

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Marco Aurélio Campetti da Silveira

**SELEÇÃO DE VARIÁVEIS PARA
CLUSTERIZAÇÃO COM VISTAS AO
APRIMORAMENTO DE PROCESSOS
PRODUTIVOS**

Porto Alegre

2013

Marco Aurélio Campetti da Silveira

Seleção de variáveis para clusterização com vistas ao aprimoramento de processos produtivos

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Profissional, na área de concentração em Sistemas de Produção.

Orientador: Prof. Michel José Anzanello, PhD.

Porto Alegre

2013

Marco Aurélio Campetti da Silveira

Seleção de variáveis para clusterização com vistas ao aprimoramento de processos produtivos

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Profissional e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof. Michel José Anzanello, PhD

Orientador PPGEP / UFRGS

Prof. José Luis Duarte Ribeiro, Dr.

Coordenador PPGEP / UFRGS

Banca Examinadora:

Prof. Danilo Marcondes Filho, Dr. (DEST/UFRGS)

Prof^a. Liane Werner, Dra. (PPGEP/UFRGS)

Profa. Márcia Echeveste, Dra. (PPGEP/UFRGS)

Dedicatória

Aos meus Avós, Etelvina e Henrique, Araci e
Armindo e ao Dindo Aldo, que lá de Cima
olham por nós.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos professores e colaboradores do Departamento de Engenharia de Produção e Transportes (DEPROT) da UFGRS, que desde a graduação participam e contribuem com a minha formação acadêmica e pessoal.

Agradeço ao meu orientador, Prof. Michel José Anzanello, pela contribuição, dedicação e paciência, fundamentais para conclusão de mais esta etapa da minha formação.

À Banca Examinadora, pela disponibilidade e avaliação da presente dissertação.

À minha família, em especial aos meus pais, Carmen e Paraguaçu, pelo incentivo e apoio incondicionais ao longo de minha vida inteira.

À família Pulvirenti, por me receber como a um filho, em especial à Lú, minha amada companheira de todas as horas.

Aos amigos desde sempre Cristiano Reis, Felipe Carvalho, Guilherme Garcia, Rafael Rimoli e Rodrigo Martin.

SILVEIRA, M. A. C. *Seleção de variáveis para clusterização com vistas ao aprimoramento de processos produtivos*. 2013. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul, Brasil.

RESUMO

A disputa por parcelas de mercado impõe condições severas às empresas sob diversas perspectivas. Dentre elas salienta-se a crescente demanda por alta variedade de produtos, que por sua vez cria um ambiente de decisões gerenciais complexas e por vezes conflitantes. Neste contexto, dois pontos relativos a processos produtivos tornam-se cada vez mais importantes na implantação de estratégias diferenciadas: a programação da produção e a gestão de estoques. Esta dissertação apresenta uma sistemática que visa embasar decisões relativas a tais pontos, aprimorando o processo produtivo. Como primeira etapa, trata-se o problema relativo à programação da produção diária. Para tanto, é apresentada uma sistemática de seleção de variáveis de clusterização para agrupamento de produtos, a qual é integrada à Simulação de Monte Carlo (SMC) com objetivo de maximizar lucro. Os cenários propostos são aplicados em clusters (famílias de produtos) e não nos produtos de forma individual, simplificando e agilizando a programação da produção. O erro percentual em relação à situação real foi de 1%. A segunda etapa desta dissertação foca na seleção de variáveis de clusterização com vistas à gestão de estoques. Desta forma, é apresentada uma abordagem de seleção de variáveis para clusterização de 76 produtos em três clusters, sendo que para cada cluster são geradas políticas simultâneas de reposição dos produtos. Tais políticas são confrontadas, em termos de custos de colocação de pedidos e guarda de estoques, com os resultados gerados pelo Lote Econômico de Compras (LEC). A redução do volume de pedidos anuais se aproximou de 90%, enquanto que o incremento de custos relativos à guarda de produtos e processamento de pedidos foi de 0,2% frente ao custo gerado pelo LEC.

Palavras-chave: seleção de variáveis, clusterização, simulação de Monte Carlo, gestão de estoques

SILVEIRA, M. A. C. *Clustering variable selection for production planning improvement*. 2013. Dissertation (Master in Industrial Engineering) - Federal University of Rio Grande do Sul, Brasil.

ABSTRACT

The dispute for larger market shares imposes hard conditions to companies in several perspectives. The growing demand for high variety of product models gives rise to complex productive scenarios, requiring precise managerial decisions. In this context, two points relating to production processes become increasingly important when implementing managerial strategies: production scheduling and inventory management. This dissertation presents an approach aimed at supporting decisions related to such points. As a first step, we tackle the daily scheduling problem presenting a systematic for selecting the most relevant variables for clustering products with similar features into groups; such groups are then integrated to a Monte Carlo Simulation (MCS) tailored to maximizing profit. In our propositions, managing clusters of products leads to simpler and faster managerial decisions regarding the production schedule. A proper training of the MCS parameters yielded a 1% deviation when compared to the real situation. The second part of this dissertation focuses on variable selection for clustering tailored to inventory management. For that matter, we present a variable selection approach for clustering 76 products into three clusters; such clusters are then integrated to a simultaneous inventory policy. The simultaneous policy aims at reducing costs of orders placement and simplifying the inventory management. When compared to the Economic Quantity Order (EOQ), our propositions reduced the number of order placements in 90%, while increasing costs related to inventory keeping in 0.2%.

Key words: variable selection, clustering, Monte Carlo Simulation, inventory management

LISTA DE FIGURAS

Figura 2.1 – Macro-fluxo sistemática proposta.....	25
Figura 2.2 – Processo Hierárquico de Clusterização: identificação de nº de clusters.....	30
Figura 2.3 – Evolução da qualidade de clusterização medida pelo IS.....	32
Figura 3.1 – Processo Hierárquico de Clusterização: identificação de nº de clusters.....	50
Figura 3.2 – IS’s gerados pelo procedimento “omita uma variável por vez”.....	50

LISTA DE TABELAS

Tabela 2.1 – Estrutura genérica das planilhas de simulação.....	28
Tabela 2.2 - Alocação dos produtos aos clusters (através de dendograma).....	30
Tabela 2.3 - Eliminação das variáveis tidas como inconsistentes pelos especialistas.....	31
Tabela 2.4 – Eliminação por coeficiente de variação.....	31
Tabela 2.5 – IS’s gerados pelo procedimento “omita uma variável por vez”.....	32
Tabela 2.6 – Alocação final produtos aos clusters.....	33
Tabela 2.7 – Simulação de demanda clusters.....	34
Tabela 2.8 – Comparação dos cenários produtivos simulados.....	35
Tabela 2.9 – Comparação simulação agrupada x individual.....	36
Tabela 2.10 – Comparação simulação agrupada x individual.....	37
Tabela 3.1 – Matriz descritiva de produtos e variáveis (parcial).....	49
Tabela 3.2 – Alocação dos produtos aos clusters.....	49
Tabela 3.3 – Frequências e custos derivados dos lotes econômico.....	51
Tabela 3.4 – Políticas de manutenção de estoques.....	52
Tabela 3.5 – Políticas Simultâneas de manutenção de estoques revisadas.....	53
Tabela 3.6 – Comparativo das Políticas Ótima, Simultânea e Simultânea Revisada.....	53

