduzir o tempo de viagem em 12,5%.

- Se o fator de decisão fosse *tamanho da fila*, os cenários Experimento 2 e Experimento 4 trariam os melhores resultados, isto é uma redução de 4,6% no tamanho da fila para ambos os casos.

Tabela 33 - Resultado dos Cenários Simulados na Linha Vermelha

LINHA VERMELHA							
	Factores				Respostas		
Experimento	Velocidade Comercial (m/min)	Headway (segundos)	Tempo abertura das portas	Tempo de Espera	Lotação Média	Tempo Viagem 1-23 Médio	Tamanho da fila Médio
Ex1	560	95	15	3,16	2.592	51,08	395
Ex2	840	95	15	**2,43**	2.582	**39,00**	**349**
Ex3	560	143	15	3,16	2.592	51,08	395
Ex4	840	143	15	**2,43**	2.582	**39,00**	**349**
Ex5	560	95	25	4,02	2.794	56,27	457
Ex6	840	95	25	2,77	2.530	44,66	374
Ex7	560	143	25	4,02	2.794	56,27	457
Ex8	840	143	25	2,77	**2.530**	44,63	374
Ex9	700	119	20	2,80	2.763	46,60	374
Ex10	700	119	20	2,80	2.739	46,60	374
Ex11	700	119	20	2,80	2.739	46,60	374
Ex12	465	119	20	2,83	2.639	67,07	379
Ex13	935	119	20	2,49	2.612	39,19	357
Ex14	700	79	20	2,82	2.760	46,60	376
Ex15	700	159	20	2,82	2.763	46,60	378
Ex16	700	119	12	2,48	**2.467**	42,13	355
Ex17	700	119	28	3,40	2.739	50,83	415
Ex18	700	95	20	2,82	2.760	46,60	376
Ex19	700	143	20	2,83	2.762	46,60	379
Atual				2,64	2.727	44,55	366

Na tabela 34, são apresentados os resultados dos cenários simulados em uma linha com *Demanda Baixa (Verde)*, os melhores resultados (comparados com o sistema atual) aparecem coloridos para serem diferenciados dos outros cenários simulados.

- Se o fator de decisão fosse *Tempo de Espera*, os melhores cenários seriam Experimento 2 e Experimento 6, a decisão na mudança destes fatores permitiria uma redução no tempo de espera de até 19,7%.
- Se o fator de decisão fosse a *Lotação*, é recomendável implementar os cenários Experimento 2 e Experimento 6, pois o primeiro, permite uma redução de 19,5% no nível de ocupação do trem, enquanto o segundo uma redução de 19,3% no mesmo indicador.

- Já para o Tempo de Viagem, o modelo comprovou que os melhores cenários seriam Experimento 4, e Experimento 13, pois o primeiro permitiria reduzir o tempo de viagem em 12,7% enquanto o segundo permitiria chegar numa redução de 19,3%
- Se o fator de decisão fosse tamanho da fila, os cenários Experimento 2 e Experimento 6 trariam os melhores resultados, isto seria, uma redução de 19,9% para o primeiro cenário e uma redução de 19,1% para o segundo.

Tabela 34 - Resultado dos Cenários Simulados na Linha Verde

LINHA VERDE							
	Fatores				Respostas		
Experimento	Velocidade Comercial (m/min)	Headway (segundos)	Tempo abertura das portas	Tempo de Espera	Lotação Média	Tempo Viagem 1-23 Médio	Tamanho da fila Médio
Ex1	491,2	95,0	15,0	1,7	710,0	43,1	96,6
Ex2	736,8	95,0	15,0	**1,1**	**473,0**	29,8	**63,9**
Ex3	491,2	227,0	15,0	2,0	826,4	36,9	112,0
Ex4	736,8	227,0	15,0	2,0	825,6	**27,2**	112,5
Ex5	491,2	95,0	25,0	1,7	710,0	44,9	96,2
Ex6	736,8	95,0	25,0	**1,1**	**474,1**	31,5	**64,6**
Ex7	491,2	227,0	25,0	2,0	825,9	39,1	112,2
Ex8	736,8	227,0	25,0	2,0	826,9	29,3	111,6
Ex9	614,0	161,0	20,0	1,4	587,1	31,0	79,5
Ex10	614,0	161,0	20,0	1,4	587,1	31,0	79,5
Ex11	614,0	161,0	20,0	1,4	587,1	31,0	79,5
Ex12	407,5	161,0	20,0	2,0	852,8	51,9	115,2
Ex13	820,5	161,0	20,0	1,4	587,5	**25,1**	79,9
Ex14	614,0	50,0	20,0	1,4	568,2	36,0	77,2
Ex15	614,0	272,0	20,0	2,4	991,8	32,9	135,1
Ex16	614,0	161,0	11,6	1,4	588,2	29,3	79,4
Ex17	614,0	161,0	28,4	1,4	587,4	32,7	79,4
Ex18	614,0	95,0	20,0	1,4	568,2	36,0	77,2
Ex19	614,0	227,0	20,0	1,4	568,9	35,6	76,9
Atual				1,4	587,5	31,1	79,8

4.7.2. Análise de Cenários – Usando *ANOVA*

Com o objetivo de encontrar os fatores que tem efeito sobre cada uma das variáveis, foram comparados os resultados de cada um dos cenários simulados usando ANOVA.

A análise foi feita para cada uma das linhas usando as seguintes Hipóteses.

- H_0: Todas as médias são iguais.
- H_1: As médias são diferentes

4.7.2.1. Simulação da Linha Verde

Nesta linha, caraterizada como uma linha de demanda baixa, foram analisadas o Tempo de Espera, Tempo de Viagem, Lotação e Tamanho da Fila.

- Tempo de Espera

Os resultados apresentados na tabela 36, apontam que o único fator que compromete o tempo de espera para um sistema com baixa demanda é a velocidade, o tempo de espera pode aumentar em 30,9% quando a velocidade se reduz de 820,5 m/min para 407,5 m/min.

Tabela 35 - Fatores que afetam o Tempo de Espera da Linha Verde

TEMPO DE ESPERA								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Espera	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***
Velocidade	4,10	6,94	Velocidade	17,30	6,94	Headway	2,43	6,94
Headway	1,84	6,94	T.Abertura Portas	0,33	6,94	T.Abertura Portas	0,64	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

- Tempo de Viagem

Na tabela 37, os resultados da simulação mostram que tanto a velocidade como o *headway* comprometem o tempo de viagem dos passageiros em uma linha de trem com demanda baixa. O maior tempo de viagem é encontrado com uma velocidade inferior a 491,2 m/min. O tempo de viagem pode a passar aumentar em 51% se a velocidade for reduzida de 820,5 m/min para 407,5 m/min.

Por outro lado, embora o impacto do *headway* não seja tão significativo, o tempo de viagem pode aumentar em 8% quando um *headway* passa de 272s para 50s.

Tabela 36 - Fatores que afetam o Tempo da Viagem da Linha Verde

TEMPO DE VIAGEM								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Viagem	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***
Velocidade	7,37	6,94	Velocidade	11,52	6,94	Headway	8,77	6,94
Headway	0,24	6,94	T.Abertura Portas	0,68	6,94	T.Abertura Portas	3,17	6,94
Velocidade	F>Fcri Rejeita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F>Fcri Rejeita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

Fonte: o autor

- Lotação

Conforme observado na tabela 38, o único fator que compromete a lotação para um sistema de demanda baixo é a velocidade. Os níveis de ocupação caem em 31,1% quando a velocidade passa de 407,5m/min para 820,5m/min

Tabela 37- Fatores que afetam a Lotação da Linha Verde

LOTAÇÃO								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Lotação	***F***	***F crit***	***Lotação***	***F***	***F crit***	***Lotação***	***F***	***F crit***
Velocidade	4,05	6,94	Velocidade	16,68	6,94	Headway	2,49	6,94
Headway	1,88	6,94	T.Abertura Portas	0,29	6,94	T.Abertura Portas	0,67	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Rejeita Ho	

Fonte: o autor

- Tamanho da Fila

Os resultados na tabela 39, mostram que a velocidade é o fator que tem maior impacto no tamanho da fila e os maiores níveis são encontrados em uma velocidade de 491,2m/min. Os resultados demonstram que uma redução na velocidade de 820,5m/min para 407,5m/min aumenta em 30,64% o tamanho médio da fila nas estações do trem.

Tabela 38 - Fatores que afetam o Tamanho da Fila da Linha Verde

TAMANHO DA FILA								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tamanho da Fila	***F***	***F crit***	***Tamanho da Fila***	***F***	***F crit***	***Tamanho da Fila***	***F***	***F crit***
Velocidade	4,06	6,94	Velocidade	17,53	6,94	Headway	2,43	6,94
Headway	1,87	6,94	T.Abertura Portas	0,31	6,94	T.Abertura Portas	0,65	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F<Fcri Aceita Ho	

Fonte: o autor

4.7.2.2. Simulação da Linha Azul

Nesta linha, caraterizada como uma linha de demanda média, foram analisadas o Tempo de Espera, Tempo de Viagem, Lotação e Tamanho da Fila.

- Tempo de Espera

Na tabela 40, as análises demonstram que a Velocidade e o *Headway* são os principais fatores que comprometem o tempo de espera dos passageiros e o menor nível encontra-se com uma velocidade de 746,4m/min e um *headway* de 95 segundos, enquanto os maiores tempos foram encontrados com *Headway* superior a 95 segundos junto com uma velocidade menor que 746,4m/min. Para este nível de demanda, o tempo de abertura de portas não representa um fator de impacto para o tempo de espera dos passageiros.

O tempo de espera é reduzido em 34,88% quando passa de uma velocidade de 412,9m/min para uma velocidade 831,2m/min, o mesmo ocorre com 45,5% quando o *headway* passa de 191,5m/min para 70,5m/min.

Tabela 39 - Fatores que afetam o Tempo de Espera na Linha Azul

TEMPO DE ESPERA								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Espera	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***
Velocidade	4,23	6,94	Velocidade	11,92	6,94	Headway	16,59	6,94
Headway	2,38	6,94	T.Abertura Portas	0,61	6,94	T.Abertura Portas	0,02	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F>Fcri Rejeita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

Fonte: o autor

- Tempo de Viagem

De igual maneira como aconteceu na análise do tempo de espera, na tabela 41, as análises demonstram que os fatores que afetam o comportamento do tempo de viagem são a Velocidade e o *Headway*, resultado que era de se esperar, pois quanto maior a velocidade, menor o tempo de deslocamento. Os maiores tempos de viagem, são encontrados para uma velocidade menor de 497,6 m/min junto com um *headway* menor de 167s.

Por outro lado, os menores tempos foram registrados para uma velocidades superior aos 746,4m/min combinado com *headway* no intervalo de 131 s até 167 s. Uma variação na velocidade de 831,2 m/min para 412,8m/min, produz um aumento de 53% no tempo de viagem dos passageiros. Comprovou-se igualmente que o tempo de viagem é reduzido em 26% quando o *headway* aumenta de 70,5s para 191,5s, resultado que se deve ao fato de que uma menor frequência na saída dos trens pode levar a bloqueios na pista, já que foi simulada uma só via para cada linha, mesmo porque os trens não podem sobrepassar os outros, causando assim atrasos no tempo de deslocamento do trem.

Para este nível de demanda, o tempo de abertura de portas não tem nenhuma influência no comportamento do tempo de viagem dos passageiros.

Tabela 40 - Fatores que afetam o Tempo da Viagem na Linha Azul

TEMPO DE VIAGEM								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Viagem	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***
Velocidade	4,23	6,94	Velocidade	11,92	6,94	Headway	16,59	6,94
Headway	2,38	6,94	T.Abertura Portas	0,61	6,94	T.Abertura Portas	0,02	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F>Fcri Rejeita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

Fonte: o autor

- Lotação

A tabela 42 mostra que os principais fatores que comprometem a lotação dos trens para um nível de demanda médio são a velocidade e o *headway*. Velocidades superiores a 633m/min apresentam os menores níveis de lotação, por outro lado, *headways* inferiores aos 131s também permitem reduzir os níveis de lotação dos trens, pois reduzem a frequência em que os trens passam pelas estações.

A lotação aumenta em 31,5% quando a velocidade passa de 831m/min para 412m/min. Dessa forma, velocidades baixas não são recomendáveis quando procura-se uma redução na ocupação dos trens. Um comportamento similar acontece quando o *headway* é reduzido de 191,5s para 70,5s, e esta redução no tempo, também gera uma redução de 34,4% no nível de ocupação dos trens.

Tabela 41 - Fatores que afetam a Lotação na Linha Azul

			LOTAÇÃO					
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Lotação	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***
Velocidade	4,36	6,94	Velocidade	10,21	6,94	Headway	24,98	6,94
Headway	3,74	6,94	T.Abertura Portas	0,28	6,94	T.Abertura Portas	0,07	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F>Fcri Rejeita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

Fonte: o autor

- Tamanho da Fila

O único fator que compromete o tamanho da fila é a velocidade. O modelo demonstrou que velocidades inferiores a 491m/min, geram um aumento no número de passageiros na fila aguardando para embarcar. Analisando o intervalo completo de movimento da velocidade, o tamanho da fila é reduzida em 32% ao passar de uma velocidade de 412,8m/min para uma velocidade de 831,2m/min, tal como apresentado na tabela 43.

Tabela 42 - Fatores que afetam o Tamanho da Fila na Linha Azul

			TAMANHO DA FILA					
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tamnho da Fila	***F***	***F crit***	***Tamnho da Fila***	***F***	***F crit***	***Tamnho da Fila***	***F***	***F crit***
Velocidade	4,06	6,94	Velocidade	17,53	6,94	Headway	2,43	6,94
Headway	1,87	6,94	T.Abertura Portas	0,31	6,94	T.Abertura Portas	0,65	6,94
Velocidade	F<Fcri Aceita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho	

4.7.2.3. Simulação da Linha Vermelha

Nesta linha, caraterizada como uma linha de demanda alta, foram analisadas o Tempo de Espera, Tempo de Viagem, Lotação e Tamanho da Fila.

- Tempo de Espera

Quando analisado o tempo de espera em uma linha com nível demanda alta tal como apresentado na tabela 44, observa-se que tanto a velocidade como o tempo de abertura das portas comprometem a eficiência do sistema. O maior tempo de espera acontece para uma velocidade inferior a 700m/min

Neste cenário, não existem evidências que confirme que o Headway tem impacto neste tipo de demanda.

Tabela 43 - Fatores que afetam o Tempo de Espera da Linha Vermelha

TEMPO DE ESPERA								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Espera	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***	***Tempo de Espera***	***F***	***F crit***
Velocidade	7,70	6,94	Velocidade	6,32	6,94	Headway	1,76	6,94
Headway	1,29	6,94	T.Abertura Portas	6,84	6,94	T.Abertura Portas	48,57	6,94
Velocidade	F>Fcri Rejeita Ho		*Velocidade*	F<Fcri Aceita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F>Fcri Rejeita Ho	

- Tempo de Viagem

Para estes níveis de demanda a velocidade e o tempo de abertura de portas também comprometem o tempo de viagem dos passageiros, sendo que uma velocidade menor de 846m/min produz os maiores tempos de viagem, incrementando o tempo em 41,57% quando passa de 935 m/min para 464,5 m/min. O TAP influencia o tempo de viagem, encontrando os maiores valores quando o tempo é maior que 20s; o aumento no TAP de 11,6s para 28,4s produz um incremento em 17,11% no tempo de viagem dos passageiros, tal como apresentado na tabela 45. Não existem indícios que demonstrem que o headway compromete o desempenho deste fator.

Tabela 44 - Fatores que afetam o Tempo da Viagem da Linha Vermelha

TEMPO DE VIAGEM								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tempo de Viagem	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***	***Tempo de Viagem***	***F***	***F crit***
Velocidade	9,34	6,94	Velocidade	9,09	6,94	Headway	0,00	6,94
Headway	0,52	6,94	T.Abertura Portas	1,77	6,94	T.Abertura Portas	12,38	6,94
Velocidade	F>Fcri Rejeita Ho		*Velocidade*	F>Fcri Rejeita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F>Fcri Rejeita Ho	

Fonte: o autor

- Lotação

Assim como aconteceu com os fatores anteriormente apresentados, na tabela 46 se nota que a lotação é comprometida por variações na Velocidade e TAP. A velocidade, só chega impactar a lotação em 5,1%, quando é levada de 840/min para 560m/min. Devido a esta pequena variação a resolução da figura 31, não permite encontrar nenhum padrão.

Por outro lado, o TAP, também compromete a lotação do trem, pois a simulação demonstrou que a lotação aumenta em 10% quando o TAP passa de 11,6s para 28,4s, isto porque, embora a capacidade do trem seja restrita, quanto maior for o tempo em que o trem permanece em uma estação, maior é a probabilidade de que os passageiros tentem entrar nos vagões.

Tabela 45 - Fatores que afetam a Lotação da Linha Vermelha

LOTAÇÃO								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Lotação	***F***	***F crit***	***Lotação***	***F***	***F crit***	***Lotação***	***F***	***F crit***
Velocidade	12,54	6,94	Velocidade	0,78	6,94	Headway	0,35	6,94
Headway	0,33	6,94	T.Abertura Portas	1,62	6,94	T.Abertura Portas	8,69	6,94
Velocidade	F>Fcri Rejeita Ho		*Velocidade*	F<Fcri Aceita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F<Fcri Aceita Ho		*TAP*	F>Fcri Rejeita Ho	

- Tamanho da Fila

Na tabela 47 é demonstrado que tanto a velocidade como o TAP continuam a serem os fatores que comprometem os resultados deste indicador, pois uma velocidade menor de 770m/min, aumenta o tamanho das filas, e a mesma tendência é observada com TAP

maiores de 20s. O tamanho da fila aumenta em 14,45% quando o TAP aumenta de 11,6s para 38,4s.