SUMÁRIO

1. INTRODUÇÃO	11
1.1 Considerações Iniciais	11
1.2 Justificativa do Tema.....	11
1.3 Objetivos.....	12
1.4 Procedimentos Metodológicos	12
1.5 Delimitações da Pesquisa	13
1.6 Estrutura da Dissertação	14
1.7 Referências	14
2. PRIMEIRO ARTIGO: SISTEMÁTICA PARA FACILITAÇÃO DA PROGRAMAÇÃO DE PRODUÇÃO COM BASE EM SIMULAÇÃO E SELEÇÃO DE VARIÁVEIS.....	16
2.1 Introdução	17
2.2 Referencial Teórico	18
2.3 Método.....	24
2.4 Resultados e Discussão.....	28
2.5 Conclusão	36
2.6 Referências	37
3. SEGUNDO ARTIGO: SIMPLIFICAÇÃO DA GESTÃO DE ESTOQUES COM BASE EM SELEÇÃO DE VARIÁVEIS PARA CLUSTERIZAÇÃO	40
3.1 Introdução	41
3.2 Referencial teórico.....	42
3.3 Método de Trabalho.....	46
3.4 Resultados e Discussões	48
3.5 Conclusão	53
3.6 Referências	54
4. CONSIDERAÇÕES FINAIS.....	56

1. Introdução

1.1 Considerações Iniciais

Diversos segmentos do mercado caracterizam-se pela ampla e variada oferta de produtos, visando primordialmente uma maior participação de mercado e a fidelização de clientes (KIM e MAUBORGNE, 2005). A opção por determinado produto, nos dias atuais, apoia-se em mais dimensões do que simplesmente preço ou qualidade de um produto (COSTA, 2007).

Segundo Ehrbar (2004) e Catelli (2010), empresas são organizações ou sistemas econômicos que visam atender as necessidades de terceiros, sendo remuneradas por isto e obtendo uma parcela de ganho na forma de lucro que possibilita a continuidade da mesma. Segundo Costa (2007), oferecer a variedade de produtos desejada pelo cliente gera incertezas e complexidades por vezes conflitantes com a premissa lançada por Ehrbar (2004) e Catelli (2010). Chopra e Meindl (2003) sinalizam que parte da incerteza advém das decisões acerca da utilização de recursos produtivos, enquanto que Wanke (2005) complementa dizendo que a precisa gestão de estoques pode contribuir para o êxito da organização ou acentuar as incertezas. Neste contexto fica evidente a necessidade de as empresas conhecerem, detalhadamente, tanto a si próprias como ao mercado no qual estão inseridas (RAFAELI, 2009), buscando ferramentas eficientes na gestão de seus processos.

Uma forma eficiente de reduzir a complexidade na gestão de recursos e insumos decorrente da ampla variedade de produtos ofertados é a formação de agrupamentos de produtos em famílias com características semelhantes (GUYON e ELISSEEFF, 2003; HAIR JR. et al., 2003; KASHEF e KAMEL, 2009). Tais agrupamentos atuam como facilitadores na gestão dos produtos, visto que decisões acerca de um grupo agilizam o processo decisório e reduzem a incidência de erros de gestão (CHOPRA e MEINDL, 2003). Como pressuposto para a correta formação de famílias de produtos, surge a questão relativa à escolha das variáveis mais relevantes para a formação de grupos de produtos consistentes.

1.2 Justificativa do Tema

Em cenários produtivos caracterizados pela necessidade de ofertar-se grande variedade de produtos, é primordial dispor-se de políticas precisas acerca do uso de recursos produtivos e gestão de estoques. De acordo com Wanke (2005, 2010), os estoques podem se tornar peça fundamental para a execução de uma estratégia diferenciada frente aos concorrentes. A má gestão de estoques e recursos produtivos, além de acarretar perdas financeiras significativas,

pode macular a imagem de uma empresa frente aos clientes, impondo perdas de longo prazo de difícil mensuração.

No contexto de amplo espectro de modelos de produtos, Ritzman e Krajewski (2003) e Christopher (2011) sinalizam como oportuno agrupar itens por semelhança e gerir a programação da produção e estoques baseando-se em tais grupos. A adoção de grupos de produtos com características similares permite simplificar a gestão de inventários e recursos, visto que decisões tomadas para um produto valem para os demais itens daquele grupo. Além disso, tais agrupamentos permitem que menor tempo seja devotado à gestão de itens de menor relevância, visto que as decisões estão vinculadas ao grupo a que aquele produto está inserido.

A coerente formação de grupos de produtos com características semelhantes, no entanto, depende da identificação das variáveis de clusterização detentoras de maior poder discriminatório, visto que a utilização de variáveis ruidosas e irrelevantes tende a agrupar produtos de maneira equivocada. De tal forma, justifica-se a iniciativa de desenvolver sistemáticas simples e eficientes para a seleção das variáveis de clusterização mais relevantes; tais grupos são então integrados a abordagens para controle de recursos produtivos e inventários, simplificando o processo de gestão empresarial.

1.3 Objetivos

O objetivo desta dissertação é propor uma sistemática de seleção de variáveis com vistas à clusterização de produtos de forma a aprimorar processos produtivos, sob a ótica da programação da produção e da gestão de estoques. Como premissas da sistemática considera-se a simplicidade e eficiência de aplicação.

Como objetivos específicos, listam-se:

- Utilizar a simulação de Monte Carlo para análise de cenários produtivos visando identificar um mix de produção que maximize a receita e minimize perdas por escassez;
- Aplicar uma política simultânea de gestão de estoques com vistas à simplificação do processo de colocação de pedidos;
- Aplicar a sistemática proposta em uma empresa do setor alimentício.

1.4 Procedimentos Metodológicos

O presente trabalho se trata de uma pesquisa aplicada de abordagem quantitativa, com procedimentos experimentais e objetivos descritivos (GIL, 1991; YIN, 2001). Trata-se de

pesquisa aplicada, pois formula uma sistemática focada na solução de problema específico. Apresenta abordagem quantitativa, pois apoia-se em ferramentas de simulação, seleção de variáveis e técnicas multivariadas de análise de dados. Por se tratar de um estudo com vistas à melhoria da programação da produção e gestão estoques, apresenta procedimentos experimentais. Como são pesquisadas e analisadas relações entre observações, pode-se afirmar que apresenta objetivo descritivo.

O primeiro artigo da presente dissertação apresenta e avalia uma sistemática de simulação de Monte Carlo aplicada a produtos clusterizados, com intuito de facilitar e agilizar a programação de produção diária assertivamente. Inicialmente, estima-se a quantidade de famílias (clusters, k) ideal através de procedimento hierárquico de clusterização (HAIR JR. et al., 2003). Com intuito de reduzir a necessidade de processamento computacional, uma pré-seleção de variáveis é realizada com base na opinião de especialistas e indicadores de dispersão (GUYON e ELISSEEFF, 2003). Na sequência, utiliza-se um procedimento iterativo de seleção de variáveis do tipo “omite uma variável por iteração” associado à clusterização não-hierárquica do tipo *k-means*, com a qualidade dos agrupamentos gerados medida através do Índice Silhouette (IS) (ANZANELLO e FOGLIATTO, 2009). Por fim cenários produtivos são analisados para os agrupamentos através de Simulação de Monte Carlo (SMC).

No segundo artigo é apresentada uma sistemática de gestão de estoques baseada em famílias de produtos (WANKE, 2005). Tais famílias são geradas através da seleção das variáveis mais relevantes para clusterização. Nesta etapa a seleção de variáveis é realizada pelo procedimento iterativo “omite uma variável por iteração”. A qualidade dos agrupamentos resultantes é avaliada através do IS. Por fim, são geradas políticas Simultâneas de gestão de estoques, as quais são comparadas à política de gestão ótima gerada através do Lote Econômico de Compras (LEC) (RITZMAN e KRAJEWSKI, 2003; CHRISTOPHER, 2011). A comparação apoia-se na análise dos custos de guarda e de estoques oriundos das duas políticas.

1.5 Delimitações da Pesquisa

A presente dissertação não aborda alterações ou melhorias em algoritmos de clusterização, utilizando métodos existentes para a formação de grupos.

São utilizados dados e informações relacionados a custos e preços de materiais. Tais dados e informações são utilizados apenas como parâmetros de entrada (originários de bases já existentes) ou saída (aplicação de dados de entradas em equações detalhadas adiante).