Tabela 46 - Fatores que afetam o Tamanho da Fila da Linha Vermelha

TAMANHO DA FILA								
Velocidade vs Headway			Velocidade vs Tempo de Abertura de Portas			Headway vs Tempo de Abertura de Portas		
Tamanho da Fila	***F***	***F crit***	***Tamanho da Fila***	***F***	***F crit***	***Tamanho da Fila***	***F***	***F crit***
Velocidade	8,25	6,94	Velocidade	6,34	6,94	Headway	2,54	6,94
Headway	1,23	6,94	T.Abertura Portas	7,31	6,94	T.Abertura Portas	81,24	6,94
Velocidade	F>Fcri Rejeita Ho		*Velocidade*	F<Fcri Aceita Ho		*Headway*	F<Fcri Aceita Ho	
Headway	F<Fcri Aceita Ho		*TAP*	F>Fcri Rejeita Ho		*TAP*	F>Fcri Rejeita Ho	

Os resultados apresentados nas três linhas podem ser resumidos na tabela 48, onde se evidenciam quais as variáveis que compromentem os resultados dos fatores para cada tipo de demanda.

Tabela 47 – Identificação das variáveis críticas obtidas a partir da análise de cenários usando ANOVA

Tipo de Demanda	T. Espera	T. Viagem	Lotação	Tamanho da Fila
Linha Verde (Demanda Baixa)	Velocidade	Velocidade, Headway	Velocidade	Velocidade
Linha Azul (Demanda Média)	Velocidade, Headway	Velocidade, Headway	Velocidade, Headway	Velocidade
Linha Azul (Demanda Alta)	Velocidade, TAP	Velocidade, TAP	Velocidade, TAP	Velocidade, TAP

Esta tabela pode ser considerada uma matriz de decisão que pode auxiliar os gestores de uma empresa de transporte em massa, na tomada de decisão, quando precisassem estabelecer quais as variáveis que devem ser trabalhadas conforme os requerimentos da empresa e o objetivo estratégico. Isto permite focalizar os esforços da operacionais, atacando as variáveis que permitiram atingir os resultados esperados.

Algumas das conclusões que podem se obter a partir desta tabela são:

- Para uma linha com demanda **baixa**, é recomendável aumentar a velocidade, isto permitiria:

  Reduzir o tempo de espera em até **19,66%**;

  Reduzir o tempo de viagem em até **12,66%;**

Reduzir a lotação em até **19,49%**;

Reduzir o tamanho da fila em até **19,90%**;

- Para uma linha com demanda **média**, comprovou-se que tanto a velocidade como o *headway* podem comprometer o resultado dos indicadores, e a variação nestes indicadores permitiria:

  Reduzir o tempo de espera em até **10,0%**;

  Reduzir o tempo de viagem em até **29,85%.**

  Reduzir a lotação em até **15,25%**

  Reduzir o tamanho da fila em até **14,98%;**

- Para uma linha com demanda **alta**, é recomendável ajustar a velocidade, mantendo-a no limite máxico, e reduzir o tempo de abertura das portas:

  Reduzir o tempo de espera em até **7,91%**;

  Reduzir o tempo de viagem em até **12,46%**

  Reduzir a lotação em até **7,23%**

  Reduzir o tamanho da fila em **4,58%**.

5. CONCLUSÃO

A simulação de eventos discretos aplicada aos sistemas de transporte em massa pode ser considerada uma poderosa ferramenta para a investigação de cenários de operação, seus resultados são potencializados com o uso da metodogia DOE, a união destas duas metodologias, demonstrou ser útil estratégicamente, pois permitiu definir o impacto dos fatores pesquisados (*Headway*, Tempo de abertura das portas e velocidade), nas variáveis tempo de viagem, lotação, tempo de espera e tamanho da fila; os resultados aqui apresentados permitem um tomador de decisões definir as variáveis sobre as quais deve focar seus esforços com base nos objetivos e os *trade-off* definidos pela compania, por tanto, a definição do melhor cenário para ser adotado, depende da estratégia da empresa, pois a melhora de um indicador, pode resultar em algum dano a outros.

Igualmente, a simulação também permite avaliar o comportamento do sistema frente a diferentes tipos de demanda e eventualmente, pode ser usado com um sistema de planejamento das operações, permitindo aos tomadores de decisão analisar rápidamente o sistema e definir estratégias alinhadas as necessidades e comportamento específicos da operação.

A simulação demonstrou que, a principal variável que compromete todos os indicadores de um sistema de transporte em massa é a Velocidade, que permitiu uma redução dos indicadores analisados de até 40%; os resultados também demonstraram que, conforme o nível de demanda aumenta, o número de viariáveis que comprometem o sistema também aumenta.

Embora não tenham sido encontrados indícios na literatura de análises feitos com o factor tempo de abertura de portas como factor crítico, a simulação demonstrou que seu maior impacto acontece em sistemas de demanda alta, pois uma maior estadia dos trens nas estações gera aumento no tamanho da fila, o tempo de espera, o tempo da viagem e a lotação.

O processo de validação operacional de modelos de simulação proposto por Leal et al. (2011), demonstrou ser uma metodologia útil que permitiu aumentar a significância estatística do modelo, se comparado com os outros métodos de validação aplicados. Entregou um resultado confiável que garantiu que o modelo computacional criado, estivesse replicando o sistema real com um nível de confiança de 95%, assim mesmo, a

validação por eventos, também constitui um método quantitativo útil para comparar os resultados da simulação frente ao sistema real.

Embora, tanto os métodos de validação operacional, como a validação por eventos demostraram ser técnicas úteis para definir a confiabilidade do modelo, estes métodos não podem ser usados separadamente. A pesquisa realizada em simulação deve conter várias etapas de validação, começando com o método *Face-to-Face*, o qual acontece durante o processo de programação do modelo de simulação e finalizando com os métodos de validação por eventos e validação operacional que permitam quantificar os resultados.

Já o método de validação preditiva, embora seja importante, não permite chegar a conclusões finais e definir a validade do modelo, pois sua utilidade está mais focada no processo de criação de cenários, que conceitualmente permitiria obter resultados de novas situações e estratégias no sistema real e, não é possível quantificar a confiança do modelo.

Pode-se concluir que a simulação de eventos discretos, permite os pesquisadores entenderem em detalhe o funcionamento do sistema e assim mesmo proporciona a flexibilidade de se trabalhar com qualquer outra metodologia para o desenvolvimento dos cenários, começando desde a simulação de premissas e definição de estratégias específicas da empresa até a busca das melhores práticas através de métodos DOE e pesquisa operacional.

A significância estatística e a confiança do modelo, irá garantir que a implementação das mudanças no sistema real gerem os resultados esperados do modelo, por tanto, pode se concluir que, uma das etapas mais críticas da simulação é a validação, a qual não pode ser nem esquecida, nem reduzida a validação de animação, esta etapa esta alinhada diretamente com o objetivo final da simulação: representar com um alto nível de confiança um sistema real a traves de um modelo de computador.

5.1. SUGESTÕES PARA TRABALHOS FUTUROS

Seguem abaixo algumas sugestões para trabalhos futuros:

- Realizar análise focada nas rotas críticas das diferentes linhas, de forma a obter novas possibilidades de operação;
- Aplicar uma nova metodologia de pesquisa operacional para o desenvolvimento de novos cenários a fim de comparar as técnicas e obter novos resultados.

- Realizar análise econômico-financeira para conhecer o impacto das mudanças dos indicadores nos resultados financeiros da compania.
- Medir o impacto nas varíaveis de operaçõa do sistema, na integração de novas linhas ou a adição de novas estações.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, G. M. **Estatística Geral e Aplicada**. 2. ed. São Paulo: Atlas, 2002.

ASSIS, W. O.; MILANI, B. E. A. Generation of optimal schedules for metro lines using model predictive control. **Journal Automatica**, v.40, n.8, p. 1397-1404. 2004.

ASSIS, W. O.; MILANI, B. E. A. Geração de Programa Horário para Linha Metroviárias utilizando Programação Linear. In: **Anais do XIII CBA**. Florianópolis, 2000.

AZADIVAR, F. Simulation optimization methodologies. In: **Proceedings of the 1999 Winter Simulation Conference. Manhattan,** USA, v.1, p. 93-100, 1999.

BANKS J.; CARSON J. S.; NELSON, B. L.; NICOL, D. M. **Discrete-event Simulation System Simulation**. 4. ed. Upper Saddle River: Prentice-Hall, p. 640. 2005.

BARROS, B. N.; SCARMINIO, L. S.; BRUNS, R. E. **Planejamento e Otimização de Experimentos.** São Paulo: UNICAMP, 1995.

BEKKER, J.; VIVIERS, L. Using computer simulation to determine operations policies for a mechanised car park. **Simulation Modelling Practice and Theory**, South Africa, v. 16, p. 613–625, 2008.

BERTRAND, J. W. M.; FRANSOO, J. C. Modeling and simulation. Operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 241-264, 2002.

BOWDEN, R. O. Simulation optimization research and development. In: **Proceedings of the 1998 Winter Simulation Conference**. Mississippi, USA, v.2, p. 1693-1698, 1998.

CARDOSO, G. O. N.; COSTA, A. G.; TADEU, L. C. Projeto de Terceirização do Transporte Metroviário de Passageiros de São Paulo fundamentado em indicadores de desempenho**. VII Congresso Nacional de Exelência em Gestão**, Brasil, 2012.

CARDOZA, E. V. G. Aplicação das técnicas de planejamento e análise de experimentos na melhoria da qualidade de um processo de fabricação de produtos plásticos. 2002. 000f. **Dissertação (Mestrado em Engenharia de Produção)** - Universidade de São Paulo, São Carlos, 2002.

CARSON, J. S. Introduction to modeling and simulation. **Proceedings of the 2003 Winter Simulation Conference**, Marietta, USA, v.1, p. 7-13, 2003.

CARSON, Y.; MARIA, A. Simulation optimization: Methods and applications. In: **Proceedings of the 1997 Winter Simulation Conference**, USA, p. 118-126, 1997.
CHAWSHEEN, T. A.; LATIF, I. S. Detection and treatment of outliers data sets. **Irak Journal of Statistical Science**, v.9, p. 58-74, 2006.

CHUNG, C. A. **Simulation Modeling Handbook**: a practical approach. Washington, 2004.

CHWIF, L. Redução de modelos de simulacão de eventos discretos na sua concepcão: uma abordagem causal. 1999. 151 f. **Tese (Doutorado em Engenharia Mecânica)** – Escola Politécnica, Universidade de São Paulo, 1999.

CÔRREA, S. S.; MILANI, B. E. A. Regulação Robusta de Tráfego em Linhas de Metrô. **Revista Controle & Automação**, v. 12, n. 2, p. 118-130, 2001.

COSTA, I. P. Proposta de sistematização da simulação para fabricação em Lotes. 2000. 138 f. **Dissertação. (Mestrado em Engenharia Produção)** - Escola Federal de Engenharia de Itajubá, 2000.

CURY, J. E.; GOMIDE, F. A.; MENDES, M. J. A methodology for generation of optimal schedules for an underground railway system. **IEEE Transactions on Automatic Control**, n. 25, v.2, p. 217–222, 1980.

DA AGÊNCIA BRASIL. Metrô de SP é o mais lotado do mundo com 3,7 mi de usuários. Folha de São Paulo, São Paulo, Abr. 2011. Disponível em: <http://www1.folha.uol.com.br/cotidiano/906394-metro-de-sp-e-o-mais-lotado-do-mundo-com-37-mi-de-usuarios.shtml>. Acesso em: 5 ago. 2012.

DING, H.; BENYOUCEF L.; XIE X. A simulation optimization approach using genetic search for supplier selection. In: **Proceedings of the 2003 Winter Simulation Conference,** Estados Unidos, v. 2, p. 1260-1267, 2003.

DHINGRA, S. L.; KRISHNA, R. K. V.; MURALIDHAR, S. Public transport routing and scheduling using Genetic Algorithms. **8th International Conference on Computer-Aided Scheduling of Public Transport,** Berlin, Alemanha, p. 21-23, 2000.

FALLON, A.; SPADA, C. Detection and Accommodation of Outliers in Normally Distributed Data Set. **Environmental Sampling & Monitoring Primer**, 1997.

FONSECA, J.S. de; MARTINS, G. de A. **Curso de estatística**. 6 ed. São Paulo: Atlas, 1996. 320p.

FRAZELLE, E. **Worldclass warehousing and material handling**. 1st ed. The United States: McGrawHill Professional, 2002.

FU, M. C. Optimization for Simulation: Theory vs. Practice. **Journal on Computing**, v. 14, n 3, p. 192-215, 2002.

GIANNETTONI, M.; SAVIO, S. Fleet management in railway freight operations for rail transportation systems: the EU project F-Man. International Conference on Design and Operation in Railway Engineering (COMPRAIL), Dresden. 2004.

GIBB, J., Back Propagation Family Album, Technical Report C/TR96-05. Macquarie University, August, 1996. Disponível em: < http://citeseer.ist.psu.edu/gibb96back.html>. Acesso em: 11 jul. 2013.

GONZAGA, R. A. F. Planejamentos fatoriais fraccionados para análise de sensibilidade de modelos de simulação de eventos discretos. 2006. 134 f. **Dissertação (Mestrado em Engenharia Produção)** - Escola Federal de Engenharia de Itajubá, 2006.

GUAN, F.J.; YANG, H.; WIRASINGHE, S.C. Simultaneous optimization of transit line configuration and passenger line assignment. **Transportation Research Part B**. v. 40, 2006. p. 885 – 902.

HASAN, S.; AL-HUSSEIN, M.; GILLIS, P. Advanced simulation of tower crane operation utilizing system dynamics modeling and lean principles. Proceedings of the 2010 Winter Simulation Conference, p. 3262-3271, 2010.

HICKS, C. R.; TURNER, K. V. J. **Fundamental Concepts in the Design of Experiments.** 5 ed. Oxford University Press, 565 p, 1999.

HIGGINGS, A.; FERREIRA, L.; KOZAN, E. Optimization of train schedules to achieve minimum transit times and maximum reliability. In: **Proceedings of the 13th International Symposium on Transportation an Traffic Theory**, Lyon, France, p. 589-614, 1996

HU, J.; SHI, X.; SONG, J.; XU, Y. Optimal Design for Urban Mass Transit Network Based on Evolutionary Algorithms. **Computer Science**, v. 3611, p.1089 -1100, 2005.

IRANZO, D. Perez. Análise de Outliers: Um estudo de caso. 2007. 382 f. **Tese (Doutorado em Economia)** – Universidade de Valecia, Espanha, 2007.

ISHIKAWA, K. Guide quality control. **Asian Productivity**, Tokio, Japão, p. 150 – 180, 1976.

JUSTEK, A.; PEÑA, D.; ZAMAR, R.. A multivariate Kolmogorov-Smirnov test of goodness of fit. **Statistics and Probability Letters**. v. 35, n. 3, p. 251-259, 1997.

KRÖSE, B.; VAN DER SMAGT, P. An Introduction to Neural Networks. The University of Amsterdam. nov. 1996. Disponível em:
< http://citeseer.ist.psu.edu/ose96introduction.html>. Acesso em: 11 jul. 2013.

LAM, K.; LAU, R. S. M. A simulation approach to restructuring call centers. **Business Process Management Journal**, v. 10, n. 4, p. 481-494, 2004.

LAW, A. M.; KELTON, W. D. **Simulation Modeling & Analysis**. 2. ed. New York: McGraw Hill, 1991.

LAW, A. M. How to build valid and credible simulation models. **Proceedings of the 38th Winter Simulation Conference**, v. 38, p. 58-66, 2006.

LAW, A. M.; KELTON, W. D. **Simulation Modeling and Analisys**. New York: Mc Graw-Hill, 1982.

_____. _____. 3 ed. New York: McGraw-Hill, 2000. 760 p.

LEE Y. H.; KIM S. H. Optimal production-distribution planning in supply chain management using hybrid simulation-analytic approach. In: **Proceedings of the 2000 Winter Simulation Conference**. Ansan, South Korea, p. 1252-1259, 2000.

LEAL, F.; SILVA, F. C; MONTEVECHO, J. A.; MARTINS, F. A. A practical guide for operational validation of discrete simulation models. **Pesquisa Operacional**, Rio de Janeiro, v. 31, n. 1, p. 57-77, 2011.

LI, J. P. Train Station Passenger Flow Study. **Proceedings of the 2000 Winter Simulation Conference**, Orlando, USA, p. 1173-1176, 2000.

LONGO, F. Design and integration of the containers inspection activities in the container terminal operations. **International Journal of Production Economics**, v. 125, n. 2, p. 272-283, 2010.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010.

MARINOV, M.; GIOVANNI, L.; BELLISAI, G.; CLEVERMANN, J.; MASTELLOU, A.; VICTÓRIA, D.; DELEVA, L.. Analysis of rail yard and terminal performances. **Journal of Transport Literature**, v. 8, n. 2, p. 178-200, 2014.

MAHFOUZ, A.; HASSAN SA; ARISHA A. Practical simulation application: Evaluation of process control parameters in Twisted-Pair Cables manufacturing system. **Simulation Modelling Practice and Theory**, 2010.

MATHEWS, P. G. **Design of experiments with MINITAB**. Wisconsin: American Society for Qualit, 2005.

MONTEVECHI, J.; PINHO A.F., LEAL, F, & MARINS, F. A. S. Application of design of experiments on the simulation of a process in an automotive industry. In: **Proceedings of the 2007 Winter Simulation Conference**, Washington, DC, USA, 2007.

MONTGOMERY, D. **Design and analysis of experiments**. 7. ed. The United States: Wiley, 2008.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros.** 2. ed. Rio de Janeiro: LTC, 2003. p. 463.

_____. **Applied Statistics and Probability for Engineers**, 3. Ed. United States of America: editora John Wiley & Sons, Inc. 2002.

MUSSONE, L.; CALVO, R. W. **An analytical approach to calculate the capacity of a railway system.** European Journal of Operational Research. v.228. p. 11-23. 2013.

NAKANO, D. Métodos de pesquisa adotados na Engenharia de Produção e gestão de operações. In: MIGUEL, P.A.C. (Coord.). **Metodologia de pesquisa em Engenharia de Produção e Gestão de Operações.** 2. ed. São Paulo: Elsevier, 2012.

NAZZAL, D.; MOLLAGHASEMI, M. & ANDERSON, D. A simulation-based evaluation of the cost of cycle time reduction in Agere Systems wafer fabrication facility – a case study. **International Journal of Production Economics**, v. 100, n. 2, p.300-313, 2006.

NEWMAN, S. T.; NEWNES, L. B.; BELL, R. A multi-level modeling system for the design of flexible machining installations. **International Journal of Production Research**. v 36, n. 9, p. 2355-2376, 1998.

NUNES, J. R. B. Simulação e Otimização de uma Linha de Manufatura em fase de projeto. 2006. 113f. **Dissertação (Mestrado em Engenharia de Produção)** - Universidade Federal de Itajubá, Itajubá, 2006.

OLIVEIRA, L. C. Uso integrado do método QFD e de técnicas estatísticas de planejamento e análise de experimentos, na etapa do projeto do produto e do processo. Belo Horizonte. 1999. 270 f. **Dissertação (Mestrado em Engenharia Produção)** – Universidade Federal de Minas Gerais, 1999.

OBAIDAR, M. S.; PAPADIMITRIOU G.I. **Applied system simulation:methodologies and applications**. Massachusetts: Springer, 2003.

PENG, D.; CHAO, L.; Zhili, L. Walking Time Modeling on Transfer Pedestrians in Subway Passages. **Journal of transportation systems engineering and information technology**. v.9. n.4. 2009, p. 103-109.

PEREIRA, I. C. Proposta de sistematização da simulação para fabricação em lotes. 2000. 138f. **Dissertação (Mestrado em Engenharia de Produção)** - Universidade Federal de Itajubá, Itajubá, 2000.

PEREIRA, M. N. Estudo Comparativo de Gráficos de Probabilidade Normal para Análise de experimentos Fatoriais Não replicados. 2010. 196f. Dissertação (**Mestrado em Matemática Aplicada e Estatística)** - Universidade Federal de Rio Grande do Norte, Natal, 2010.

PINTO, O. P. F. Simulação e otimização: Desenvolvimento de uma ferramenta de análise de decisão para suprimento de refinarias de petróleo através de uma rede de oleodutos. 2001. 131f. **Dissertação (Mestrado em Eletrônica)** - Universidade Federal de Santa Catarina, Florianópolis, 2001.

POTTER A, YANG B.; LALWANI C. A simulation study of despatch bay performance in the steel processing industry. **European Journal of Operational Research**, v. 179 n. 2, p. 567-578, 2007.

REKAB, K.; SHAIKH, M. **Statistical design of experiments with engineering applications**. Danvers: CRC Press, 2005.

SAHIN, I. Railway traffic control and train scheduling based on inter-train conflict management. **Transportation Research**, v. 33, p. 511-534, 1998.

SHANNON R. E. Introduction to the art and science of simulation**. Proceedings of the Winter Simulation Conference**, Washington DC, v.1, p. 7-14, 1998.

SCHOLZ, F. W.; STEPHENS, M. A., K-Sample, Anderson–Darling Tests. **Journal of the American Statistical Association,** v. 82, p. 918 -924, 1987.

VILLAMIL, E. G.; GARCÍA, M. H. **Introducción al proyecto de ingeniería**. Buenos Aires: Universidad Nacional de Buenos Aires, 2003.

YALÇINKAYA, O.; MIRA, B. G. Modelling and optimization of average travel time for a metro line by simulation and response surface methodology. **European Journal of Operational Research**, v. 196, n.1, p. 225-233, 2009.

WILK, M. B.; GNANADESIKAN, R. Graphical Methods for internal comparisons in multiresponse experiments. **The annals of Mathematical Statistics**, v. 35, n. 2, p. 613-631, 1964.

ZHAO, F. Large-scale transit network optimization by minimizing transfers and user cost. **Journal of Public Transportation**. v. 9. n. 2, p. 107-129, 2006.

ZHAO, F.; ZENG, X. Optimization of transit network layout and headway with a combined genetic algorithm and simulated annealing method. J**ournal of Engineering Optimization**. v. 38. n. 6, p. 701-722, 2006.

_____. Optimization of transit route network, vehicle headways and timetabelas for large-scale transit networks. **European Journal of Operational Research**, v. 186, p. 841-855, 2008.

THE METROS. The Metros Awards 2011 winners, Italy, France. mar./2011. Disponível em: <http://www.terrapinn.com/awards/the-metros/winners.stm>. Acesso em: 5 ago. 2012.

PORTAL DO GOVERNO DO ESTADO DE SÃO PAULO. Metrô de São Paulo é escolhido como o melhor das Américas. mar. 2010. Disponivel em: <http://www.saopaulo.sp.gov.br/spnoticias/lenoticia.php?id=208783>. Acesso em: 5 ago. 2012.

7. ANEXOS

7.1. Anexo A - Distribuições Estatísticas

Para traduzir os dados coletados numa linguagem conhecida pelo simulador, é necessario encontrar a distribuição de probabilidade na qual podem ser ajustados.

Os softwares de simulação, normalmente vêm acompanhados de um pacote de análise estatístico, tal o caso do que acompanha o Software ProModel e o InputAnalyzer que acompanha o Software Arena.

Estes dos programas de análise matemático, fornecem resultados semelhantes tais como estatística descritiva, distribuições de probabilidade e testes de ajuste. Os items a seguir fornecem informações importantes de algumas distribuições de probabilidade:

7.1.1. Distribuição Normal

Foi introduzida pela primera vez por Abraham De Moivre em 1753, e redescoberta por Laplace (1774) e por Gauss (1809), por tanto a distribuição normal é também conhecida como Distribuição de Gauss, de Laplace ou Laplace-Gauss. Esta distribuição constitui a base de toda inferência estatística e pode ser usada em sistemas onde ocorrem filas para representar o intervalo e o número de chegadas, tempos de processo e processos com um grande número de valores.

A função de densidade de probabilidade de uma *x* con distribuição normal é dada pela equação 29:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{1/2\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (29)$$

para todos os número reais *x*

A forma da curva normal está representada na figura 19

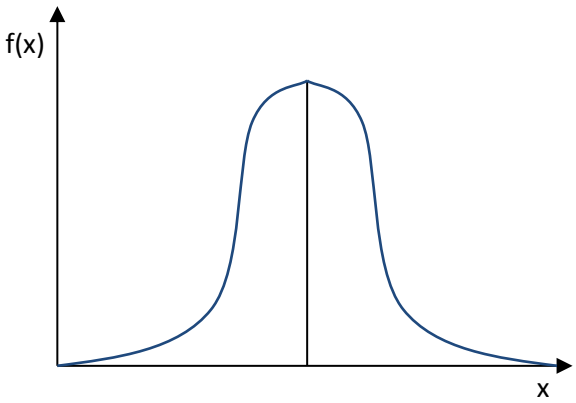


Figura 19 - Forma da curva da distribuição normal

7.1.2. Distribuição Exponencial

Distribuição continua de probabilidade que pode representar o tempo entre falhas de equipamentos, tempo entre chegadas de clientes, área entre dois defeitos consecutivos de uma peça de tecido, etc (Andrade, 2002).

A função de densidade esta representada pela equação 30

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad (30)$$

para x>=0, caso contrario 0

Desvio padrão = media = β

E a função distribuição pela equação 31

$$f(x) = 1 - e^{-x/\beta} \quad (31)$$

A forma da está representada na figura 20.

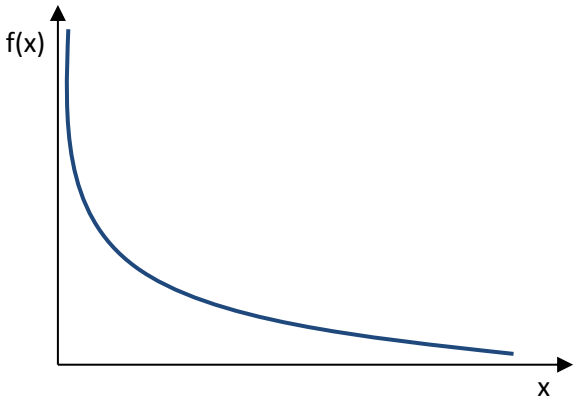


Figura 20 - Forma da curva da distribuição exponencial

7.1.3. Distribuição Weibull

Distribuição continua de probabilidade usada principalmente para definir o tempo de realizar alguma tarea ou o intervalo de tempo de falha de uma peça. Numa estação de metrô poderia representar o tempo de abertura das portas, ou o intervalo de tempo de falha dos carros.

A curva da distribuição Weibull está representada na Figura 21.

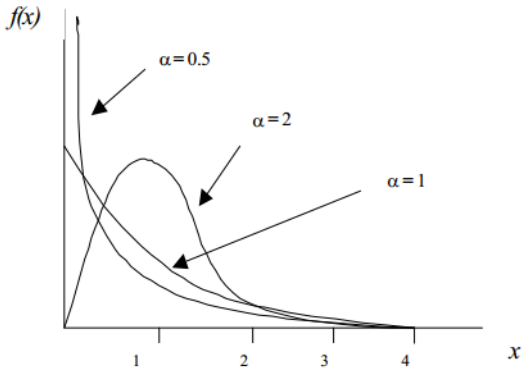


Figura 21 - Forma das curvas de distribuição de Weibull para diversos α

A função densidade de probabilidade é representada pela equação

$$f(x) = \frac{\alpha}{\beta} x^{\alpha-1} e^{-(\alpha/\beta)^{\alpha}} \quad (32)$$

para x>0, caso contrario 0

E a Função de distribuição acumulada esta representada pela equação 33

$$f(x) = 1 - e^{(x/\beta)^{\alpha}} \quad (33)$$

para x>0, caso contrario 0

7.1.4. Distribuição Lognormal

Distribuição continua com formato parecido à distribuição Weibull, os usos mais comuns são tempos de processo, ou tarefas com um grande número de dados.

A curva da distribuição Lognormal esta representada na figura 22.

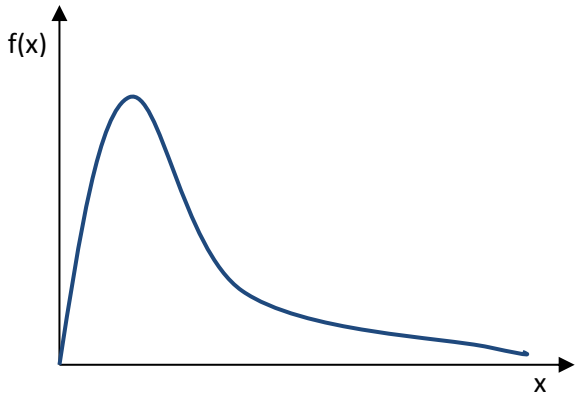


Figura 22 - Forma das curvas de distribuição de Lognormal

A função de densidade esta representada pela equação 34

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{(lnx-\mu)^2/(2\sigma^2)} \quad (34)$$

para x>0, caso contrario 0

8. APÊNDICES

APÊNDICE A

Acta Scientiarum

http://www.uem.br/acta
ISSN printed: 1806-2563
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v35i4.17468

Improvement of the supply chain for the sugar cane exportation process employing discrete events simulation techniques

Maria Alejandra Guerrero Hernandez* **and André Felipe Henriques Librantz**

*Programa de Pos-graduação em Engenharia de Produção, Universidade Nove de Julho, Av. Francisco Matarazzo, 612, São Paulo, São Paulo, Brazil. *Author for correspondence. E-mail: malejandragh@gmail.com*

ABSTRACT. Current paper proposes a reduction on all logistic costs involved in the sugar cane exportation process for an important medium-size cargo harbor in South America. The strategy consisted in studying and improving the efficiency of the logistics transportation flow of merchandise between the sugar cane mills and the cargo ports by employing discrete events simulation as a methodology and the ProModel software as a simulation platform. The simulated scenarios show 89% reductions in additional costs for inventory management; 50% of the resources used for loading the containers; and an 11.4% increment in the efficiency of operations. In addition, the new planning strategy was also evaluated with either a positive or negative variation of the demand. It should be underscored that the improved scenarios (out of those simulated) do not involve investments in facilities or in machinery purchasing.

Keywords: simulation, sea port, logistics, transport.

Improvement of the supply chain for the sugar cane exportation process employing discrete events simulation techniques

RESUMO. O objetivo desse trabalho é encontrar uma redução de todos os custos logísticos envolvidos no processo de exportação do açúcar de cana de um importante porto de carga de porte médio na América do Sul. A estratégia consistiu em estudar e melhorar a eficiência do fluxo logístico de transporte de mercadorias entre os engenhos de cana de açúcar e os portos de carga, utilizando como metodologia a simulação de eventos discretos e o software ProModel como plataforma de simulação. Os cenários simulados, levam-nos a obter reduções de 89% nos custos na gestão de estoques e 50% dos recursos utilizados para carregar os contêineres, também permitiu obter um incremento de 11,4% na eficiência das operações. Além disso, foi possível avaliar uma nova estratégia de planejamento com variação positiva ou negativa da demanda. É importante destacar que os cenários melhorados (daqueles simulados) não envolvem investimento em instalações nem na compra de equipamentos.

Palavras-chave: simulação, porto marítimo, logística, transporte.

Introduction

In today's highly competitive companies it is imperative to face the challenge of reducing loading duration, stocks and costs to improve the process in a global environment. Depending on the complexity of the process and the type of solution adopted, each company has its own decision-making process. Discrete event simulation has become an important method and provides companies with several advantages like personnel training, better visualization and understanding of the industrial productive processes.

Banks et al. (2005) state that a simulation model is a mathematical model studied by using the simulation itself, this lets you design the system in a computational model in order to perform tests and experiments besides to understand the behavior of the system, applying different strategies of operation.

Law and Kelton (2000) complement the previous description stating that discrete event simulation consists of creating a model that evolves over time where the simulation variables change at specific points in time. In most systems, the no implementation of simulation experiments may be economically unviable or simply impossible to be executed.

Current article provides an alternative for researching. In fact, it involves the simulation of scenarios conducting to the planning of logistics operations for the sugar exportation process in a port with great strategic importance in Colombia. The research is delimited to the supply chain of container exportation and focuses on finding an alternative that reduces loading time, improves delivery duration and conducts to the reduction of stocks.

The planning of brand sugar production by the mills and ship assignments is common to most large sugar exporting countries, though the specific complexities within the supply chain are different in different countries (HIGGINS et al., 2006). The literature abounds with papers that investigate logistical activities at the port and shipping. Models for scheduling activities at the port primarily focus on the handling of containers from trucks to ships (GAMBARDELLA et al, 2001; HARTMANN, 2004; KOZAN, 2000). Models for shipping berth planning address the schedule of activities that a ship undertakes once in port (BROWN et al., 1994), berthing positions within large ports (PART; KIM, 2002) and chain integration with sugar production employing heuristic methods to reduce costs of sugar production and obtain a better management of logistics for ship berthing (HIGGINS et al., 2004, 2006; SELLITO et al., 2009).

Production of sugar cane in Colombia

The sugar cane industry in Colombia is concentrated in the southwestern region of the country called the Cauca Valley. This geographical area shares its characteristics only with two other places, namely, Hawaii and the plains of northern Peru, characterized by the highly relevant differences related to thermal amplitude between day and night, essential conditions to get the best concentration of sucrose in sugar cane plants.

This privileged condition guarantees sugar production throughout the year, which doesn't occur in the rest of the world, where the harvest of sugar cane only lasts 4 to 6 months on an average. Consequently, fixed costs and the cost of capital are about half and up to a third of those for other sugar cane production areas in the world. Although production costs are lower, the logistic ones are very high, affecting seriously the revenues of the companies and their competitiveness on the international markets.

During the analyzed period (one month), 93,750 tons of sugar were exported: 33.4% in containers, 26.57% in bulk and the remaining 33.4% in sacks. To load the referred 33.4%, (31,350 tons of sugar), 1,183 containers were used.

Lack of planning for the dispatching of the product from the sugar mills to the maritime port generated a cost overrun of US$ 8,190 month^{-1} and a resource utilization of 30.81%.

Under the same circumstances of demand, current study demonstrated that it was possible to reduce fines up to US$ 822 month^{-1} and increase the resource utilization to 60.85%.

The Maritime Port

Cargo sea ports are fundamental in Colombia's business development since more than 90% of exports currently occur by sea port facilities.

Colombia has nine important sea ports, comprising seven on the Caribbean coast: Guajira, Santa Marta, Cienaga, Barranquilla, Cartagena, Golfo de Morroquillo, Urabá and St. Andrew; and two on the Pacific coast: Buenaventura and Tumaco.

Cartagena, Barranquilla and Santa Marta are the main cargo ports on the Caribbean's coast. The port of Buenaventura on the Pacific coast has been established as the country's main multipurpose port and is actually the epicenter of most imports and exports of non-traditional products.

In 2010, 9.02 million tons of different products were transported from the port of Buenaventura. Table 1 shows that 1,178 ships were docked.

Table 1. Tons transported through the port of Buenaventura in 2010.

Type of Cargo	Exported Tons	Imported Tons	Total Tons
General Cargo	60,212	686,422	746,633
Solid bulk	53,659	2,855,151	2,908,810
Liquid bulk	-	324,870	324,870
Coal in bulk	619,871	-	619,871
Container 20L1	638,774	1,661,165	2,299,939
Container 40L1	488,005	1,631,609	2,119,614
Total	1,860,521	7,159,217	9,019,737

The operations of the maritime port of Buenaventura may be comparable to some small ports in Brazil such as Tup Usiminas, São Paulo State, Porto de Salvador, Bahia State, Porto de Belém, Pará State or Porto de Fortaleza, Ceará State and others, which exported an average of 3 millions of tons in 2009. However, these ports transport the amount of tons corresponding to the 9.7% of the total tonnage transported through Brazilian ports, as shown in Table 2.

Table 2. Tons transported through the Brazilian ports in 2009.