Técnicas de custeios e levantamento de custo ou precificação não são abordadas na sistemática.

1.6 Estrutura da Dissertação

A dissertação é composta por quatro capítulos. No primeiro são apresentadas as considerações iniciais, justificativa do tema, objetivos, procedimentos metodológicos, delimitações da pesquisa e estrutura da dissertação.

O segundo capítulo traz o primeiro artigo. Neste é apresentada a aplicação de uma sistemática de seleção de variáveis para clusterização associada à Simulação de Monte Carlo, com vistas à facilitação da programação da produção com base em cenários simulados. A sistemática é avaliada em termos da minimização dos índices de escassez e maximização dos lucros; por fim, são sugeridos desdobramentos futuros.

No terceiro capítulo é apresentado o segundo artigo, o qual traz uma sistemática de clusterização de produtos baseada em procedimento iterativo de seleção de variáveis. Os grupos formados são integrados à política Simultânea de reposição de estoques. Os resultados alcançados através de política Simultânea são comparados com os oriundos do Lote Econômico de Compras (LEC).

No quarto e último capítulo são apresentadas as conclusões. Neste capítulo também são abordados possíveis desdobramentos desta pesquisa.

1.7 Referências

- ANZANELLO, M. J.; FOGLIATTO, F. S.** Selecting the best variables for grouping mass-customized products involving worker's learning. *Int. J. Production Economics* 130 (2011) 268–276, 2011.
- ANZANELLO, M. J.** Seleção de variáveis com vistas à classificação de bateladas de produção em duas classes. *Gestão e Produção*, São Carlos, v. 16, n. 4, p. 526-533, out.-dez. 2009.
- CATELLI, A.** Controladoria: Uma abordagem da Gestão Econômica – GECON. São Paulo. Atlas, 2010.
- CHOPRA, S.; MEINDL, P.** Gerenciamento da cadeia de suprimentos: estratégia, planejamento e operação. São Paulo. Prentice Hall, 2003
- CHRISTOPHER, M.** Logística e gerenciamento da cadeia de suprimentos. São Paulo. Cengage Learning, 2011
- COSTA, F. J.** A influência do valor percebido pelo cliente sobre comportamentos de reclamação e boca a boca. Tese (Doutorado em Administração de Empresas). Fundação Getúlio Vargas - EAESP. 2007
- EHRBAR, A.** EVA: Valor Econômico Agregado: A Verdadeira Chave para a CRIAÇÃO DE RIQUEZA. Rio de Janeiro. Qualitymark, 1999.

GIL, A. C. Como elaborar projetos de pesquisa. 3.ed. São Paulo: Atlas, 1991.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3 (2003) 1157-1182, 2003.

HAIR JR., J. F.; ANDERSON, R.E.; TATHAM, R.L.; & BLACK, W.C. Análise multivariada de dados. Prentice-Hall. São Paulo. 2003.

KASHEF, R.; KAMEL, M. S. Cooperative clustering. *Pattern Recognition* 43 (2010) 2315–2329. Disponível em: www.elsevier.com/locate/pr

KIM, W. C. MAUBORGNE, R. A estratégia do Ocenô Azul. São Paulo. Campus, 2005.

NAGATANI, T.; OZAWA, S.; ABE, S. Fast variable selection by block addition and block deletion. *Journal of Intelligent Learning Systems and Applications*, 2010, 2, 200-211.

RAFAELLI, L. Análise envoltória de dados como ferramenta para avaliação de desempenho relativo. Dissertação (Mestrado em Engenharia). Universidade Federal do Rio Grande do Sul. 2009.

STEINLEY, D.; BRUSCO, M. A new variable weighting and selection procedure for K-means cluster analysis. *Multivariate Behavioral Research* 43 (1), 77–108, 2008.

YIN, R. K. Estudo de Caso: planejamento e métodos. 2.ed. Porto Alegre: Bookman, 2001.

WANKE, P. F. Metodologia para gestão de estoques de peças de reposição: um estudo de caso em empresa brasileira. *Revista Tecnológica*. Centro de Estudos em Logística, Coppead, UFRJ. Dez., 2005.

WANKE, P. F. The impact of different demand allocation rules on total stock levels. *Revista Pesquisa Operacional*, v. 30, n.1, p. 33-52, Jan/Abr 2010.

2 PRIMEIRO ARTIGO

SISTEMÁTICA PARA FACILITAÇÃO DA PROGRAMAÇÃO DE PRODUÇÃO COM BASE EM SIMULAÇÃO E SELEÇÃO DE VARIÁVEIS

Resumo

Este artigo apresenta uma sistemática que combina seleção de variáveis para clusterização e Simulação de Monte Carlo com objetivo de agilizar e facilitar a programação da produção em cenários multiprodutos. A sistemática é composta por quatro etapas: (i) estruturação e validação de dados, (ii) seleção de variáveis para clusterização, (iii) simulação de demanda através de Simulação de Monte Carlo, e (iv) análise de resultados e aderência da sistemática à realidade. A sistemática proposta foi aplicada em uma empresa do segmento alimentício com grande variedade de produtos ofertados, permitindo simplificar a programação da produção e maximizar os lucros decorrentes da comercialização dos produtos.

Palavras-chaves: Seleção de variáveis, clusterização, Simulação de Monte Carlo

Abstract

This paper proposes an approach that combines variable selection for clustering and Monte Carlo simulation in order to simplify the production scheduling in multiproduct scenarios. The system consists of four steps: (i) data collection, (ii) variable selection for clustering products with similar productive and physical features, (iii) simulation of demand using Monte Carlo simulation, and (iv) analysis of results and adherence of systematic to real scenarios. The methodology proposed was applied in a food industry with a wide variety of products, making production planning simpler and maximizing profits.

Keywords: *Variable Selection, clustering, Monte Carlo Simulation*

2.1 Introdução

A disputa pela preferência do cliente não depende somente de preço ou qualidade de um produto, mas se baseia em um série de outras dimensões que caracterizam tal produto. Segundo Kim e Mauborgne (2005), em muitos segmentos, tais dimensões são apenas qualificadoras, isto é, apenas permitem que o produto seja percebido pelo consumidor, mas não necessariamente escolhido. Costa (2007) reforça que, para o cliente, o conceito de valor vai muito além de qualidade e preço justo. A agregação de diversas dimensões no momento da escolha cria um cenário complexo e de incerteza para as empresas. Esta incerteza, em parte, é gerada pela variedade de produtos ou famílias de produtos ofertadas, que demanda esforços extras da área de operações (em particular, do planejamento de produção).

Desta forma, torna-se cada vez mais necessário às empresas conhecerem detalhadamente o mercado, bem como a si próprias (RAFAELI, 2009). A variedade de produtos necessária para atender o espectro de consumidores alvo traz à tona o problema relativo à utilização dos recursos produtivos, matérias-primas e insumos, o que normalmente vem à tona quando a capacidade produtiva apresenta um limite físico. No caso de empresas que têm elevada gama de linhas de produtos, faz-se necessária priorização de determinado produto em detrimento a outro (CHOPRA e MEINDL, 2003), o que pode ser facilitado através da formação de grupos com características semelhantes e alocação das famílias resultantes aos recursos disponíveis. A geração de tais famílias usualmente se vale de variáveis que descrevem características físicas e necessidades de processamento dos diversos modelos de produto.

Neste contexto, a identificação das variáveis mais relevantes para a inserção dos produtos em famílias de acordo com suas demandas produtivas e características físicas aparece como fundamental para aprimorar a programação da produção. Conforme Hair Jr et al. (2003), a utilização da totalidade de variáveis disponíveis em procedimentos de clusterização tende a reduzir a qualidade dos grupos formados, visto que variáveis ruidosas e irrelevantes comprometem a eficiência dos algoritmos de clusterização, conduzindo a alocações equivocadas. Para tanto, a utilização de uma sistemática estruturada de identificação das variáveis mais relevantes com vistas à formação de grupos viabiliza maior assertividade na programação da produção. Tais agrupamentos podem então ser integrados a outras ferramentas (por exemplo, simulação de eventos discretos) que viabilizem a geração de cenários voltados à maximização do desempenho dos recursos produtivos.