Millions of Tons	Quantity of Ports	% Total of Ports	Total Tons	% Total Tons
75-90	3	6	247,217,674	33.7
30-50	6	12	232,632,344	31.7
10-20	8	16	116,242,879	15.9
5-10	10	20	71,094,321	9.7
0-5	24	47	65,743,923	9.0
Total	51	100%	732,931,141	100

Restrictions and definition of the problem

The logistic process includes the following activities: 1. Products loading at the 12 sugar mills of the region. 2. Freight (via roads). 3. Cargo admission to the port of Buenaventura. 4. Quality control. 5. Weight-in. 6. Unload. 7. Storage. 8. Loading the ships (containers, sacks and bulk).

The above system is fraught with multiple restrictions of a social, political and environmental nature, which together affect its performance and involve high logistic costs due to excess of inventory, poor use of resources and delivery failures. These restrictions may be summarized as follows:

- The travel time of trucks from the sugar mills to the port of Buenaventura may take between 5 and 24 hours. Variations are mainly due to transit restrictions related to cargo type, hours, vehicles, highly curvy roads and behavioral factors, with difficulties to have a good control on truck drivers. The latter factor will not be dealt with in current study and only travel time will be analyzed as a system constraint. An average speed of 24 km h^{-1} was considered for simulation purpose.

- The sugar is exported mainly in three forms: bulk, sacks and containers. In the case of sugar in bulk, approximately 40% of the total time available is lost since the operation is interrupted by rain or foul weather. The port of Buenaventura has a relative humidity of 51%.

- At the period of current study, the loading operation of containers for the United States and Canada had to be undertaken in a protected area with cameras and in the presence of anti-narcotics police officers. They only had three available anti-narcotics agents who could only monitor up to two containers simultaneously. Nowadays the conditions are different since the company has a warrant allowing it to operate by itself.

- The company owns three warehouses at the port of Buenaventura. One is only used for sugar in bulk and has a storage capacity of 40,000 tons; the other two warehouses with a capacity of 20,000 tons each are areas of free storage where sacks can be stored.

- Contracts between the company under study and the shipping companies on failure in the estimated load times feature rewards and punishments paid by both parties in the following scenarios: when the loading is done in less time than estimated, the shipping company pays the company under analysis a sort of 'reward'; on the other hand, when the product is not ready for loading, the company pays the shipping company a sort of 'penalty'.

- Lack of planning in the transport of the product from the sugar mills. In many cases since the sugar is dispatched from the sugar mills several days before ETA (Estimated Time of Arrival), there is an increase in storage days and sometimes the company has to rent an additional warehouse. Although the product is available for the immediate loading of the ships and thus assets are obtained, in the long run excessive storage costs do not compensate the losses.

- Low utilization of the cargo's resources of the producing companies. Capacity allows the loading of 16 trucks day^{-1} with 32 tons each; however, in the current scenario, only between 8 and 10 trucks with 32 tons each loaded per day.

The need for companies to compete within a faster, globalized and demanding market and at the same time to comply with the principles of quality, service and delivery time at minimum costs requires the development of a new management model to reduce the aggregated costs and improve the response time of supply chain.

Material and methods

The simulation model was developed with ProModel V 4.0 on Windows XP Operating System.

ProModel software has high flexibility and ability to interact with the system. It is one of the most advanced softwares for discrete event simulation to evaluate, plan and design production, storage, logistics and other operational and strategic situations, and thus generate considerable money savings to improve the planning and control systems.

The technique of Discrete Event Simulation makes possible the development of virtual scenarios which emulates the behavior of virtually any type of process.

The model includes the logistical process from 12 sugar mills to the maritime port of Buenaventura, Colombia, taking into consideration each of the system´s segments and constraints.

The simulation model and the scenarios for improvement are analyzed from two perspectives:

1. Loading operations in containers.
2. Sacks and bulk.

It is important to underscore that current paper only analyzes the operations of loading containers.

Operations of loading containers

The input data for simulation software refer to a month considered representative and characteristic. The simulation is based on two routes: northern and southern.

Southern route: for this particular month, 11 cargo ships, distributed along the month, were loaded with 696 containers.

Northern route: 15 cargo ships were loaded, also distributed along the month, with 487 containers.

Analyzed indicators

- Value of awards and penalties.
- % Compliance Cut Off (Cut Off: the date and time in which the entire load must be physically

within the port, ready, properly documented and approved for shipment).

- % Northern Route penalized containers.
- % Southern Route penalized containers.
- Maximum inventory of pallets.
- Average inventory of pallets.
- Maximum number of loaded containers per day.
- Average quantity of loaded containers per day.
- % Utilization of resources per day.

Input data and assumptions

All time intervals in each stage of the supply chain, namely, travel time, weighing time, quality control, inspection and loading of containers, were analyzed separately for the northern and southern routes.

- The Information and Statistical Analyses were performed by StatFit, a tool provided by ProModel. Information was collected from the physical documents where the company keeps the control records of its operations.

- Penalties for non-compliance are calculated by the staying time of the ships after the fifth day of berth at the port; moreover, containers must be completely loaded two days before ETA.

- Current cargo capacity has an average of 56 containers per day for the southern route and 48 for the northern one.

- Capacity of container: 12 pallets per container.

- Number of teams of operators for loading containers: Northern route: 6 teams working from 8:00 to 18:00; Southern Route: 9 teams, one of which works from 8:00 to 18:00 and the other eight from 18:00 to 8:00.

- Each loading team is composed of six people who carry the sacks to the pallets; the pallets are then lifted and loaded onto the containers.

Results and discussion

Since the process started when the trucks were loaded at the mills, the scenarios were designed by defining the optimal time to start loading the truck so that the cargo was ready at the port of Buenaventura by the arrival time of the ships.

The definition of the scenarios was based on three alternatives: 1. Improved scenarios defined by the company. 2. Internal Benchmarking. 3. Trial and error.

Twelve main scenarios were simulated by varying the date of dispatch from the mills, as follow:

1 - Dispatch 7 days before the northern route and ETA to 5 days before ETA to south route.

2 - Dispatch 6 days before ETA for the north and south routes.

3 - Dispatch 5 days before ETA for the north and south routes.

4 - Dispatch 4 days before ETA for the north and south routes.

5 - Dispatch 3 days before ETA for the north and south routes.

6 - Dispatch 2 days before ETA for the north and south routes.

7 - Date of dispatch for each route calculated by the formulas: Northern route: Cut Off - (# Containers/48) - 1; Southern route: Cut Off - (# Containers / 56) – 1.

8 - Dispatch 7 days before ETA for the north and south routes.

9 - Dispatch 7 days before ETA for both northern and southern routes, maximum load capacity and dispatch (16 trucks day^{-1}).

10 - Dispatch 6 days before ETA for both northern and southern routes; maximum load capacity and dispatch (16 trucks day^{-1}).

11 - Dispatch 7 days before ETA for the northern route and 5 days before ETA for the southern route; maximum load capacity at the sugar productions facilities (16 trucks day^{-1}).

12 - Dispatching date for each route calculated by the equations: Northern route: Cut Off – (Quantity of containers (#) /48) – 1; Southern route: Cut Off – (Quantity of containers (#) /56) – 1, with a maximum loading capacity and dispatch (16 trucks day^{-1}).

Within the context of these scenarios, simulation demonstrated that Scenario 12 would provide the best performance for the defined indicators, reduce the additional non-budgeted costs for storage and decrease penalties. In this scenario, the operations were scheduled and took into account that the starting date of the loading process must be established some days before the Cut Off and the efficiency in loading containers as well, which was 56 containers for the southern and 48 for the northern route.

Starting of the loading process (Northern route) = (Cut Off – (# Containers/48) – 1) days [1]

Starting of the loading process (Southern route) = (Cut Off – (#Containers /56) – 1) days [2]

With this improvement, the cost overrun for storage was reduced by 88.89% (penalties) and a 99.75% fulfillment was obtained for the cut off dates. Within a total of 1,183 containers loaded during the month, the number of penalized containers had a reduction of 85.4%; that is, it went down from 30 to 4%, with costs decrease

US$ 7,280 month^{-1} and US$ 87,360 year^{-1}, under the current conditions of demand. Figure 1A and B show the monthly results that could be obtained with the new scheduling.

Improvement allowed increase of efficiency of the loading process for the containers, as Figures 2 and 3 show.

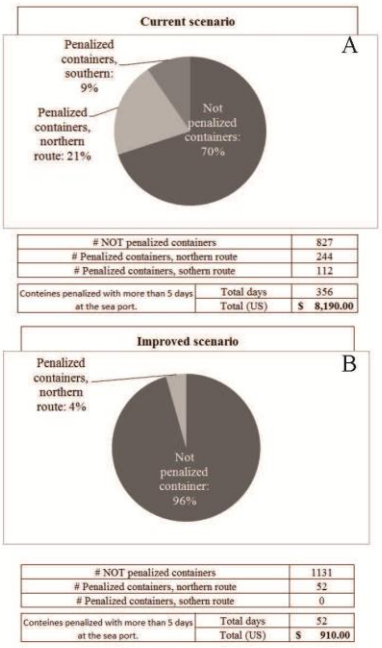


Figure 1. (A) Storage costs of the current situation, (B) storage costs of the improved scenario.

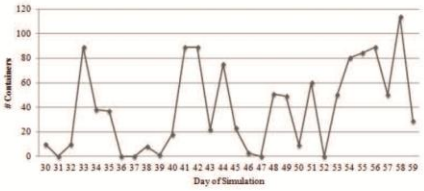


Figure 2. Number of containers loaded per day (current scenario).

Whereas in current situation, the number of loaded containers per day was 39, the average number in the improved scenario increased to 44, a 4.9% improvement in the efficiency of the loading operations. A decrease in penalties and a better utilization of resources might also be observed.

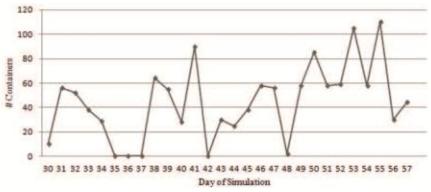


Figure 3. Number of containers loaded per day (improved scenario).

Planning of the dispatches with the current demand

The scheduling of dispatches at the sugar mills was done by formulas [1] and [2] for each case.

The working shifts will not change and they will remain the same as follows:

North: Monday thru Friday, 6:00 to 20:00; Saturday, 6:00 to 12:00.

South: Monday thru Friday, 2 shifts of 12 hours, Saturdays: 6:00 to 12:00.

In order to appraise the efficiency of the resource's utilization, the behavior of the two different routes was analyzed (Northern route: 6 morning shifts, Southern Route: 8 night shifts, 1 morning shift) by numerical simulation.

1 - Four teams for loading the containers for each shift (day and night); it is mandatory that the loading of the northern route is done during the day (this is a system's restriction)

2 - Four teams for loading the containers during the morning shift.

Table 3 shows the results of these simulated scenarios.

Table 3. Reduction of resources with current demand.

Scenario	Penalties	% Penalized containers	Average of Containers loaded per day	% Resources Utilization
Improved scenario Current Resources	$ 910,00	4.4	44	34.23
Scenario 1: 4 Teams Shift^{-1}	$ 822,00	4.0	44	60.85
Scenario 2: 4 Teams Shift^{-1} Daily^{-1}	$ 1,820.00	7.0	44	34.23

Simulation showed that scenario 1: 4 teams shift^{-1} could reduce even more the penalties: US$ 822 month^{-1} representing additional savings per year, besides those by the first improvement of US$ 1,000 year^{-1}. The better use of the resources triggered a 167% increase in productivity and a reduction of downtime from 65.77 to 34.15%. Since salaries and compensations are not part of current analysis, it is not possible to calculate their cost reduction. However, a working hypothesis using the

minimum salary would lead to savings of US$ 140,000 year^{-1}. Total savings may reach US$ 228,360 per year when they are coupled with penalties reduction.

Planning of dispatches to reduce demand by 40%

Two main scenarios were simulated for this purpose:

1 - Current teams (6 for the northern route; 9 for the southern route), with 40% reduction in demand.

2 - Four teams for loading the containers at each shift (day, night); the packaging is a priority for the Northern Route, to be done during daytime and with a reduction of 40% in demand.

Table 4 shows results obtained in the simulation of the proposed scenarios:

Table 4. Reduction of demand in 40%.

Scenario	Penalties	% Penalized containers	Average of Containers loaded per day	% Resources Utilization
Scenario 1: Current teams	$ 0	0.0%	26	20.69%
Scenario 2: 4 Teams Shift^{-1}	$ 0	0.0%	26	36.78%

The simulated scenarios show that the new planning model is also applicable if it has a demand reduction since the latter doesn't generate penalties. It also demonstrates that the number of teams may be reduced to even less than 4 teams shift^{-1}.

Planning of dispatches with a 20% increase of demand

Two main scenarios were simulated for this purpose:

1 - Current teams (6 for the northern route; 9 for the southern route) increase demand by 20%.

2 - Four teams for loading the containers for each shift, a 20% increase in demand.

Table 5 lists the results obtained in the simulation of the proposed scenarios:

Table 5. Increase of demand by 20%.

Scenario	Penalties	% Penalized containers	Average of Containers loaded per day	% Resources Utilization
Scenario 1: Current teams	$ 1,137.5	5.49	53	41.23
Scenario 2: 4 teams shift^{-1}	$ 227.5	1.0	51	70.68

The new operation plan generates good results even with a greater demand. However, to achieve the projected objective, or rather, better alternatives having low costs, Scenario 2: 4 teams shift^{-1} not only reduces the penalties to US$ 227.50 month^{-1} but also increases the utilization of resources with 70.68%.

Conclusion

The golden goal was to redefine the strategy of the entire sugar supply chain, minimizing costs but aligned with the operational restrictions. Simulations provided crucial information to the high management for decision making, appraising new alternatives and understanding the impact of certain changes. New planning policies reduced penalized containers from 356 to 52; a reduction of 85.4%, with costs reduction of US$ 7,280 month^{-1} or US$ 87,360 year$_{-1}$. Simulated scenarios with changes in the loading working crews demonstrated that half of the resources with increased utilization (167%) resulted in US$ 140,000 year^{-1}; global saving of US$ 228,360 year^{-1}.

Acknowledgements

The authors would like to thank the Nove de Julho University of São Paulo for the financial support.

References

BANKS, J.; CARSON J. S.; NELSON, B. L.; NICOL, D. M. **Discrete-event Simulation System Simulation**. 4th ed. Upper Saddle River: Prentice-Hall, 2005.

BROWN, G. G.; LAWPHONSPANICH, S.; PODOLAK-THURMAN, K. Optimizing ship berthing. **Naval Research Logistics**, v. 41, n. 1, p. 1-15, 1994.

GAMBARDELLA, L. M.; MASTROLILLI, M.; RIZZOLI, A. E.; ZAFFALON, M. An optimization methodology for intermodal terminal management. **Journal of Intelligent Manufacturing**, v. 12, n. 5-6, p. 521-534.2001.

HARTMANN, S. Generating scenarios for simulation and optimization of container terminal logistics. **OR Spectrum**, v. 26, n. 2, p. 171-192, 2004.

HIGGINS, A.; BEASHEL, G.; HARRISON, A. Scheduling of brand production and shipping within a sugar supply chain. **Journal of the Operational Research Society**, v. 57, p. 490-498, 2006.

HIGGINS, A.; ANTONY, G.; SANDELL, G.; DAVIES, I.; PRESTWIDGE, D.; ANDREW, B. A framework for integrating and optimizing a complex harvesting and transport system for sugar production. **Agricultural Systems**, v. 82, n. 2, p. 99-115, 2004.

KOZAN, E. Optimizing container transfers at multimodal terminal. **Mathematical and Computer Modelling**, v. 31, n. 10, p. 235-243, 2000.

LAW, A. M.; KELTON, D. W. **Simulation modeling and analysis**. 3rd ed. New York: McGraw-Hill, 2000.

PART, K. T.; KIM, K. H. Berth scheduling for container terminals by using a sub-gradient optimization technique.

Journal of the Operational Research Society, v. 53, n. 9, p. 1054-1062, 2002.

SELLITO, M. A.; BORCHARDT, M.; PEREIRA, G. M. Analysis of a load operation logistics and shipment of cement by computer simulation. **Journal of Industrial Management**, v. 5, n. 4, p. 130-151, 2009.

Received on June 4, 2012.
Accepted on January 14, 2013.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

APÊNDICE B

Ingeniare. Revista chilena de ingeniería

Simulación de eventos discretos de la cadena logística de exportación de commodities

Discrete event simulation of the commodities' supply chain exportation

María Alejandra Guerrero Hernández André Felipe Henriques Librantz [1]

RESUMEN

El presente trabajo utiliza la técnica de simulación de eventos discretos como metodología de apoyo a la decisión en un problema complejo con múltiples variables; el objetivo es modelar las operaciones logísticas involucradas en la exportación marítima de commodities en las presentaciones de carga suelta y granel desde doce empresas productoras hasta el puerto marítimo. La simulación permite analizar la sensibilidad de los principales parámetros y variables del sistema, con el fin de definir diferentes modelos de planeación de despachos. Los escenarios muestran que es posible obtener una reducción de los inventarios y de los costos de mantenimiento, así como también mejorar los indicadores de servicio y como consecuencia aumentar la tasa de utilización de los recursos.

Palabras clave: Simulación de Eventos Discretos, Puertos Marítimos, ProModel, exportación marítima.

ABSTRACT

This paper uses the technique of discrete event simulation as methodology for decision support in a complex problem with multiple variables; the objective is to model the logistics operations involved in the maritime exportation of commodities in the presentations of bulk and bags since twelve producers until the seaport. The simulation allows analyzing the sensitivity of the key parameters and system variables in order to define different models of planning offices. Results showed that it is possible to reduce the inventory and maintenance cost, as well as improving the service performance indicators and as consequence increasing the resources utilization rate.

Keywords: Discrete Events Simulation, Seaport, ProModel, maritime exportation.