Este artigo propõe uma sistemática de seleção de variáveis com vistas à formação de famílias

de produtos com demandas produtivas similares, a qual é integrada a uma ferramenta de simulação com o intuito de facilitar e agilizar a programação da produção diária. A primeira etapa é constituída por um procedimento de seleção do tipo “omite uma variável por iteração” em conjunto com um índice para avaliação da qualidade do agrupamento gerado, o Índice Silhouette. À cada variável omitida, a qualidade do agrupamento é avaliada; a variável que, quando omitida, conduzir ao maior Índice Silhouette, é eliminada (visto que tal variável é a que menos contribui na formação de agrupamentos precisos). O número de clusters é inicialmente estimado através de clusterização hierárquica (apoiada em dendograma), e a efetiva inserção dos produtos às famílias é realizada através do algoritmo *k-means* (ferramenta de clusterização do tipo não-hierárquica) (HAIR JR. et al., 2003). Os grupos gerados pelas variáveis selecionadas são então analisados em cenários produtivos diversos através da simulação de Monte Carlo.

Além da presente introdução, o artigo é composto por um referencial teórico acerca dos fundamentos de clusterização, seleção de variáveis e simulação de monte Carlo na seção 2.2. O método é apresentado na seção 2.3, sendo seus resultados descritos na seção 2.4. A seção 2.5 traz as conclusões do estudo e sugestões de trabalhos futuros.

2.2 Referencial Teórico

Essa seção traz os fundamentos da ferramenta de clusterização, métodos para seleção de variáveis e simulação de Monte Carlo.

2.2.1 Clusterização

Clusterização é o processo de alocação de objetos com características similares em grupos, de tal forma que objetos alocados a outros grupos apresentem características distintas. Tal similaridade é usualmente mensurada através de métricas apropriadas, destacando-se as medidas de distância entre as observações (MIMAROGLU; ERDIL, 2010; SANTHISREE; DAMODARAM, 2011).

Em diversos cenários produtivos e gerenciais, é mais oportuno gerir produtos e recursos através da definição de grupos homogêneos. A construção destes grupos requer ferramentas apropriadas para garantir a similaridade entre os objetos agrupados, dentre as quais destaca-se a clusterização (HAIR JR. et al., 2003, RODRIGUES e SELLITTO, 2009). Cheznian et al. (2011) corroboram a utilidade do processo de clusterização com vistas à identificação de similaridade ou diferenças entre as observações analisadas.

Segundo Kashef e Kamel (2009), a clusterização permite a abstração e interpretação de

grandes quantidades de dados pela construção de um significado comum não aparente para cada grupo ou cluster. Apesar de não aparente, Jain (2010) salienta que a meta da técnica é revelar o agrupamento natural que existe em uma série de dados. Anzanello e Fogliatto (2011) afirmam que, por vezes, a alocação de recursos em cenários produtivos é mais eficiente através do agrupamento, visto que determinadas características dos agrupamentos formados demandam abordagens específicas.

Segundo Hair Jr. et al. (2003), existem dois procedimentos tradicionais de clusterização, os hierárquicos e os não hierárquicos. A diferença entre os métodos está na forma como as observações são alocadas aos grupos. Os procedimentos hierárquicos constroem os agrupamentos através de árvore hierárquica (dendograma), avaliando progressivamente a similaridade entre os grupos e observações. Os procedimentos não-hierárquicos, por sua vez, alocam observações em uma único movimento baseado nas distâncias entre as observações.

Os métodos hierárquicos apresentam relações de hierarquia entre agrupamentos formados em estágios subsequentes, isto é, os resultados de um estágio anterior de agrupamento são considerados no estágio seguinte (SANTHISREE e DAMODARAM, 2011). Dividem-se em dois grupos: os aglomerativos e os divisivos, diferenciando-se pela sequência de execução. Enquanto procedimentos aglomerativos consideram cada observação como um agrupamento individual, os divisivos consideram um único agrupamento contendo todas as observações. À medida que o procedimento é executado, os aglomerativos diminuem o número de clusters, através de agrupamento por maior semelhança. Já nos divisivos, o processo é inverso: a partir do aglomerado inicial, observações são extraídas por critérios de diferença, formando agrupamentos menores e mais homogêneos (HAIR JR. et al, 2003; SANTHISREE e DAMODARAM, 2011).

Nos procedimentos não hierárquicos, o ponto de partida é a definição do número de agrupamentos a serem gerados (k , número de clusters). Um número k de “sementes randômicas” é então gerado, e distâncias Euclidianas entre tais sementes e as observações a serem alocadas são estimadas. Distância Euclidiana é o tamanho do segmento de reta entre dois objetos (HAIR JR. et al., 2003), que pode ser calculado através da Equação (1), na qual os termos compostos pelas coordenadas x e y dos pontos. O método busca, aleatoriamente e iterativamente, distribuir as observações a k grupos, de forma que a distância total entre os dados de um grupo e o seu respectivo centroide, somada por todos os grupos, seja minimizada (LIU et al., 2008). Segundo Hair Jr. et al. (2003) e Liu et al. (2008), o centroide é valor gerado pela média das observações contidas no grupo. Ao iniciar o procedimento, são geradas

aleatoriamente k sementes, isto é, grupos a partir dos quais são calculados valores médios para cada grupo (centroides). Como próximo passo, as observações são realocadas aos k grupos de acordo com maior proximidade aos centroides iniciais, que na sequência são recalculados. Este processo iterativo acontece até que as realocações não sejam mais necessárias, gerando o valor mínimo na função objetivo.

$$D_{\text{Euclidiana}} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

De forma geral, os métodos não hierárquicos apresentam melhores resultados em relação a dados atípicos e desempenho de clusterização do que os métodos hierárquicos, porém tais resultados estão diretamente ligados à qualidade dos dados analisados e ao número de agrupamentos a ser gerado. Deve-se salientar que cada cenário gerado é independente dos outros, e deriva somente do número de clusters definido inicialmente (HAIR JR. et al., 2003; SANTHISREE e DAMODARAM, 2011).

Uma prática recomendada na formação de agrupamentos é a combinação de procedimentos, fazendo uso das virtudes de cada um dos métodos acima descritos. Normalmente, inicia-se a busca pela identificação da quantidade ideal de clusters valendo-se de um procedimento hierárquico. Uma vez definido o número de clusters a ser considerado, executa-se o processo de agrupamento através do procedimento não hierárquico. Esta combinação aprimora o procedimento de clusterização (HAIR JR. et al, 2003). Aplicações recentes de ferramentas de clusterização podem ser encontradas em sistemas produtivos, estudos de marketing, estudos biológicos e filtragem de emails do tipo spam (LIU et al. 2008; ANZANELLO e FOGLIATTO, 2011; MOHAMMAD, 2011; CAI E SUN, 2011; MIMAROGLU e ERDIL, 2010; CHEZHIAN, SUBASH e SAMY, 2011).