[1]Universidad Nove de Julho (UNINOVE), Avenida Francisco Matarazzo, 612, Código Postal: 05001-001, Sao Paulo-Brasil, malejandragh@gmail.com; librantz@uninove.br

INTRODUCCIÓN

El modelado de sistemas reales cuyo desempeño depende en gran medida de parámetros aleatorios, como la demanda o el tiempo de servicio, es hoy en día un amplio campo de investigación, tanto para los desarrollos teóricos como para las aplicaciones. Un Modelo de Simulación es un modelo descriptivo de un proceso o sistema, que usualmente incluye parámetros para representar diferentes configuraciones del sistema o proceso.

El uso de modelos de simulación puede reemplazar la realización de experimentos en sistemas reales y aquellos proyectos que aún se encuentran en fase de desarrollo [1], permitiendo experimentar, evaluar y comparar muchos sistemas alternativos [2].

Para modelar y simular un sistema real es necesario implementar un modelo matemático aplicado a la simulación, pueden ser utilizados lenguajes de programación como C++, Visual Basic, Pascal; o programas de simulación como por ejemplo, ProModel, Arena, Extend y SLAM. [3]

ProModel, es un software de simulación de eventos discretos altamente flexible, que permite la interacción con otras herramientas computacionales como Excel, y su integración con Excel lo vuelve altamente flexible y amigable. [4]

Los escenarios o estrategias definidos usando la simulación, permiten obtener un análisis de sensibilidad de las diferentes variables, al alterar todos o algunos de los parámetros de entrada al sistema, permite obtener medidas de desempeño que reflejen la calidad del servicio prestado y la sensibilidad de los indicadores ante las variaciones del sistema. La técnica de elaboración de escenarios es un tema que ha sido tratado por varios autores,

La simulación, es una metodología que se ha popularizado en los últimos años, diversos procesos logísticos, productivos y de servicios han sido mejorados con la ayuda de la simulación y el modelaje matemático [5-10].

METODOLOGÍA

El sistema modelado corresponde a los procesos desarrollados por un operador logístico, que tiene como objetivo planear, programar y realizar las operaciones logísticas de exportación de commodities en las presentaciones de carga suelta y granel, por vía marítima, para 12 empresas del sector. La carga suelta, consiste en bultos sueltos o individuales, manipulados y embarcados como unidades separadas, paquetes, sacos, cajas, tambores, piezas y el producto a granel es una carga constituida por sólidos que vienen sin empaque o envase y no pierden su condición en ninguna de las distintas fases de la operación portuaria.

El operador logístico realiza la programación de los pedidos y coordina junto con las 12 empresas proveedoras la salida de los camiones desde sus instalaciones hasta el puerto marítimo, así mismo se encarga de todos los procesos administrativos y operativos desarrollados dentro del puerto marítimo.

Este operador, cuenta con instalaciones alquiladas dentro del puerto marítimo, 3 bodegas, la primera de granel, con capacidad para almacenar 40.000 toneladas, y la segunda y tercera para carga suelta, con capacidad de almacenar 7.140 estibas, conjuntamente. La operación logística tanto para la exportación de carga suelta y granel incluye los siguientes procesos:

1. Programación de los despachos desde las empresas proveedoras. 2. Embarque del producto en camiones.3. Transporte por vía terrestre. 4. Entrada de los camiones al puerto marítimo y realización de las documentaciones necesarias. 5. Control de calidad. 6. Almacenaje. 7. Carga del producto en la motonave como carga suelta, granel y contenedores. (Es importante resaltar que en el presente trabajo solo serán abordadas las operaciones de carga suelta y granel). El cargue de los sacos y el granel en

las motonaves se realiza separadamente y con equipo especializado diferente.

Uno de los principales problemas para este operador logístico es la planeación de las operaciones y la sincronización del envío de los camiones desde las 12 empresas proveedoras. La falta de sincronización ha generado operaciones poco eficientes que se reducen en dos problemas principales y opuestos:

- Salida de los camiones de las empresas proveedoras con muchos días de anticipación antes del atraque de la motonave que irán a atender, lo que desemboca en altos niveles de inventario dentro de las bodegas alquiladas dentro del puerto marítimo.
- Falta de carga en puerto cuando se encuentra la motonave atracada y los camiones aún no han llegado con el producto. Estos dos problemas son usuales, y generan altos costos para la empresa, para el primer caso, el costo asociado corresponde al mantenimiento del inventario, y el segundo problema acarrea penalizaciones, el operador logístico debe pagar a la naviera una multa por cada día de demora en el proceso de carga del producto.

Por su parte, la eficiencia en las operaciones es premiada con el pago de compensaciones, el operador logístico recibe estos premios con base en la reducción de días de cargue con relación a lo planeado.

PROCEDIMIENTOS

Dada la naturaleza no determinística que presenta el proceso sujeto a análisis, además de su estructura dinámica y compleja, se hace recomendable el modelaje de la situación problema a través de herramientas de simulación de eventos discretos. Este esquema permite representar la naturaleza estocástica del sistema y simplificar la representación de todas las variables y elementos que a él pertenecen. Para esto, fue seleccionado el software de simulación de eventos discretos ProModel, por su alta flexibilidad para el modelaje de procesos.

Como estrategia, el análisis de las operaciones, se realizó con base en la información correspondiente a un mes característico de la operación, mes en el cual fueron atendidas 5 motonaves con un total de 62.400 toneladas de producto, considerada una muestra adecuada, de acuerdo a la información estudiada, la cantidad de motonaves atendidas por mes es estable, por lo tanto, esta muestra garantiza que cualquier estacionalidad en el proceso estará incluida.

Para las variables de entrada del modelo de simulación, la representatividad de los resultados se comprobó a través de pruebas de bondad de ajuste, Chi-Cuadrado, usando el software Stat Fit, paquete de análisis estadísticos que acompaña ProModel.
Para todos los casos, las pruebas fueron realizadas considerando una confianza de 95%, α=0,05.
El ajuste fue comprobado con el factor P-value. La hipótesis nula, considera que los datos se ajustan a las distribuciones calculadas.
Así, la hipótesis nula es aceptada si P-value>α, de lo contrario se rechaza y se asume que los datos no se ajustan a la distribución calculada.
Los indicadores representados como variables discretas, fueron usados de esta forma debido a la falta de datos históricos para su análisis estadístico.
A continuación se presentan las variables, sus distribuciones y el resultado del cálculo de p-value.

- Tiempo de Transporte de los camiones desde cada una de las empresas proveedoras resultado de las variables: distancia (desde los proveedores hasta el puerto) y velocidad promedio de los camiones (40km/h).
- Tiempo de Movimiento de las estibas con carga suelto hasta el equipo de carga de la motonave 1,57 minutos/estiba.

- Tiempo de Movimiento de la carga a granel hasta las bandas transportadoras que desembocan en la motonave: 2,36 minutos para cada 34 toneladas.
- Tiempo de Análisis de laboratorio: ~ N(7,98;1,97) minutos. P-value = 0.12.
- Tiempo de Pesaje en el puerto ~N(2;1) minutos. P-value = 0,062.
- Tiempo Segundo Pesaje en el Puerto ~N(2;1) minutos. P-value = 0,16.
- Tiempo de descargue de producto para camiones de granel ~N(19,51;6,81) minutos, P-value = 0,21.
- Tiempo de descargue de producto para camiones de carga suelta ~N(52,721; 9,030) minutos. P-value = 0,089.

Situación Actual

La información de toneladas movidas y de premios ganados para el mes analizado es presentada en la Tabla 1.

A partir de esta información, se buscó mejorar los resultados con el establecimiento de nuevos posibles escenarios y la definición de parámetros que permitan dar inicio a la simulación. Los indicadores fueron analizados de manera independiente para las operaciones de carga suelta y granel.

Indicadores Analizados

- Fila media en la entrada al puerto Marítimo (F)
- Tiempo promedio de un camión en el sistema – Desde que sale de la empresa proveedora hasta que sale del puerto – (TS).
- Premios ganados por reducción de días de embarque para carga suelta (Premios_CS).
- Premios ganados por reducción de días de embarque para grane (Premios_GR).
- Media de camiones descargados por día para carga suelta (Prom_Cam/día CS).
- # Máximo de camiones Descargados por día para carga suelta (Max_Cam/día CS).
- Inventario Promedio de carga suelta – Estibas– (Prom_Inv_CS).
- Máximo Inventario de carga suelta – Estibas– (Max_Inv_CS).
- Media de camiones descargados por día: GRANEL (Prom_Cam/día GR)
- # Máximo de camiones Descargados por día para Granel (Max_Cam/día GR).
- Inventario Promedio de granel – Toneladas– (Prom_Inv_GR)
- Máximo Inventario de carga suelta – Toneladas– (Max_Inv_GR).

Tabla 1. Motonaves atendidas el mes de noviembre.

Toneladas por Buque Atendido	Días antes del ETA para iniciar la salida de los camiones desde las empresas proveedoras	Días empleados para cargar el buque	Premios Reales (USD)	
15,50	7,40	9,80	$23.575,00	CARGA SUELTA
3,50	6,90	6,10	$1.012,00	
14,00	7,50	10,30	$12.845,00	
4,40	7,60	2,80	$13.255,00	
25,00	6,30	1,50	$23.750,00	GRANEL
	TOTAL PREMIOS $74.437,00 USD			

RESULTADOS

Estrategia de Definición de Escenarios

En el marco de la discusión sobre la disminución de costos y el incremento de beneficios en las operaciones portuarias de exportación de commodities, se definió que la variable más relevante y que genera un mayor impacto en las operaciones es el número de días antes de la llegada del buque, variable que llamaremos ETA (Estimated time arrival), variable que define el número de días en que los camiones deberían iniciar su salida desde las empresas proveedoras

después de la fecha estimada de llegada de una motonave.

La planeación de las motonaves no puede realizarse de manera aislada y las estrategias seleccionadas deben apuntar a que el producto fluya, aún con recursos limitados y compartidos, por tanto, los escenarios simulados, deben considerar, ajustar y encontrar el mejor balance que produzca el mayor beneficio al estudiar estas dos alternativas:

1. En caso de que los camiones salgan desde los proveedores muchos días antes del ETA de la motonave, el nivel de inventarios acumulado el día de la llegada efectiva del barco es muy alto, causando sobrecostos por inventarios excesivos. Sin embargo, el haber mucho inventario en las bodegas genera la posibilidad de que los problemas por "falta de carga" sean mínimos, incrementando los beneficios por premios.

2. Si los camiones salen desde las empresas proveedoras con la mínima cantidad de días antes del ETA, acumularía menos inventarios permitiendo una disminución de los costos de mantenimiento de inventarios, lo cual aumentaría la posibilidad de que exista falta de carga en el momento de la llegada del barco, perdiéndose la posibilidad de ganar mejores premios o teniendo incluso penalidades.

Escenarios Simulados
Situación Actual - Carga Suelta: 7 días antes del ETA; Granel: 15 días antes del ETA.
Escenario 1 - Carga Suelta: 6 días antes del ETA; Granel: 14 días antes del ETA.
Escenario 2 - Carga Suelta: 5 días antes del ETA; Granel: 13 días antes del ETA.
Escenario 3 - Carga Suelta: para buques entre 10.000 y 15.000 toneladas, 5 días antes del ETA, para Buques entre 5.000 y 10.000 toneladas 8 días antes del ETA, para Buques < 5000 toneladas, 10 días antes del ETA.
Granel: para buques > 10.000 toneladas, 40% de la carga en puerto antes del ETA, para Buques entre 5.000 y 10.000 toneladas, 50% de la carga en puerto antes del ETA, para Buques < 5.000 toneladas, 50% de la carga en puerto antes del ETA.

La Tabla 2 presenta los resultados de los escenarios simulados, permitiendo una comparación de cada una de las variables definidas frente a la modificación del parámetro de entrada Días Antes del ETA.

Carga Suelta, al analizar estas operaciones, en el escenario actual se observa que el inventario promedio fue 28% de la capacidad de las bodegas, sin embargo, se presenta un pico máximo de inventarios que en cierto momento sobrepasa la capacidad instalada de almacenamiento dentro de las bodegas del puerto (7.140 estibas), requiriéndose el uso de bodegas alternativas.
Todos los escenarios presentan una mejora en los resultados de la empresa, los escenarios en 1 y 2, permitieron disminuir el inventario máximo de sacos llegando hasta 6.715 estibas, obteniéndose un aplanamiento de la curva de inventarios, sin afectar los premios de las motonaves de sacos.
Este beneficio, también se observa en el Escenario 3, en donde la planeación de las motonaves a través de reglas por diferenciación de tamaño de la motonave permite reducir el inventario máximo en las bodegas de carga suelta para 5.991 estibas, lo cual hace pensar en la posibilidad de prescindir una de las bodegas y representando un ahorro importante impactando directamente en los costos fijos de la compañía.

Tabla 2. Cuadro comparativo de los escenarios simulados

Indicador	Situación Actual	Escenario 1	Escenario 2	Escenario 3
Fila Media de camiones en la entrada del Puerto	28	32	34	8
Tiempo promedio de un camión en el sistema(horas)	26	28	29	13
Premios_CS ($USD)	50.697	50.490	48.400	55.894
Premios_GR ($USD)	23.750	20.238	15.407	14.598
Número Promedio de Camiones /día Carga Suelta	43	39	35	41
Máximo Número de Camiones /día Carga Suelta	72	71	71	72
Inventario Promedio Carga Suelta (Estibas)	2.044	1.808	1.572	1.598
Inventario Máximo Carga Suelta (Estibas)	7.421	6.715	6.715	5.991
Número Promedio de Camiones /día Granel	46	46	46	46
Máximo Número de Camiones /día Granel	60	59	61	62
Inventario Promedio Granel (Toneladas)	3.332	2.856	2.488	1.800
Inventario Máximo Granel (Toneladas)	23.358	23.358	21.896	19.830

Granel, en el escenario actual, el máximo inventario obtenido fue de 25.000 toneladas (Capacidad de la bodega=40.000 toneladas), y el inventario promedio de 3.332 toneladas, presentándose una baja utilización de la bodega para este fin.

En el escenario 1, al disminuir de 15 para 14 el número de días antes del ETA, el nivel de inventario promedio disminuyó (en 884 toneladas), al igual que los premios (US$3.512 por mes). Para estas operaciones se observa la alta sensibilidad de los costos ante la planeación de despachos, al no presentarse mejoras en ninguno de los escenarios propuestos, se hace pertinente que el análisis de las operaciones de granel sea realizada de manera independiente para cada nueva motonave programada y mediante la utilización de un modelo de simulación, que permita, tener cuantificada la sensibilidad del sistema ante cambios en las variables.

CONCLUSIONES

A través del análisis los diferentes escenarios desarrollados con la simulación, puede concluirse que:

La utilización de herramientas de simulación permite el análisis de sistemas complejos y de las variables implicadas en el proceso, ya que permite la investigación de la sensibilidad e impacto de los diferentes parámetros en el resultado del proceso.

Para el caso de la carga libre, puede encontrarse la alta sensibilidad del sistema a las modificaciones de fechas de envío de producto desde las empresas productoras, obteniéndose una reducción en los niveles de inventarios que podrían permitir elimina una bodega y así reducir los costos fijos del proceso. Igualmente, la sensibilidad de este parámetro, permite también que se incremente la ganancia de premios obtenidos por cumplimento en las fechas y reducción de problemas de falta de carga.

Para el caso de granel, el análisis no es tan sencillo, debido al volumen de los pedidos y la alta cantidad de camiones que deben ser movilizados para cumplir con los pedidos; cada navío es un nuevo caso y debe ser analizado como un sistema independiente, es allí que la herramienta de simulación como apoyo a la planeación es útil para ayudar a la empresa a mejorar la planeación de sus operaciones por medio de una programación

asertiva y flexible que permita minimizar el costo del inventario y maximizar los premios.

Futuros trabajos pueden desarrollarse uniendo la simulación a alguna técnica de optimización que permita encontrar nuevos resultados y evaluar diferentes escenarios para tomar mejores decisiones de forma rápida, efectiva y al menor costo.

AGRADECIMIENTOS

Los autores agradecen a la Universidad Nove de Julho (UNINOVE) por el apoyo financiero.

REFERENCIAS

[1] M.A. LAW. "How to build valid and credible simulation models". Proceedings of the 38th Winter Simulation Conference, v. 38, p. 58–66, 2006.

[2] Carson, J.S.II, Introduction to modelling and simulation, Actas de 2003 Winter Simulation Conference, 1, pp. 7-13, New Orleans, USA, 2003.

[3] R.Shannon "Introduction to the art and science of simulation". Proceedings of the 1998 Winter Simulation Conference, p. 7-14, 1998.

[4] HASAN, S., AL-HUSSEIN, M., Advanced simulation of tower crane operation utilizing system dynamics modeling and lean principles, Proceedings of the 2010 Winter Simulation Conference. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds. 2010

[5] M.C. Velez Gallego, D.A. Valencia Ramirez, C.A. Castro Zuluaga. "Heurístico de simulación-optimización para la configuración de un sistema de estantería selectiva simple". Ingeniare. Revista chilena de ingeniería. Vol. 20 N° 1, pp. 17-24. Abril de 2012. ISSN 0718-3305. DOI:10.4067/S0718-33052012000100003.

[6] M.J. Contreras, F. Bastleri, L. Maldonado. "Optimization of the harbor operations using simulation and surface response methodology". Revista Ingeniería Industrial. Vol. 9 N° 2, pp.73-92. 2010. ISSN 0717-9103.

[7] O. Yalcinkaya, G. Bayhan. "Modelling and optimization of average travel time for a metro line by simulation and response surface methodology". European Journal of Operational Research. Vol. 196, pp. 225-233. Julio de 2009. ISSN: 0377-2217. DOI:10.1016/j.ejor.2008.03.010.

[8] E.M. Rubio, M.A. Sebastian, A. Sanz, "Simulación de Sistemas Flexibles de Fabricación Mediante Modelos de Realidad Virtual". Información Tecnológica. Vol. 15 N° 1, pp. 49-54. 2004. ISSN 0718-0764 DOI:10.4067/S0718-07642004000100008.