Concluído o processo de clusterização, é fundamental avaliar a qualidade do agrupamento gerado (HAIR JR. et al., 2003). Uma forma de medição consiste no Índice Silhouette (IS) (MIMAROGLU e ERDIL, 2010; ANZANELLO e FOGLIATTO, 2011), o qual oscila entre $[-1; +1]$; quanto mais próximo de $+1$, maior é a qualidade do agrupamento. Observações com índice próximo a $+1$ indicam adequada alocação ao cluster de tais observações, visto que as mesmas estão distantes dos outros clusters; valores próximos de -1 indicam que a observação foi alocada ao cluster errado. O IS é calculado através da Equação (2), composta por dois elementos: um referente a distância média entre a observação e demais observações alocadas ao mesmo cluster $a(j)$, e outro referente à distância média entre a observação em questão e as observações no agrupamento mais próximo $b(j)$.

$$IS(j) = \frac{b(j) - a(j)}{\text{Max}\{b(j), a(j)\}} \quad (2)$$

Cabe ressaltar que, a escolha das variáveis utilizadas na formação dos agrupamentos deve estar diretamente relacionada aos objetivos da análise (NAVEIRO e PEREIRA FILHO, 1991; MIMAROGLU e ERDIL, 2010). Anzanello e Fogliatto (2011) e Mohammad (2011) recomendam parcimônia na escolha das variáveis de agrupamento, visto que a inserção de variáveis irrelevantes no procedimento pode comprometer a qualidade dos agrupamentos gerados. Abordagens para seleção de variáveis são agora apresentadas.

2.2.2 Seleção de variáveis em Clusterização

Segundo Anzanello (2009), a seleção de variáveis de processo é fundamental para controle de processos produtivos, assim como para caracterização correta de produtos. Da mesma forma, Steiner et al. (2008) sinalizam que grande dificuldade na avaliação de bens se deve à heterogeneidade de variáveis e atributos que os caracterizam. A eficácia na seleção de variáveis para caracterização dos objetos responde por importante parcela do sucesso no agrupamento formado. Desta forma, os objetivos da clusterização não podem ser separados da seleção de variáveis (HAIR JR. et al., 2003), uma vez que a escolha de uma variável significa selecionar uma dimensão específica do objeto em estudo em detrimento a outras (SENRA et al., 2007). Villanueva (2006) define seleção de variáveis como a identificação de subconjunto de variáveis que levam a resultados satisfatórios no reconhecimento de padrões em bases de dados consistindo de elevado volume de informações.

Guyon e Elisseeff (2003) sinalizam que há duas abordagens clássicas para os métodos de seleção de variáveis, *Forward Selection* e *Backward Elimination*. O primeiro parte da incorporação progressiva das variáveis ao modelo; já a sistemática *Backward* consiste em um método regressivo, partindo do grupo total de variáveis candidatas, e então eliminando sequencialmente aquelas tidas como irrelevantes. Segundo Anzanello e Fogliatto (2011), ambos os procedimentos de seleção de variáveis têm provado eficiência e estabilidade. Uma variação das abordagens acima, o *Stepwise*, baseia-se na inserção e remoção alternada das variáveis ao modelo de acordo com a contribuição das mesmas para desempenho do agrupamento.

Anzanello e Fogliatto (2011) ressaltam que variáveis de natureza qualitativa requerem atenção adicional no que tange a seus escores ou pesos, pois podem comprometer a precisão da representação das observações. Por se tratarem de variáveis qualitativas, logo não apresentam cardinalidade natural, faz-se necessário atribuir tais pesos às variáveis. De acordo com a

ponderação resultante entre os pesos, diferentes variáveis podem ficar em evidência em um mesmo cenário. Como alternativa para este inconveniente, pode-se usar mão da premissa de que variáveis com grandes variações apresentam maior poder de clusterização (STEINLEY e BRUSCO, 2008), requerendo mesmo nível ou maior de atenção já sugerida por Anzanello e Fogliatto (2011).

Dentro do campo de seleção de variáveis, são possíveis duas abordagens: filtragem e envoltória (NAGATANI et al., 2010). Guyon e Elisseeff (2003) incluem uma terceira (*wrapper*), semelhante à envoltória, porém com menor necessidade de recursos computacionais. Na abordagem de filtragem, a ideia central é uma pré-seleção e exclusão de variáveis irrelevantes, segundo critérios definidos pelo usuário. Já nas abordagens envoltória e embutida a escolha das variáveis faz parte de um algoritmo de aprendizagem que demanda recursos computacionais. Este algoritmo usa de uma taxa de reconhecimento que busca uma característica preditora fornecida pelo usuário. Esta busca tem por objetivo encontrar o menor subgrupo de variáveis que melhor caracteriza o conjunto geral de dados de acordo com a característica preditora (GUYON e ELISSEEFF, 2003; VILLANUEVA, 2006; NAGATANI, 2010; HORTA e ALVES, 2012).

Em termos de desempenho, os envoltórios apresentam capacidade de generalização maior, mas a um custo maior. Por outro lado, os métodos de filtragem têm custos menores e maior facilidade de operacionalização, podendo comprometer o desempenho da seleção resultante (ABE, OZAWA e NAGATANI, 2010). Guyon e Elisseeff (2003) e Villanueva (2006) sugerem um uso misto de abordagens de seleção, isto é, fazer uso de metodologia de filtragem como pré-processamento e então uso de técnica envoltória ou embutida. A lógica desta estratégia é simples, eliminar origens de ruído e de *overfitting* através de filtragem, para então usar mecanismo de melhor desempenho, envoltório.

Através de uso de metodologia de seleção de variáveis, Costa Filho e Poppi (2001) constataram significativa melhora nos resultados em modelos multivariados. Nagatani et al. (2010) atribuem melhoria de desempenho em subconjunto selecionados à redução de complexidade dos modelos gerada pela escolha correta das variáveis mais relevantes. Entretanto, é difícil medir o desempenho de todos subconjuntos de variáveis possíveis (ABE et al., 2010). No mesmo sentido, Senra et al. (2007) propõem a necessidade de uma análise prévia por parte de especialistas das variáveis disponíveis, antes mesmo de definição do método.

2.2.3 Simulação de eventos discretos

A simulação é a estruturação de um modelo que visa representar uma operação ou situação do mundo real. Este modelo utiliza diversos parâmetros, detalhando o sistema em análise com determinada fidelidade. O intuito destas técnicas é suportar decisões quando a realização de pilotos ou testes reais é inviável, seja por questões de segurança, financeira, recursos tecnológicos ou temporais (AMANIFARD et al., 2011). Entretanto, a qualidade das análises geradas por modelos simulados, assim como seus resultados, está diretamente ligada à qualidade dos dados de entrada e estruturação do modelo (CATELLI, 2010; SARAIVA JÚNIOR et al., 2010).

Dentre as técnicas de simulação disponíveis na literatura, destaca-se a Simulação de Monte Carlo (SMC). Esta técnica é baseada na geração de números aleatórios e probabilidade de ocorrência de valores associados ao fenômeno em análise. Em casos de difícil modelagem ou formulação, dados de entrada podem ser substituídos e representados por padrões estatísticos, sobre os quais a SMC é aplicada (ZAPATA et al., 2004; SARAIVA JÚNIOR, RODRIGUES e COSTA 2010).

A SMC é operacionalizada como segue. Através de um processo iterativo, são gerados, aleatória e sucessivamente, N valores de uma variável de entrada específica. Os valores desta variável de entrada são então aplicados ao modelo em análise. Os valores finais obtidos no modelo analisado, pela incidência dos valores gerados para variável de entrada, resultam em uma distribuição de probabilidade com média e desvio padrão de ocorrências para o evento (modelo) estudado (ZAPATA et al., 2004). Variáveis aleatórias são então geradas e rebatidas contra a função de distribuição acumulada. Tal conversão é repetida por um número elevado de vezes, de forma que os valores gerados possam representar a frequência de ocorrência do fenômeno em análise. Os dados gerados são então inseridos na modelagem de interesse, e cenários alternativos são avaliados de acordo com o propósito da análise. Cabrer et al. (2007) enfatizam a necessidade comprovação de que as variáveis aleatórias, isto é, dados de entrada, são independentes. Em termos práticos, os resultados obtidos para uma variável aleatória não devem condicionar/influenciar ou ser condicionados/influenciados pelos resultados de outras variáveis aleatórias. Faz-se necessário, também, conhecer precisamente as distribuições de probabilidade dos dados de entrada do sistema modelado (ZAPATA et al., 2004).