[9] A. Higgins, G. Beashel , A. Harrison. "Scheduling of brand production and shipping within a sugar supply chain". Journal of the Operational Research Society. Vol. 57, pp. 490–498. Junio de 2005. ISSN 0160-5682. DOI: 10.1057/palgrave.jors.2602025.

[10] J.L. Calderón, F.C. Lario. "Simulación de Cadenas de Suministro: Nuevas Aplicaciones y Áreas de Desarrollo". Revista Información Tecnológica. Vol. 18 N°1, pp.137-146. 2007. ISSN 0718-0764. DOI: 10.4067/S0718-07642007000100018.

APÊNDICE C

REDUCTION OF THE INVENTORY IN A PACKAGING INDUSTRY USING DISCRETE EVENTS SIMULATION

Maria Alejandra Guerrero Hernandez (UNINOVE)
malejandragh@gmail.com
Andre Felipe H. Librantz (UNINOVE)
librantz@uninove.br

The simulation is a tool used mainly for analyzing complex systems, which can allow achieving significant cost reduction and improving the planning and control of productive systems, considering the constraints of the real system. In this ccontext, this work aimed to present a detailed process of a ProModel simulation for a packaging industry making possible, to determine and reduce the lead time and the inventory level. It also presented the advantages of simulation for decision support processes in complex systems with multiple variables. The results indicated a reduction up to 60% at current levels of inventory without affecting the productivity of machines, supporting financial indicators of the company and effectiveness in the supply chain management.

Keywords: Discrete Events Simulation, Inventory, Supply Chain

1 Introduction

The main purpose of Supply Chain Management (SCM) is to create value by means of the efficient operation of the processes involved; in this scenario, inventory management is highlighted because it has been a decisive factor for the success or failure of many organizations (GUNASEKARAN et. al 2009).

Inventories are a product of uncertainty between supply and demand, a perfect synchronization between this factors, could lead to zero inventory process, according Bayraktar et al. (2008), the sources of uncertainty are found in cases where there is the need to synchronize the demand with the supply.

Inventories have a high impact in the efficiency of the production and the level of service offered to customers; about 50% or more of the current asset of companies is composed by stocks (Christopher, 1997). Reduce inventories and maintain high service levels are the main challenges for organizations.

Huan et. al. (2004), concluded that studies and papers published in the SCM' literature are focused mainly in three specific categories: Operational, Design and Strategic Management. Mathematical tools, simulation, optimization and economical models are the most popular tools used for supporting SCM researches.

The simulation is nowadays a fundamental methodology for supporting the decision making; its flexibility allows users interact with the system, understand the whole process and test scenarios; these experiments would be impossible to implement in the real system and would imply high costs and unnecessary risk for businesses.

The use of optimization techniques in the Demand & Supply management begins with the definition of the classic model of economic order quantity (Wilson; 1934); multiple strategies have been used, i.e. The Predictive control Model (Garcia et al; 1989) and Deterministic Models using Linear Programming (Braun et al. 2003), however, some authors like Schwartz et. al (2006) argues that inventory organization's problems are solved mainly by SPSA technique (Simultaneous perturbation stochastic Approximation).

In literature; the simulation has been used in several areas; Calderon and Lario (2007) analyzed the simulation as a tool to improve the supply chain by searching in more than 70 articles, the main variables and parameters involved; Contreras et. al. (2010) used simulation to improve processes in a seaport and evaluated the tools required to fulfill the indicators demanded by customers; Librantz et. al (2010) used numerical simulation for maximizing the gain medium of a laser to a specific crystal. Ebert et al (2009) developed a simulation model of the dynamics of the operations of a slaughterhouse for poultry. Neiverth et. al (2010) developed a simulation model to improve the performance in a production line of Yerba Mate; Jesus (2010).

In this context, the aim of this work was to evaluate the optimal inventory levels for an important Packaging factory using discrete event simulation. It presented the construction of the simulation model, afterward an analysis of the results of the simulated scenarios and the advantages of using simulation tools for decision support in supply chain management.

2 Methodology

In this work, a case study for a packaging company was developed. This industry is responsible for producing food packaging; cleaning and personal hygiene' products. The production process includes: Extrusion, Thermoforming, Printing and Stamped. The company has three stocking points:

1. Raw material (material for extrusion);
2. Sheets and rolls; products that have passed through the extrusion process;

3. Thermoformed products waiting for been printed and stamped (customization of the product).
The research was conducted by interviews, plant visits and analysis of production reports for 12 months, it was found that the deterministic model in which could sort the inventory management of the company is Economic Order Quantity (EOQ). The warehouse resupply point (WRP) for the first point of inventory (sheets and rolls) was 60 tons; for the second and third inventory points (thermoformed product) the WRP was 2687 boxes.
The sheets, rolls and thermoformed products are manufactured and stored until the production order comes, after that, the packages are customized according to customer specifications.
The demand forecast is made considering the company's sales goals established annually by the company.
Based on the researches of Olhager and Ostlund (1990), the production system which this company can be classified is Assemble-to-order (ATO); hybrid system that combines the strategies: make to stock (MTS) and make to order (MTO), the successful of this production system depends on a high integration of all virtual drives in the supply chain and needs full alignment and a well-designed demand forecasting.
According to Chen et al. (2003) and Gunasekaran et al. (2005), the ATO system relies on a high integration of all levels of the supply chain: the manufacturer, the middleman, the assembler and distributor of the final product.
A diagnostic was made by visiting the factory and interviewing the managers. It was found that the stocks occupied areas that had not been allocated for this purpose and the managers agreed with this practice, arguing that this excessive amount of raw materials increased the flexibility and helped the company to fulfill the customer needs.
For Duclos et al. (2003) and Lummus et al. (2003); chain flexibility can be seen as the ability of a company to adapt to changing circumstances and thus remain a profitable enterprise, but although the need to be flexible was vital for company's survive; this strategy was harming the final business results, endangered the company staff (because all halls including evacuation were occupied); incremented the product lead time and reduced the confidence in the company inventory control.

3 Simulation of the Process

The strategy of inventory management gave the company high flexibility and helped shorten delivery times, but increasing the manufacturing costs because of unnecessary inventory levels, this situation together with the complexity and multiple variables involved, conducted to the decision of applying the methodology of discrete events simulation as the best strategy for analyzing the process and finding the significance of all the variables involved. The Software ProModel was selected due to its flexibility and the ability to interact with the system.

3.1 Parameters of the Simulation

In ProModel, all is set to paradigm Locations, Entities, and arrivals process, it can use for any system.

- Locations

Correspond to workstations; represent fixed places in the system where entities are directed to processing, storage, or making decisions. 146 Locations were defined; the main locations (machines), and storage areas like pallets for each machine were considered as locations.

- Entities

Entities are any object that is processed for the model.

38 entities were created. These entities are represented by requests for production, raw materials, sheets, rolls, boxes. Each product was identified with attributes: Product i.e. Industrial Glass, On-cup, Blister, dish. Material i.e. polypropylene, polystyrene. Weight (considered important since the performance of the machines depends on this parameter).

- Process

Describes the operations that are performed in each station, such operations can be: process time, resources required to perform the work, and the selection of the path for the entity.
This simulation model had 189 rows, and eight subroutines.

- Arrivals

Are the entities that feed the system and ensure the continuity of the model. This model began with the arrival of a request for production, and the medium stock of Raw Material and Product in Process.

3.2 Statistical Analysis

Performance by product for each machine was collected for a period of 12 months from the company production reports, this sample is considered significant and can cover any seasonality of the industry itself, the data were evaluated with StatFit Software, this software allows to adjust automatically continuous distributions, and translates the results into distribution functions appropriate to the programming language simulator.
The performance of the machines was adjusted in a Normal Distribution. This distribution is represented for two factors: N(mean, standard deviation).

Extrusion Process
Performance Measurement unit: hours / kilogram.
The machine performance is dependent of the material being processed (PP or PS):

Machine1_PS = ~ N (0.0021, 0.0007) h / kg
Machine 1_PP = ~ N (0.0045, 0.0008) h / kg
Machine 2_PS = ~ N (0.0036, 0.0007) h / kg
Machine 2_PP = ~ N (0.0036, 0.0007) h / kg

Thermoforming Process
Performance Measurement unit: minutes /gram
The performance of all thermoforming machines are equal, the variation depends on the weight of the product to be produced. We analyzed 38 references weight.
i.e. Performance of a product weighing 7g = ~ N (0.0022,0.0004) min/gr
The resulting products are packaged in boxes, for this product, the packaging capacity is 1200 units / box.

Stamp Process
The time stamping is defined with the functions
Stamp Machine 1 = ~ N (0.22, 0.052) min / blade
Stamp Machine 2 = ~ N (0.25, 0.10) min / blade
The resulting products are packaged in boxes i.e, Industrial Glass: packaging capacity = 1200 units / box.

Printing Process
The performance of the printing machines is analyzed by type of product and machine.
Performance measurement unit: time / box.
i.e Industrial Glass, the printing performance is defined as ~ N (0.16, 0.0634) h/box
The resulting product was stored on pallets and delivered to the client.

4 Final Parameters

The simulation started with an inventory of roll and sheets of 60 tons and with an inventory of 2,687 boxes of thermoformed product unprinted, (average company' stocks according to reports analyzed).
154 hours were simulated with 20 replications for each scenario in order to increase the accuracy of the model and understand the system behavior during many events.

5 Results and Discussion

The inventory management was analyzed with the EOQ. The model should deliver a level of inventory that does not compromise the factory productivity. The demand forecasts are made by considering the production performed in 12 continuous months.

5.1 Inventory of Sheets and rolls

For achieving the minimum acceptable level of inventory that ensures customer deliveries within the time limits and demand forecasts, the indicators analyzed were mainly: Percentage of uptime of the machines and percentage of time waiting for product.
Eight (8) scenarios were simulated with systematic inventory reductions of 17% beginning with the current stock of 60 tons.
Figure 1 showed the simulated scenarios demonstrating the impact of the lack of material in the machines downtimes.

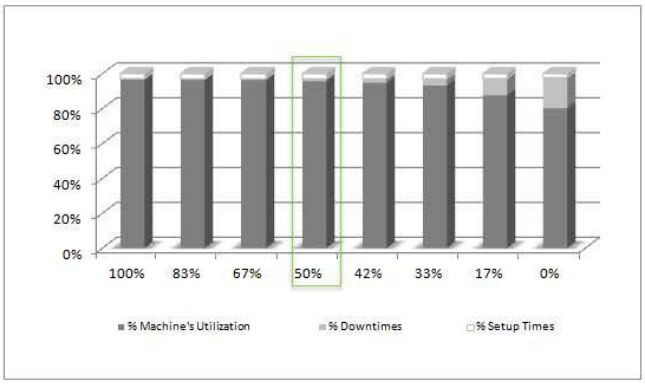
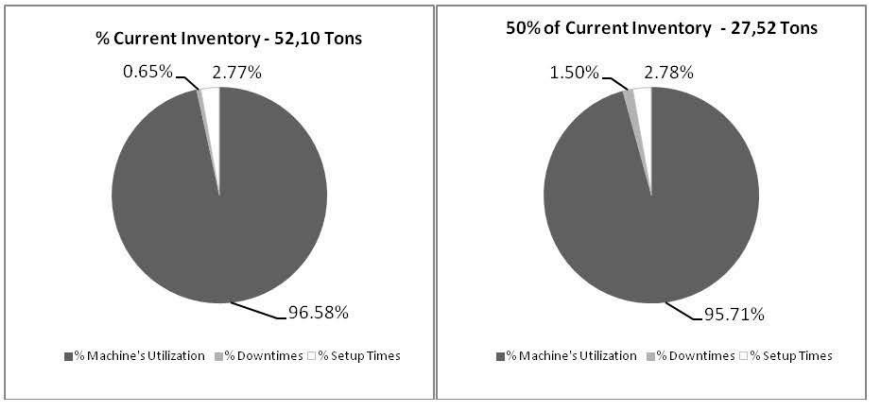


Figure 1 - Inventory reduction of Sheets and rolls

Table 1 - Inventory level in Tons when simulation is finished

% Inventory Reduction	Inventory in Tons when simulation is finished
100%	52.1
83%	45.4
67%	36.2
50%	27.5
42%	22.8
33%	18.4
17%	13.8
0%	10.1

The economic order quantity can be reduced up to 50% of the current (27.5 tons), as shown in Table 2,
Having 50% of the inventory, the machines downtime reached 1.5%, due of the lack of raw material, even with this slight lost of productivity, coming from 0.65% on the current scenario, the advantage in reducing the cost of capital was very important.
For the other simulated scenarios, the productivity reduction began to be more dramatic when then inventories were lower than 22.8 tons, reaching a downtime of 17.58% with zero inventories. This behavior can be observed in figure 2.



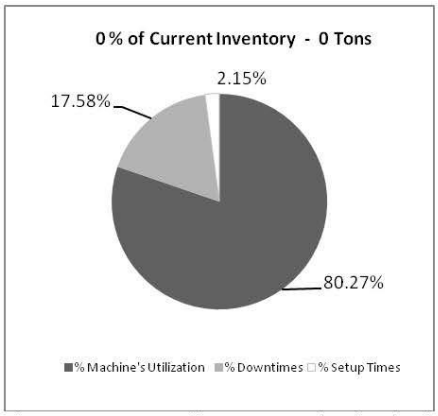


Figure 2 - Impact of inventory reduction in the productivity of the machines

5.2 Inventory of non-printed product

After the level of inventories of rolls and sheets were defined in 27.5 tons, new scenarios were tested to find the best level of inventory in the warehouse of product without printing. These scenarios were defined by performing successive reductions of 10% in the initial stock (2,687 boxes) until achieving the zero inventory level.

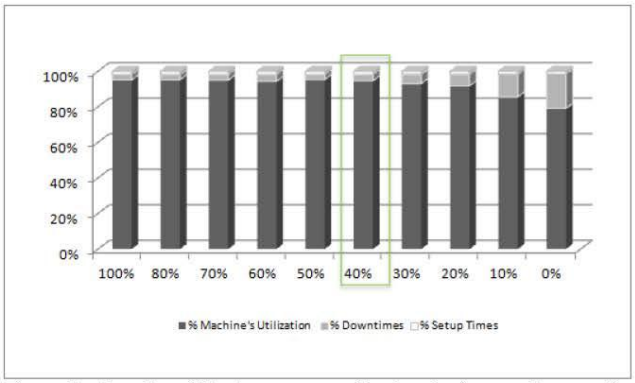


Figure 3 - Results of the inventory reduction in the warehouse of products without printing

The inventory can be reduced up to 40% of the current, so the EOQ will be of 1,074 boxes, the machine productivity is not affected, and the room for this product at its warehouse will be sufficient, having no need to use any more corridors or safe areas as temporary warehouse.
The impact of inventory reduction in the printing presses downtimes increased, starting at 3.01% with current inventory and reaching 3.54% with 1,074 boxes. The productivity dropped 0.48%; acceptable reduction, considering the gains in the cost of capital, the accuracy in the control of stocks and the reduction in security risks.
In the other simulated scenarios, the reduction of the productivity went down 19.40% when working without stock. This behavior can be observed in figure 4.

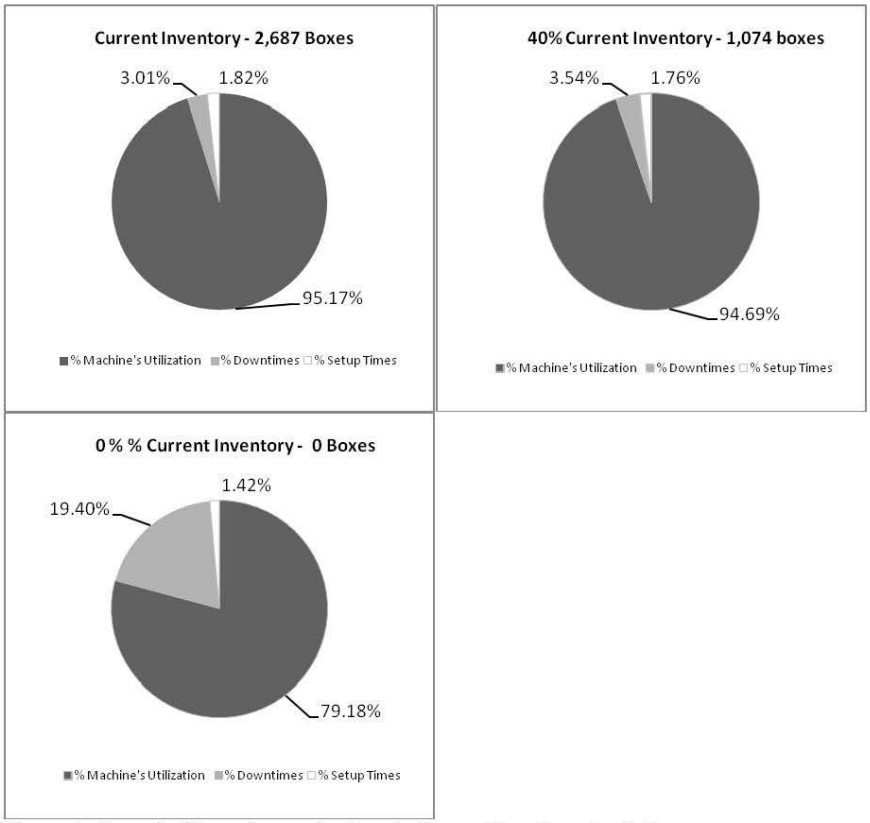


Figure 4 - Impact of inventory reduction in the machines' productivity

With the reduction in inventory, the product cycle time is reduced by 41%, from an average of 58.8 hours to 34.8 hours thereby it is possible to improving the response time to the client with the minimum inventory (Table 2)

Cycle Time (Hours)

100% Inventory	50% inventory	
58.8	34.8	*Mean*
1.8	2.9	*Standard Deviation*
55.7	30.9	*Maximum*
60.9	40.2	*Minimum*

Table 2. Reduction of cycle time after the inventory reduction

6 Conclusion

In this work, the simulation allowed to analyze all the parameters involved in the manufacture of plastic packaging, in order to find the best inventory levels, so as to keep a high percentage of machines in use, reducing product lead time and capital invested. This kind of analysis allows creating synergies between all areas, supporting the improvement of financial results by improving the performance indicators in the supply chain.