Dentre as vantagens da SMC, destaca-se que os dados de entrada podem apresentar qualquer tipo de distribuição, pode-se analisar cenários de forma ágil (alterando somente dados de

entrada) e, devido a não existência de um algoritmo único para SMC, pode-se ajustar o procedimento de simulação da forma mais oportuna à situação. Complementarmente, Pamplona e Silva (2005) e Saraiva Júnior et al. (2010) entendem a SMC como um método numérico estocástico universal para a solução de problemas matemáticos, propondo sua utilização ao gerenciamento de riscos.

A SMC vem sendo usada nas mais diversas áreas. Zapata et al. (2004) utilizaram a técnica para avaliar a confiabilidade de um sistema de distribuição de energia elétrica. Através deste, foi possível constatar a versatilidade e flexibilidade do sistema. Já Saraiva Júnior et al. (2010) constataram a utilidade da SMC ao utilizá-la no auxílio à definição de mix de produção de plásticos. Através da simulação foram feitas análises econômico-financeiras relativas à utilização de recursos produtivos, as quais viabilizaram definir a melhor política de mix de produtos. Como ponto forte do método, além de flexibilidade e simplicidade de aplicação, os autores ressaltam a possibilidade de utilização de conceitos de áreas de conhecimento distintas de forma integrada.

2.3 Método

Segundo Yin (2001) e Gil (1991), este artigo constitui-se em uma pesquisa aplicada, visto que é orientado à solução de um problema específico. Do ponto de vista da abordagem, é um trabalho caracterizado como quantitativo, visto que vale-se de ferramentas multivariadas para seleção de variáveis e simulação. Quanto aos objetivos do estudo, o mesmo apresenta-se como descritivo, já que são pesquisadas as características e relações entre variáveis selecionadas. Em relação ao procedimento adotado, o que melhor descreve é pesquisa experimental, uma vez que o objetivo geral do estudo é a otimização do processo de programação da produção. Para isto são avaliadas e quantificadas as variáveis que têm influência sobre a caracterização do processo.

A metodologia proposta é dividida em quatro etapas: (i) coleta de dados, (ii) seleção de variáveis para clusterização de observações em grupos, (iii) simulação baseada nos grupos gerados, e (iv) análise dos resultados obtidos. Tais etapas são apresentadas na Figura 1 e detalhadas na sequência.

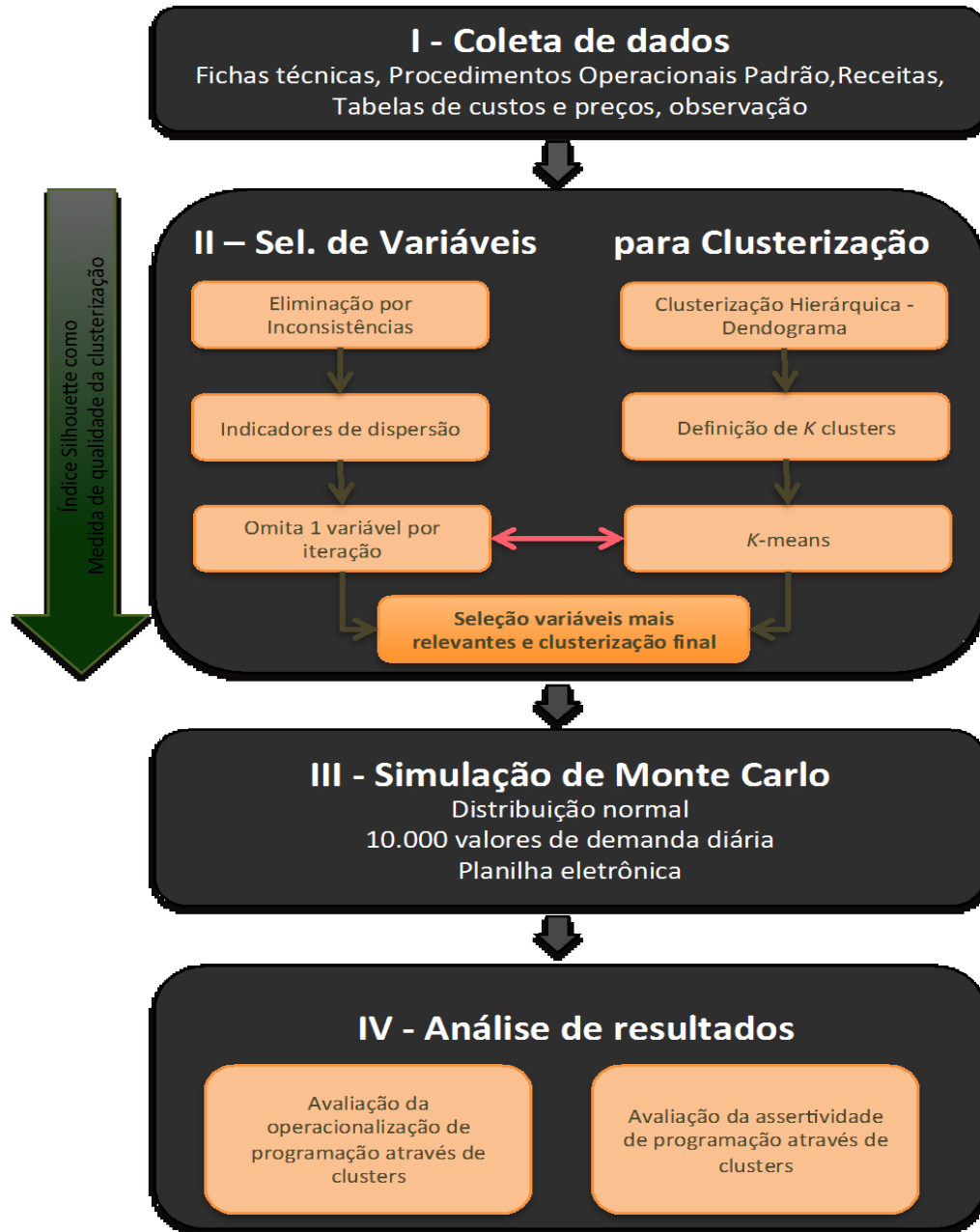


Tabela 2.1 – Macro-fluxo sistemática proposta

2.3.1 Coleta de dados

Na primeira etapa, deve-se coletar dados que descrevam características gerenciais e operacionais dos modelos de produtos analisados. Tais variáveis permitem a formação de grupos de produtos com necessidades de processamento similares. Exemplos incluem variáveis associadas à forma de consumo do produto em análise, estocagem, características de ordem financeira, processos de preparo, recursos operacionais necessários e tipo de produto, entre outras. Tais variáveis podem ser coletadas de bancos de dados históricos da empresa analisada e opiniões de especialistas.

Para dados de ordem financeira, processos de preparo, recursos operacionais e estocagem, os dados são coletados de procedimentos operacionais padrão (POP) e tabelas de custo padrão. As dimensões de consumo e tipo de produto são extraídas diretamente da análise de propriedades físicas dos produtos.

2.3.2 Seleção de variáveis para clusterização

Esta etapa é subdividida em três passos. O primeiro aplica um procedimento hierárquico de clusterização sobre os dados com intuito de estimar o número adequado de clusters. O segundo refere-se a uma pré-seleção de variáveis através de indicadores de variabilidade e opiniões de especialistas, com objetivo de reduzir o número de variáveis candidatas. Por fim, ocorre a etapa de seleção de variáveis para clusterização definitiva, utilizando uma abordagem do tipo “omita uma variável por iteração”. Esses passos são agora detalhados.