The simulation provided a high flexibility to develop new scenarios and test the behavior of the system with different inventory levels. This flexibility made possible to reduce the inventory of sheets and rolls up to 50%, and the stock of product in process by up to 40% of the current level. The multiple scenarios simulated was based on statistical analysis and probability distributions with high accuracy of the reality that leads to the increase of the results confidence level.

The results pointed that the discrete event simulation can be a good tool for analyzing complex and dynamic systems with multiple variables and it can be used for resolving problems in any type of system (production, services, etc.).

7 References

BAYRAKTAR, E.; KOH, S. C. L.; SARI, K.; TATOGLU, E. The role of forecasting on bullwhip effect for E-SCM applications. International Journal of Production Economics, v. 113, p. 193-204, 2008.

BRAUN, M. W.; RIVERA, D. E.; FLORES, M. E.; CARLYLE, W. M.; KEMPF, K. G. A model predictive control framework for robust management of multi-product, multi-echolon demand networks. Annual Reviews in Control, v. 27, p. 229–245, 2003.

CALDERON, J.L. & LARIO, F.C. Simulación de Cadenas de Suministro: Nuevas Aplicaciones y Áreas de Desarrollo. Revista Información Tecnológica, v. 18, p.137-146, 2007.

CHEN, R.S.; LU, K.Y.; YU, S.C.; TZENG, H.W.; CHANG, C.C. A case study in the design of BTO/CTO shop floor control system, Information and Management, v. 41, n. 1, p. 25-37, 2003.

CHRISTOPHER, M. Logística e Gerenciamento da Cadeia de Suprimento: estratégias para a redução de custos e melhoria dos serviços. São Paulo: Pioneira, 1997.

CONTRERAS, M.J.; BASESTLERI, F. & MALDONADO, L. Optimization of the harbor operations using simulation and surface response methodology. Revista Ingenieria Industrial, v.2, n.9 p.73-92, 2010.

DUCLOS, L.K; VOKURKA, R.J.; LUMMUS, R.R. A conceptual model of supply chain flexibility. Industrial Management & Data Systems, v. 103, n. 6, p. 446-56, 2003.

EBERT , D.C; SILVA, L.C & VILAS BOAS, M.A. Simulação da dinâmica operacional de um processo industrial de abate de aves. Ciência e Tecnología de Alimentos. v. 29, p.305-309, 2009.

GARCIA, E. S.; REIS, L.; MACHADO, L.R. & FILHO, V. F. Gestão de estoques, otimizando a logística e a cadeia de suprimentos. Rio de Janeiro: Petrobras, 2006.

GUNASEKARAN, A.; NGAI, E.W.T. Build-to-order supply chain management: a literature review and framework for development. Journal of Operations Management, v. 23, n. 5, p. 423-51, 2005

GUNASEKARAN, A.; NGAI, E.W.T. Modeling and analysis of build-to-order supply chains. European Journal of Operational Research, v. 195, p. 319-334, 2009.

HUAN, S.H., SHEORAN, S.K.M; WANG, G. A review and analysis of supply chain operations reference (SCOR) model. Supply Chain Management: An International Journal. v.9, p. 23-29.

JESUS, D.A. A análise da sinergia entre o mapeamento do fluxo de valor e a simulação computacional para o aumento da produtividade em sistemas de manufatura: Um estudo de caso em uma lina de montagem multi-modelos. XXX Encontro Nacional de Engenharia de Produção, 2010.
LIBRANTZ, A.F.; ALVES, W.A.; ARAUJO, S.A.; GOMES, L. & LIMA, A.C. Optimização da operação de meios lasers com aplicações em engenharia por medio de simulação numérica. XVII Simposio de Engenharia de Produção, 2010.
LUMMUS, R.R.; DUCLOS, L.K.; VOKURKA, R.J. Supply chain flexibility: building a new model. Global Journal of Flexible Systems Management, v. 4, n. 4, p. 1-13, 2003.
NEIVERTH, C.A; FREITAS, E.R & COLMENERO, J.C, Simulation for analiysis and performance of the production line of the industrial drying of Yerba Mat. XVI International conference on industrial engineering and operations management, 2010.
OLHAGER, J.; OSTLUND, B. An integrated push-pull manufacturing strategy. European Journal of Operational Research, v. 45, n. 2/3, p. 135-42, 1990.
SCHWARTZ, J. D.; WANG, W.; RIVERA D.E. Simulation-based optimization of process control policies for inventory management in supply chains. Automatica (Journal of IFAC), v. 42, n. 8, p. 1311-1320, 2006.
WILSON, R. H. A scientific routine for stock control. Harvard Business Review, v. 13, p. 116–128, 1934.

APÊNDICE D

International Conference on Industrial Engineering and Operations Management
ICIEOM
Guimarães - Portugal
July 9 - 11, 2012

Maximizing Profits with the Improvement in Product Composition, Using Genetic Algorithms and K-Means Application to a Company of the Printing Industry

María A. Guerrero, Rodrigo A. Batistelo and André F. H. Librantz

Industrial Engineering Post Graduation Program, Nove de Julho University (UNINOVE), São Paulo, Brazil

Email: malejandragh@gmail.com, rodrigo.batistelo@gmail.com, librantz@uninove.br

Abstract

In recent years, the printing industries have struggled to remain competitive due to increased concurrence and the appearance of new electronic media aimed at replacing the printed paper. In this context arises the need, as not only high technology, as optimization methods for such industries that can support decision making aiming cost reduction and profit improvements. In this context, it was proposed to maximize the profits of a Colombian printing industry by means of the best composition of their products to be sold, in order to improve its contribution margin (financial remainder of production of each product) and have a strategic vision for the sale of its many products. For this, the products were grouped by similarity, using clustering algorithm called K-Means and it was applied genetic algorithms technique for maximize the Contribution Margin. Results pointed that it was possible to achieve good increments in the revenues and contribution margin, as well.

Keywords: printing industry; optimization methods; genetic algorithm; k-means algorithm.

1 Introduction

1.1 Definition of the Problem

Over the last decade, the printing industry has been faced numerous problems, mostly because of the fierce competition, and the emergence of new communications, graphic and electronic media, which have replaced the printed product. However, lots of companies that are still working in this industry are day-to-day seeking to become highly competitive and efficient. Investments in technology are often not sufficient and do not solve entirely the problem, it is also necessary to develop and apply methodologies in decision making to better use the companies' productive resources. The high quality standards are not enough for competing. For most of the companies, their management boards have sophisticated quality control equipments that help to produce similar and high quality products making very difficult to obtain a remarkable differentiation from the competitors.

The best way to compete is by offering the best prices, even sacrificing utility. Sales budgets are usually not specific, considering only sales by customer or geographic location, no matter in which products the company has the best efficiency and productivity, or those which the company has the bigger contribution margin.

1.2 Company features

The experiments were performed on a printing company located in Colombia.

According to the processes classification defined by (Davis, Aquilano, Nicholas, & Chase, 2001), the type of process where this company can be classified is: Make to Order Production System, which focuses the production in highly customized products (the products are produced with the customer's specifications). According to the same author, these types of processes require more flexibility than the Make to Stock Production System and as a result, tend to be slower, inefficient, and consequently more expensive. Due

to the high diversity in customer orders and the long list of products generated / produced, the tasks of classifying the products, taking into consideration the characteristics of contribution margin and productivity, become very difficult if only traditional methods are adopted by the company, i.e. Excel.

The company has already made some classifications, based on the physical characteristics of the products; defining thirty-five possible product groups (clusters), but each one of these clusters could have important variations of productivity and contribution margins.

2 Scope

The goal of this study is to maximize the utility of the company and to determine which is the best composition of products, that provides the best contribution margin, and thus offer the Sales & Marketing departments, a solid basis for defining sales budgets and also guide them about the products to be sold, based on the best relation between costs and productivity impacting greatly the financial results.

2.1 Relevant Variables for Analysis

For establishing the best composition of products, are considered the two following variables:

- Contribution Margin ($ - %)

Contribution margin is equal to sales revenue minus variable expenses (both manufacturing and non-manufacturing). In other words, the contribution margin means: A measurement of the profitability of a product is the financial leftover of each product's production; it is used for the amortization of fixed costs. Its understanding allows discussions and actions focused on the achievement of the expected utility.

- Production Efficiency: Time Adding Value to the Product

It is measure of the efficiency used by the company. It is the amount of time that the company employs for adding value to the product.

3 Bibliographic Review

3.1 K-Means Algorithm

K-means is a well known clustering algorithm. It consists of a method for clustering analysis whose goal is to partition a set of samples into k clusters in which each sample belongs to the cluster with the nearest mean (Mitchell, 1997). Given a set of samples (x1, x2, ..., xn), where each sample is a n-dimensional real vector, k-means algorithm aims to partition the set of samples into k sub-sets S = {S1, S2, ..., Sk}, optimizing an objective function F that can be defined as:

$$F = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j \quad \mu_i \right\|$$

where μi is the mean of points in Si.

The most common algorithm uses an iterative technique that, in general, considers Euclidean distance as a similarity measure of vectors and variance as a measure of cluster scatter. Since the number of clusters *k* is an input parameter of the algorithm to define the *k* centroids, an inadequate choice of *k* can generate poor results. Thus, for using k-means, it is important to run diagnostic checks for determining the suitable number of clusters for the considered data set.

3.2 Genetic Algorithms

The Genetic Algorithms (GA) consists in a technique of optimization inspired by natural evolution theory. It has been used for solving optimization problems in several areas over the last decades, mainly because its efficiency in irregular search spaces (Goldberg, 1989; Haupt & Haupt, 1998; Mitchell, M., 1998). The GA generally uses binary strings called chromosome or individual to represent solutions of the problem. At the beginning, a population of individuals is randomly generated. The size of population depends on the problem to be solved (Haupt & Haupt, 1998). By means of competition, the most able chromosomes of the population are selected and crossed each other, to generating new chromosomes better than those ones of the previous population. So, at each generation the probability of one or more individual to be a solution of the problem is increased (Goldberg, 1989; Haupt & Haupt, 1998; Librantz, Coppini, Baptista, Araújo, & Rosa, 2011; Santana, Araújo, Librantz, & Tambourgi, 2010). A GA can find the global optimum solution in a complex multi-modal search space without requiring specific knowledge about the target problem that it was developed for solve. Besides, a GA operates over the population in parallel, yielding various solutions at a time. Hence, this method has found applications in engineering problems involving complex combinatorial optimization (Librantz, Coppini, Baptista, Araújo, & Rosa, 2011; Santana, Araújo, Librantz, & Tambourgi, 2010). GA procedure involves four main operations: Evaluation, selection, crossover and mutation. In the evaluation operation fitness function is used to measure the aptitude of the individuals of the population, providing information such as the number of new individuals each one can generate according with its aptitude. The selection consists of the choice of the best individuals for reproduction. Hence, the individuals with better aptitudes are selected while the other ones are discarded, that is, each individual has a probability to be selected according to its aptitude. In the crossover operation, the genetic material of the best chromosomes is mixed to generate the individuals of the next population. Finally, a random change in a small number of bits, with some small probability, is performed in order to preventing the population of chromosomes becoming too similar to each other or, in other words, to preserve the diversity of the population. This operation is essential to avoid the premature convergence.

It is important to remember that the suitable convergence of the GA depends on some parameters such as: size of chromosome, size of population, crossover points and mutation rate (Librantz, Coppini, Baptista, Araújo, & Rosa, 2011). Unfortunately, the determination of these parameters is a difficult task because and, in general, they are empirically defined (Pacheco, 1999).

4 Methodology

Here, it was described how we work with the data available, the application of optimization techniques, the objective function definition and the tools used in this work.

4.1 Data

In order to develop the experimental part of this project it was used historical data from 10 months of a company in the printing industry. These data were obtained from a corporate database ORACLE (Figure 1) and add a total of 2618 records with data on products, customers, vendors, business, revenue, product costs, contribution margin and time adding value to the product (sec / und). The classification of products was defined by the analysts of the product and classified into 35 groups (clusters), taking into account basically the shape of the product.

CLUSTER	PRODUCT	MONTH	REVENUES	SALES VOLUME (UNITS)	PRODUCTION COSTS	TOTAL COSTS	CONTRIBUTION MARGIN (%)	EGP	EFFICIENCY (SEC/UNIT)
26	LIBROS - INTERES GENERAL	1	4.871.988	1.000	2.679.593	3.410.392	0,45	61%	0,23
51	PAPELERIA - CUADERNO	1	2.479.997	200	1.363.998	1.735.998	0,45	53%	0,28
80	REVISTA - REVISTA	1	70.985.654	600.200	49.689.958	60.337.806	0,30	36%	0,38
65	POP - POP	1	974.400	12.000	623.616	769.776	0,36	63%	0,22
24	LIBROS - ENCICLOPEDIA/COLECCION	1	159.067.854	3.167	119.300.891	143.161.069	0,25	12%	0,53
82	REVISTA - CIRCULACION ABIERTA	1	12.377.200	19.400	8.787.812	10.644.392	0,29	37%	0,38
83	REVISTA - CIRCULACION CERRADA	1	4.455.104	69.611	3.341.328	4.009.594	0,25	32%	0,41
82	REVISTA - CIRCULACION ABIERTA	1	5.342.700	32.380	4.007.025	4.808.430	0,25	21%	0,47
82	REVISTA - CIRCULACION ABIERTA	1	29.200.000	40.000	21.024.000	25.404.000	0,28	37%	0,38
80	REVISTA - REVISTA	1	11.772.585	23.783	8.711.713	10.477.601	0,26	23%	0,46
82	REVISTA - CIRCULACION ABIERTA	1	24.038.055	247.815	17.067.019	20.672.727	0,29	38%	0,37

Figure 1: Sample of data from Oracle database (Colombian printing industry).

4.2 Tools and Procedures

Because of the wide variety of products and to better analyze the product portfolio, the project was divided in two stages:

- Grouping by similarity of products, using clustering algorithm K-Means, which allowed grouping the products in nine different clusters. For this, it was used function K-Means of the MATLAB software.
- Minimization of the objective function using Genetic Algorithms Toolbox of the MATLAB software.

4.2.1 Application of K-Means Algorithm

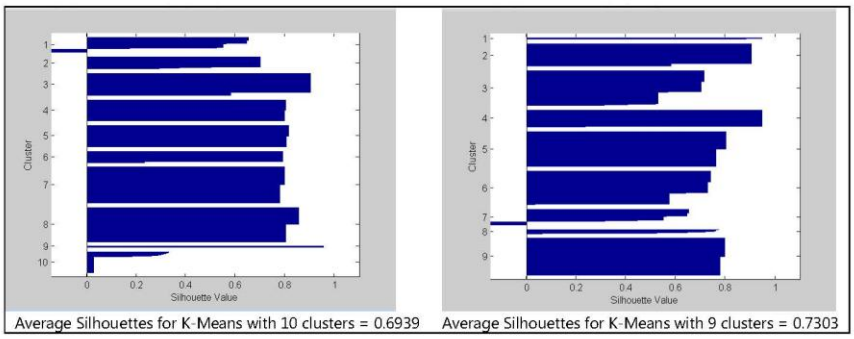
The number of clusters was defined and evaluated through graphics of silhouettes given by K-Means.

- **Graphics of Silhouettes**

The graphic silhouette is a graphical representation of the quality of generated clusters. This chart can give an idea if the resulting number of clusters was set properly. These graphics are shown in Figure 2.

The silhouettes show a measure of how close is the point of a given cluster of points of neighboring clusters. This measure ranges from +1, indicating points which are distant neighboring clusters, through 0 indicating which are difficult to classify, according to their cluster, to -1 indicating points which area assigned to the cluster probably wrong.

The database was analyzed to be clustered in 10, 9, 8 and 7 clusters.



Average Silhouettes for K-Means with 10 clusters = 0.6939 Average Silhouettes for K-Means with 9 clusters = 0.7303

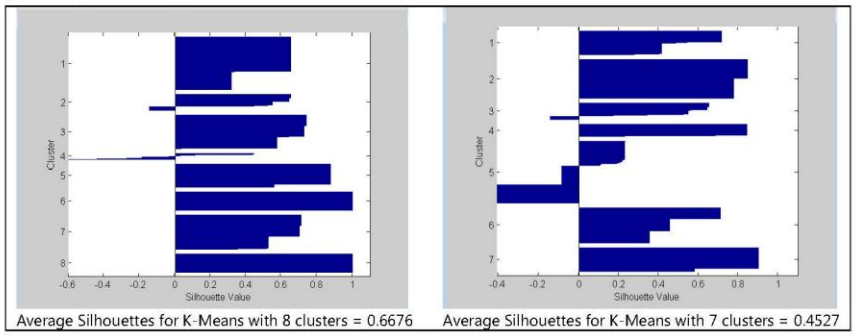


Figure 2: The graphic silhouette for several cluster numbers.

- **Number of Clusters Definition**

Based on the results evaluated by the graphics of silhouettes for 7, 8, 9 and 10 clusters, it was obtained the following results (Table 1):

Table 1: Average silhouettes for several cluster numbers.

CLUSTER	7	8	9	10
AVERAGE SILHOUETTES	0,4527	0,6676	0,7303	0,6939

It was chosen the cluster nine, once it showed the best average silhouettes values.

- **MATLAB K-Means function**

Based on the results evaluated by the graphics of silhouettes for 7, 8, 9 and 10 clusters, it was executed in MATLAB environment the command lines:

1) X = textread ('C:/ICIEOM2012/Database_K-Means.txt');
2) v_options = statset ('Display', 'final');
3) [value, centroid] = kmeans (X, 9, 'Distance', 'city', 'Replicates', 10, 'Options', v_options)

After that, it was possible to obtain the clustering by K-means. In Table 2, it was shown the labels for the values of centroids for each cluster, returned after the execution of MATLAB commands listed above.

Table 2: Labels for the centroids.