Inicialmente, identifica-se o número recomendado de clusters, k , a serem formados através de um dendograma (ferramenta típica em procedimentos hierárquicos de clusterização), no qual visualiza-se agrupamentos progressivos das observações (HAIR JR. et al., 2003). Tal valor é utilizado como parâmetro de entrada na clusterização não-hierárquica *k-means* (HAIR JR. et al., 2003; SANTHISREE e DAMODARAM, 2011). Na sequência, agrupa-se as observações utilizando-se todas as variáveis através do algoritmo *k-means*. A qualidade da clusterização gerada é medida através do Índice Silhouette (IS) (ANZANELLO e FOGLIATTO, 2011), o qual será utilizado como valor de referência para avaliar aprimoramentos nos procedimentos de clusterização decorrentes da seleção de variáveis.

Como próximo passo, realiza-se uma pré-seleção de variáveis. São dois os objetivos deste passo: o primeiro consiste na redução do número de variáveis que serão investigadas e, conseqüentemente, no número de iterações realizadas nos passos seguintes. O segundo é mais sutil, porém, de acordo com qualidade dos dados coletados, pode se tornar fundamental, visto que avalia a consistência e qualidade dos dados e variáveis candidatas. Tal seleção pode ser feita de duas formas: através de opinião de especialistas quanto à consistência das variáveis candidatas, ou utilizando-se indicadores de variabilidade das variáveis coletadas. Tais técnicas podem ser realizadas em conjunto ou individualmente.

O indicador de variabilidade utilizado neste trabalho é o coeficiente de variação, calculado pela razão entre desvio padrão e média, conforme sugerem Guyon e Elisseeff (2003) e Stanley e Brusco (2008). Os mesmos autores propõem que outras medidas, como amplitude e variância, podem indicar o poder de clusterização de uma variável. O princípio é simples:

variáveis que apresentam os maiores valores de amplitude, variância, desvio-padrão e coeficiente de variação tendem a ter melhor desempenho de clusterização. À cada eliminação de variáveis do grupo de candidatas, é repetida a clusterização através do *k-means* e calculado o IS médio para o agrupamento gerado. Este procedimento é executado até que o IS resultante do agrupamento seja inferior ao anteriormente calculado; nesse instante, inicia-se um procedimento exaustivo de seleção das melhores variáveis remanescentes através do procedimento “omite uma variável por iteração”.

O procedimento “omite uma variável por iteração” visa identificar o menor conjunto possível de variáveis relevantes para a formação dos grupos de produtos, sem perder qualidade no agrupamento. Nesse procedimento, uma variável é momentaneamente omitida a cada iteração e uma sistemática de clusterização (do tipo *k-means*) é realizada. A cada omissão de variável, a qualidade de clusterização gerada pela ausência daquela variável é medida através do IS. A variável responsável pelo maior IS ao ser omitida é eliminada do banco de dados, visto que os resultados do agrupamento são melhores quando tal variável não é incluída na análise. Na sequência, o mesmo procedimento de omissão de uma variável por vez é executado sobre o conjunto de variáveis remanescentes; esse processo é repetido até que apenas uma variável reste. O procedimento acima pode ser repetido para um intervalo de valores de k (número de clusters) considerado adequado por especialistas de processo, caso seja diferente daquele encontrado com o procedimento hierárquico acima descrito.

2.3.3 Simulação baseada nos grupos gerados

Nesta etapa, utiliza-se a Simulação de Monte Carlo (SMC) para identificação de mix de produção que maximize a receita ao menor custo médio de mercadoria vendida possível, através da minimização de escassez e perdas. Tal função-objetivo, apresentada na Equação (3), é testada nos grupos de produtos formados na etapa anterior.

As simulações são implementadas em planilha eletrônica. Na Tabela 2.1 ilustra-se a estrutura de simulação para cada grupo formado no passo anterior: dígitos aleatórios são gerados e então convertidos em demanda diária através da SMC. Através da variação do lote de produção diário, pode-se avaliar o melhor cenário para minimizar escassez e perdas e maximizar vendas e receitas.

Cluster 1						Cluster 2						Cluster 3						Cluster 4						
Iteração	Produção (un)	ITENS DISP.	Demanda (un)	Venda (un)	Ñ Venda (un)	Perda (un)	Produção (un)	ITENS DISP.	Demanda (un)	Venda (un)	Ñ Venda (un)	Perda (un)	Produção (un)	ITENS DISP.	Demanda (un)	Venda (un)	Ñ Venda (un)	Perda (un)	Produção (un)	ITENS DISP.	Demanda (un)	Venda (un)	Ñ Venda (un)	Perda (un)
1	130	130	86	86	0	44	38	38	40	38	2	0	21	21	13	13	0	8	160	160	104	104	0	56
2	130	174	115	115	0	15	38	38	28	28	0	10	21	29	10	10	0	11	160	216	131	131	0	29
3	130	145	103	103	0	27	38	48	25	25	0	13	21	32	26	26	5	0	160	189	138	138	0	22
4	130	157	101	101	0	29	38	51	25	25	0	13	21	21	24	21	3	0	160	182	99	99	0	61
9.997	130	172	63	63	0	67	38	52	21	21	0	17	21	25	16	16	0	5	160	240	69	69	0	91
9.998	130	197	79	79	0	51	38	55	39	39	1	0	21	26	18	18	0	3	160	251	123	123	0	37
9.999	130	181	91	91	0	39	38	38	32	32	0	6	21	24	16	16	0	5	160	197	121	121	0	39
10.000	130	169	124	124	0	6	38	44	26	26	0	12	21	26	13	13	0	8	160	199	95	95	0	65

Tabela 2.1 – Estrutura genérica das planilhas de simulação

Os cenários – Conservador, Misto e Agressivo - a ser testados utilizam o tamanho do lote de produção diária como variável de entrada e demanda diária como variável aleatória, conforme a Equação (3). Tal relação representa o lucro ($L(c)$) gerado pela venda dos produtos, receita ($V \times P$), custo de escassez ($\tilde{N}V \times P$) e custo de mercadoria disponível ($D \times C$), onde V significa quantidade de venda efetiva, P preço de venda, $\tilde{N}V$ não-venda ou diferença entre demanda e disponibilidade de mercadoria (nos casos em que a demanda é maior), D a quantidade de itens disponíveis e C o custo de produção destes. Para cada cluster, são calculados valores de preço e custo médios, ponderados de acordo com a participação de cada produto no histórico de vendas.

$$L(c) = V \times P - \tilde{N}V \times P - D \times C \quad (3)$$

2.3.4 Análise e revisão de resultados

O objetivo principal de agrupar produtos em famílias é facilitar o processo de programação da produção para uma grande quantidade de produtos com diferentes características. A programação através de famílias deve gerar resultados - maximização de vendas e minimização de perdas - semelhantes aos obtidos caso o procedimento fosse realizado considerando os produtos individualmente.

2.4 Resultados e Discussão

A sistemática proposta foi aplicada em uma empresa do segmento alimentício que dispõe de aproximadamente 80 produtos, entre doces e salgados, quentes e resfriados, bebidas e alimentos. Alguns destes produtos são fabricados a partir de ingredientes base, enquanto que outros são preparados através da combinação de insumos e matérias-primas pré-manufaturados. Há, ainda, um pequeno grupo (5% sobre o total de produtos ofertados), no qual a fabricação e preparação são terceirizados, sendo realizada apenas a comercialização pela empresa. Para oferecer tais produtos, são necessários mais de 250 itens de matéria prima

e insumos. Tais materiais podem ser específicos de um produto ou compartilhados por diversos. Além da variabilidade entre produtos, insumos e matérias-primas no que tange a volumes, pesos e condições de armazenamento, outro ponto crítico é o *shelf-life*, ou tempo para consumo. O *shelf-life* médio dos produtos é de 2,4 dias, ou seja, após o preparo ou fabricação, os produtos devem ser consumidos no prazo máximo de 2,4 dias, em média. Adicionalmente, mais de 80% dos produtos devem ser consumidos no dia de fabricação ou preparo. Este cenário faz com que o planejamento de produção seja crucial para minimizar perdas e desperdícios sem implicar em perdas de vendas por escassez.