CENTROID1	CENTROID2	CENTROID3	CENTROID4	CENTROID5
0.4500 0.3300	0.5500 0.2700	0.4900 0.3100	0.2500 0.4500	0.2900 0.4300
CENTROID6	**CENTROID7**	**CENTROID8**	**CENTROID9**	
0.8400 0.0900	0.2100 0.4700	0.4600 0.3200	0.5800 0.2500	

The K-Means ranks with numbers between 1 and 9, the data whose centroids were calculated. The chat below (Figure 3) shows the clustering obtained by K-Means.

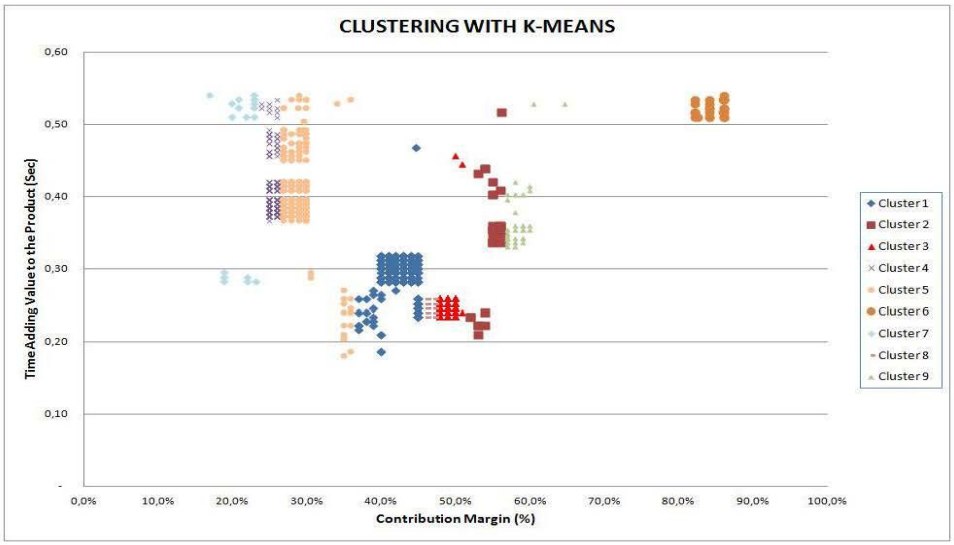


Figure 3: Clustering obtained by K-Means.

These numbers were taken back to the database, to label the data and have a good classification to continue with the phase of minimization with Genetic Algorithms (GA).

4.2.2 Application of Genetic Algorithms (GA)

The project´s final goal is to maximize the profit of a printing company doing the best composition of the products to be sold to improve the contribution margin of the company. As the Genetic Algorithms Toolbox of the MATLAB works only by minimizing functions, we seek to minimize costs so as to maximize the profit.

- **Definition of Fitness Function / Objective Function**

Taking into consideration the purpose of the present work, understanding the problem as much as the business, we chose the following variables (Table 3) to compose our fitness function:

Table 3: Variables chosen to compose the fitness function.

P	**PRICE ($/unit)**
M	**CONTRIBUTION MARGIN ($)**
R	**EFFICIENCY (sec/unit)**
V	VOLUME (unit)
I	**CLUSTER**

Then the fitness function that minimizes the costs of the company was defined as:

$$\sum_{i=1}^{i=9} \frac{P_i(1-M_i)}{R_i} \times V_i$$

The volume for each cluster (Vi) is the variable that will be found by the GA and which will provide an ideal composition of groups of products, in other words, it will inform the volume of each group of products to form the preferred portfolio sales. This result could guide the sales department at the moment of product offering.

- **Fitness Function in MATLAB**

Below is the function fitness.m, created in MATLAB for execution of its toolbox of genetic algorithms. It is that allows us to obtain, through a data file, the appropriate volumes of each of the groups (clusters) of product.

```
function result = fitness (V)
   X = textread ('C:/ICIEOM2012/Database_GA.txt');
   result = 0;
   tam = length (X(:,1));
   for i=1:tam
      R = X (i, 1);
      M = X (i, 2);
      P = X (i, 3);
      result = result + (((P*(1-M)) / R) * V(i));
   end
end
```

- **Input Data**

The input data of the file used in the fitness function, are prepared as follows (Table 4):

Table 4: Sample data used in applying the fitness function.

EFFICIENCY (sec/unit)	CONTRIBUTION MARGIN (%)	PRICE PER UNIT PRODUCED ($/unit)
0.26	0.464	2.29
0.24	0.441	1.00
0.38	0.253	0.44
0.25	0.493	1.66
0.24	0.482	2.26
0.42	0.562	5.16
0.43	0.271	0.58
0.38	0.287	0.44
0.30	0.421	2.61

The principal historical data are summarized in the Table 5 below:

Table 5: Principal historical data.

REVENUES (BRL)	PRODUCTION COSTS (BRL)	CONTRIBUTION MARGIN (%)	SALES VOLUME (units)
$ 90,194,579.1	$ 51,012,304.6	43%	83,934,774

- **Restrictions**

In this work we have established a single primary constraint, which is the sum of the volume of the nine clusters limited to 84 million units, which is the maximum production capacity of the factory according to the information of the production's analyst.

$$\sum_{i=1}^{i=9} V_i = 84.000.000$$

- **Definition of Chromosome**

The chromosome of the objective function is composed of 9 genes; each gene represents the volume of each cluster, which corresponds to the objective variables, those found by the genetic algorithm.
The way in how the genetic algorithms are presented is based on the binary system, the number of alleles of each gene is defined by the function: 2^k, where k is the size of the gene, making the conversion to the decimal system should contain the value of 84 million (restriction of number of production units described above). In this case, the k value for the gene studied is 26, so the number of alleles for the gene would be 27 (2^{26}, where $0 \leq k \leq 26$). All genes in the chromosome have the same structure as presented in Figure 4.

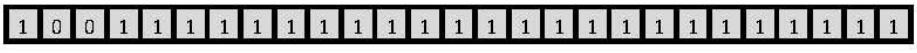


Figure 4: Representation of the gene with k=26

In general it can be stated that all variables in this chromosome may represent a maximum of 84,000,000 units, meanwhile the other variables are 0.
The chromosome is composed of 9 genes (cluster), each with 27 alleles, totalizing 243 alleles, which can take values 1 or 0 of the binary system, respecting the constraint that the sum in decimal is less than or equal to 84,000,000.

4.3 Results and discussion

Some scenarios were analyzed by changing the values of the operators of the Genetic Algorithm. Each of the proposed scenarios is presented below. They show the improvements obtained with respect to the current scenario / position of the company.

4.4 Scenario One

For scenario one, the operators were chosen, as shown in Table 6, both; the value of fitness function and the best individual of Genetic Algorithms for this scenario are shown in Figure 5.

Table 6: Operators of genetic algorithms, referring to scenario one.

Population Size: 50, **Selection**: Tournament, **Crossover Fraction**: 0,8, **Crossover**: Two Point, **Mutation**: Use constraint dependent default

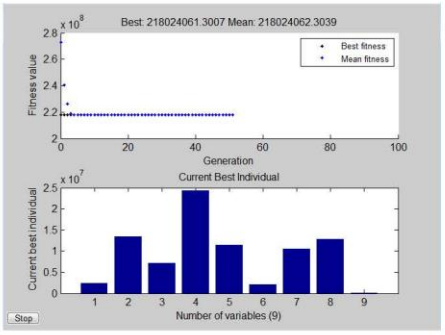


Figure 5: Value of fitness function and best individual obtained with genetic algorithms in scenario one.

Below (Figure 6), we are presenting a comparison between the company's current situation and the result obtained in scenario one.

	Current Scenario				Scenario One			
CLUSTER	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)
1	1.040.438	$ 2.386.111,18	$ 1.107.155,59	46,40%	2.445.932	$ 5.609.431,99	$ 2.602.776,45	46,40%
2	5.099.756	$ 5.111.750,75	$ 2.254.282,08	44,10%	13.432.721	$ 13.464.313,73	$ 5.937.762,35	44,10%
3	12.488.002	$ 5.505.058,63	$ 1.392.779,83	25,30%	7.173.454	$ 3.162.257,95	$ 800.051,26	25,30%
4	3.194.999	$ 5.295.899,88	$ 2.610.878,64	49,30%	24.285.046	$ 40.253.897,65	$ 19.845.171,54	49,30%
5	2.894.923	$ 6.545.844,40	$ 3.155.097,00	48,20%	11.385.257	$ 25.743.730,68	$ 12.408.478,19	48,20%
6	6.833.058	$ 35.235.058,03	$ 19.802.102,61	56,20%	2.012.081	$ 10.375.413,08	$ 5.830.982,15	56,20%
7	28.977.042	$ 16.914.815,88	$ 4.583.915,10	27,10%	10.524.630	$ 6.143.559,32	$ 1.664.904,57	27,10%
8	22.044.625	$ 9.643.743,09	$ 2.767.754,27	28,70%	12.740.879	$ 5.573.683,77	$ 1.599.647,24	28,70%
9	1.361.930	$ 3.556.297,31	$ 1.497.201,17	42,10%	0	$ 1,18	$ 0,50	42,10%
	83.934.774	**$ 90.194.579,14**	**$ 39.171.166,29**	**43,43%**	**84.000.000**	**$ 110.326.289**	**$ 50.689.774**	**45,95%**

Figure 6: Comparison between the current scenario and scenario one evaluated.

This scenario generated an improvement of $ 20.131.710BRL in billing and a contribution margin increase of $ 11.518.607BRL, which corresponds to an improvement of 5.48%.

4.5 Scenario Two

For scenario two, the operators were chosen, as shown in Table 7 and both; the value of fitness function and the best individual of Genetic Algorithms for this scenario are shown in Figure 7.

Table 7: Operators of genetic algorithms, referring to scenario two.

Population Size: 20, **Selection**: Stochastic uniform, **Crossover Fraction**: 0,8, **Crossover**: Scattered, **Mutation**: Use constraint dependent default

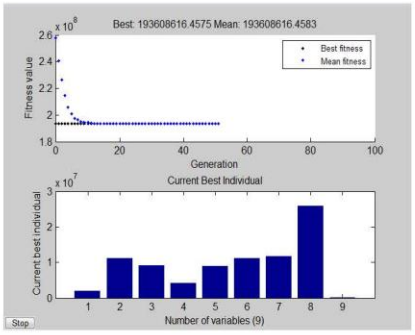


Figure 7: Value of fitness function and best individual obtained with genetic algorithms in scenario two.

Below (Figure 8), we are presenting a comparison between the company's current situation and the result obtained in scenario two.

	Current Scenario				Scenario Two			
CLUSTER	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)
1	1.040.438	$ 2.386.111,18	$ 1.107.155,59	46,40%	1.978.189	$ 4.536.723	$ 2.105.040	46,40%
2	5.099.756	$ 5.111.750,75	$ 2.254.282,08	44,10%	11.107.635	$ 11.133.760	$ 4.909.988	44,10%
3	12.488.002	$ 5.505.058,63	$ 1.392.779,83	25,30%	9.025.165	$ 3.978.544	$ 1.006.572	25,30%
4	3.194.999	$ 5.295.899,88	$ 2.610.878,64	49,30%	4.223.245	$ 7.000.279	$ 3.451.137	49,30%
5	2.894.923	$ 6.545.844,40	$ 3.155.097,00	48,20%	8.960.041	$ 20.259.963	$ 9.765.302	48,20%
6	6.833.058	$ 35.235.058,03	$ 19.802.102,61	56,20%	11.107.635	$ 57.277.161	$ 32.189.764	56,20%
7	28.977.042	$ 16.914.815,88	$ 4.583.915,10	27,10%	11.773.092	$ 6.872.326	$ 1.862.400	27,10%
8	22.044.625	$ 9.643.743,09	$ 2.767.754,27	28,70%	25.824.997	$ 11.297.522	$ 3.242.389	28,70%
9	1.361.930	$ 3.556.297,31	$ 1.497.201,17	42,10%	0	$ 0	$ 0	42,10%
	83.934.774	**$ 90.194.579,14**	**$ 39.171.166,29**	**43,43%**	**84.000.000**	**$ 122.356.278**	**$ 58.532.592**	**47,84%**

Figure 8: Comparison between the current scenario and scenario two evaluated.

This scenario generated an improvement of $ 32.161.698BRL in billing and a contribution margin increase of $ 19.361.426BRL, which corresponds to an improvement of 9.21%.

4.6 Scenario Three

The operators were chosen, as shown in Table 8, and both; the value of fitness function and the best individual of Genetic Algorithms for this scenario are shown in Figure 9.

Table 8: Operators of genetic algorithms, referring to scenario three.

Population Size: 20, **Selection**: Tournament, **Crossover Fraction**: 0,8, **Crossover**: Intermediate, **Mutation**: Use constraint dependent default

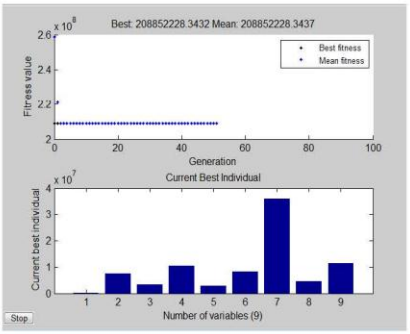


Figure 9: Value of fitness function and best individual obtained with genetic algorithms in scenario three.

Below (Figure 10), we present a comparison between the company's current situation and the result obtained in scenario three.

	Current Scenario				Scenario Three			
CLUSTER	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)	SALES VOLUME (UNITS)	REVENUES	CONTRIBUTION MARGIN ($-BRL)	CONTRIBUTION MARGIN (%)
1	1.040.438	$ 2.386.111,18	$ 1.107.155,59	46,40%	0	$ 0	$ 0	46,40%
2	5.099.756	$ 5.111.750,75	$ 2.254.282,08	44,10%	7.467.111	$ 7.484.674	$ 3.300.741	44,10%
3	12.488.002	$ 5.505.058,63	$ 1.392.779,83	25,30%	3.386.922	$ 1.493.050	$ 377.742	25,30%
4	3.194.999	$ 5.295.899,88	$ 2.610.878,64	49,30%	10.384.749	$ 17.213.336	$ 8.486.174	49,30%
5	2.894.923	$ 6.545.844,40	$ 3.155.097,00	48,20%	2.732.452	$ 6.178.473	$ 2.978.024	48,20%
6	6.833.058	$ 35.235.058,03	$ 19.802.102,61	56,20%	8.212.544	$ 42.348.455	$ 23.799.831	56,20%
7	28.977.042	$ 16.914.815,88	$ 4.583.915,10	27,10%	35.887.014	$ 20.948.385	$ 5.677.012	27,10%
8	22.044.625	$ 9.643.743,09	$ 2.767.754,27	28,70%	4.624.897	$ 2.023.229	$ 580.667	28,70%
9	1.361.930	$ 3.556.297,31	$ 1.497.201,17	42,10%	11.304.311	$ 29.518.034	$ 12.427.092	42,10%
	83.934.774	**$ 90.194.579,14**	**$ 39.171.166,29**	**43,43%**	**84.000.000**	**$ 127.207.634**	**$ 57.627.284**	**45,30%**

Figure 10: Comparison between the current scenario and scenario three evaluated.

This scenario generated an improvement of $ 37.013.055BRL in billing and a contribution margin increase of $ 18.456.117BRL, which corresponds to an improvement of 4.13%.

5 Conclusions

In this work it was proposed to maximize the profits of a Colombian printing industry by using k-means clustering algorithm combined to GA technique.

The printing industry is characterized by a high variety of products, because they are responding to customer specifications. Appraising the best mix of products and the quantity of units produced per product type is extremely important, but this analysis is technologically unfeasible and would result in a long list of objective variables to be be analyzed by the seeking algorithm, in this case, the genetic algorithm. The classification of these products requires that the patterns' identification is accurate and that leads to an effective as well as assertive classification to assist the decision taking.

The use of clustering techniques with K-Means proved to be a good alternative when it is necessary to group this type of "population" as well as reducing the number of variables to be used and of course, evaluated by the genetic algorithm. This genetic algorithm is a robust and flexible approach to evaluate alternatives that improve and optimize the performance of the company and also adapt them to market conditions.

These works contribute to improve the rentability of the companies, providing important information such as: What products to promote/sell considering, not only their contribution margin, but also, which products have the best performance in production, so the company can improve both; financial and productive indicators.

The results of this research demonstrated the great importance of the combination of these two techniques for supporting the decision taking. The experiments allowed offering different alternatives for the mixture of products that enable the company to improve its financial performance and adapt its production to the market needs.

References

Davis, M. M., Aquilano, N. J., & Chase, R. B. (2001). Fundamentos da Administração da Produção, 3ª Edição, Artmed Editora.

Goldberg, D.E. (1989). Genetic algorithms in search, optimization, and machine learning. Addison-Wesley: Massachusetts, 432 pp.

Haupt, R.L. & Haupt, S.E. (1998). The Binary Genetic Algorithm, In: Haupt, R.L; Haupt, S.E. Pratical Genetic Algorithms (1 ed.). Wiley-Interscience: New York, 276 pp.

Librantz, A. F. H., Coppini, N. L., Baptista, E. A., Araújo, S. A., & Rosa, A. F. C. (2011). Genetic Algorithm Applied to Investigate Cutting Process Parameters Influence on Workpiece Price Formation. Materials and Manufacturing Processes, v. 26, p. 550-557.

Mitchell, M. (1998). An introduction to genetic algorithms, First MIT Press paperback edition.

Mitchell, T. (1997). Machine Learning. McGraw-Hill, USA.

Pacheco MA. (1999). Algoritmos Genéticos: Princípios e Aplicações. In: V Congreso Internacional de Ingeniería Electrónica, Eléctrica y Sistemas, Lima, 11-16.

Santana, J. C. C., Araújo, S. A., Librantz, A. F. H., & Tambourgi, E. B. (2010). Optimization of Corn Malt Drying by Use of a Genetic Algorithm. Drying Technology, v. 28, p. 1236-1244.