2.4.1 Coleta de dados

Os dados coletados foram organizados em uma matriz descritiva dos produtos, a qual apresenta variáveis que descrevem os produtos em análise com vistas ao seu agrupamento. Inicialmente foram levantadas 45 variáveis para 76 produtos.

Em relação aos dados de vendas, foi coletado um histórico de 90 dias para todos os produtos. Estes dados são importantes em duas frentes. A primeira diz respeito à operacionalização da simulação, visto que a variável aleatória é a demanda. De tal forma, é imprescindível conhecer profundamente o comportamento desta variável para correta estruturação das tabelas de geração de dígitos aleatórios e incidência de demanda. Em segundo plano, percebe-se a demanda como potencial variável relevante para o procedimento de clusterização.

2.4.2 Seleção de variáveis para clusterização

A etapa de seleção de variáveis para clusterização foi iniciada através de procedimento hierárquico e opiniões de especialistas (proprietária da empresa, coordenadora e duas colaboradoras), identificando-se quatro como número adequado de clusters, conforme o dendograma apresentado na Figura 2.1. Neste dendograma, os produtos a serem agrupados encontram-se no eixo vertical, enquanto que o eixo horizontal representa a distância entre os centros dos clusters formados. Cada ramificação indica um cluster e as observações que o compõem. Embora uma primeira análise sugira três como um número adequado de grupos, tal valor foi ajustado para quatro com base na opinião de especialistas de processo, a proprietária da empresa e mais três colaboradoras. Na Tabela 2.2 é apresentada a alocação preliminar dos produtos aos clusters resultantes do processo hierárquico sobre as 45 variáveis iniciais. Pode-se perceber que o Cluster 1 concentra mais da metade das observações (41 produtos), ao passo que o 4 contém apenas cinco produtos. Tais agrupamentos possivelmente serão alterados, já que as etapas subsequentes do método procederão a novas clusterizações com base nas

variáveis selecionadas.

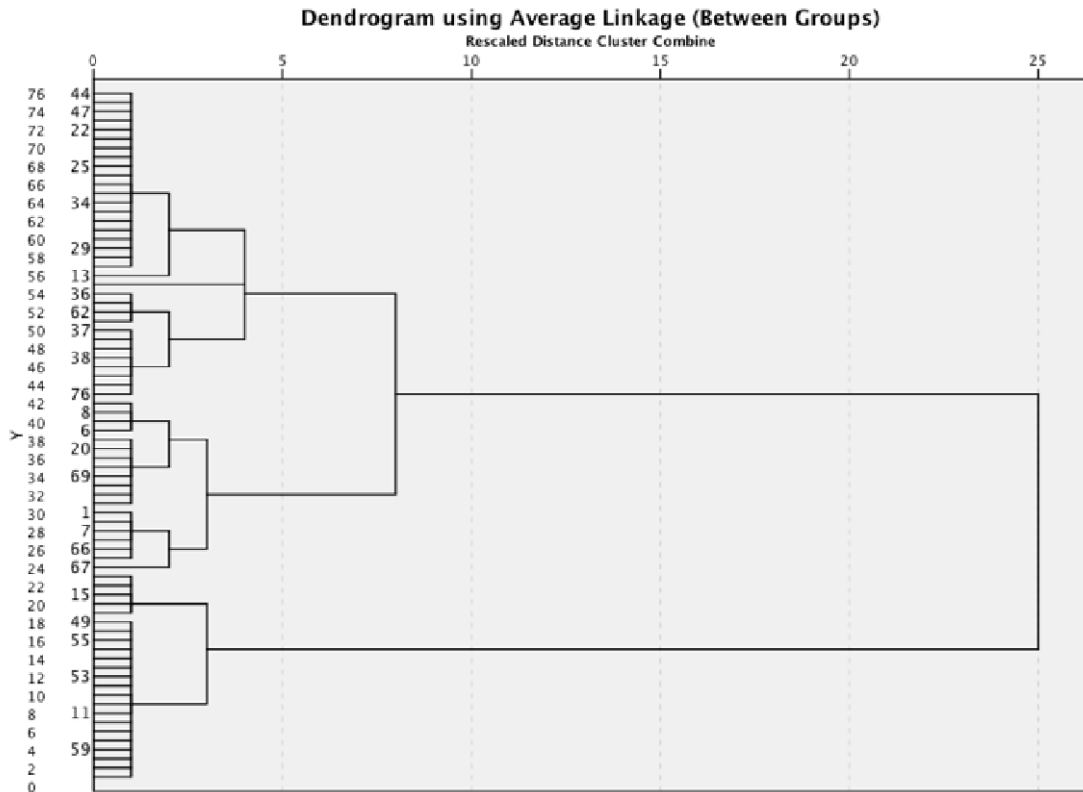


Figura 2.2 - Processo Hierárquico de Clusterização: identificação de nº de clusters

Cluster	Produtos																				
1	1	2	3	4	5	6	7	8	9	13	19	20	48	49	50	51	52	53	54	55	56
	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	
2	10	11	12	21	22	23	24	25	26	44	45	46	47								
3	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42					
4	43	14	15	16	17	18															

Tabela 2.2 – Alocação dos produtos aos clusters (através de dendograma)

A etapa de clusterização não-hierárquica foi iniciada considerando-se as 45 variáveis via software Matlab 7.8. Inicialmente, foram identificadas 10 variáveis com possível inconsistência na opinião de especialistas, restando assim 35 variáveis. Dentre tais inconsistências, ressalta-se a ausência de dados para determinados produtos, não consenso entre os respondentes, irrelevância da variável e magnitudes duvidosas das variáveis (provavelmente em decorrência da inserção de informações equivocadas nos bancos de dados). A cada execução do *k-means*, cada uma das dez variáveis subjetivamente apontadas como inconsistentes pelos especialistas foi eliminada uma a uma e o IS resultante calculado. O IS inicial com as 45 variáveis foi de 0,3723, sinalizando uma baixa qualidade nos agrupamentos; o IS alcançado após a exclusão das 10 variáveis foi 0,7693, indicando que tais variáveis contribuíam negativamente para a qualidade da clusterização (confirmando a

opinião dos especialistas). Na Tabela 2.3 é apresentada a evolução do IS a cada eliminação subjetiva das 10 variáveis tidas como inconsistentes.

Qtde. Variáveis	Eliminação por especialistas e inconsistências nos dados										Início Coef. Var.
	45	44	39	43	38	42	41	37	40	36	35
IS	0,3723	0,3885	0,3921	0,4036	0,4068	0,4073	0,4187	0,4259	0,4288	0,4368	0,7693

Tabela 2.3 – Eliminação das variáveis tidas como inconsistentes pelos especialistas

Concluída a primeira fase de eliminação, calculou-se o coeficiente de variação para as 35 variáveis remanescentes, conforme a equação (4). Tal coeficiente foi utilizado como ordenador de sequência de exclusão de variáveis; a variável com menor coeficiente a cada rodada foi eliminada (partindo-se da premissa que variáveis com maior dispersão permitem agrupamentos mais precisos, como afirmando por Stanley e Brusco (2008)). A cada eliminação, os produtos foram reagrupados em quatro clusters e o IS calculado.

$$\text{Coeficiente de variação} = \frac{\hat{\sigma}}{\hat{\mu}}, \begin{cases} \hat{\sigma} = \text{desvio padrão} \\ \hat{\mu} = \text{média aritmética} \end{cases} \quad (4)$$

O uso do coeficiente de variação possibilitou a redução do número de candidatas de 35 para 10 variáveis. Na Tabela 2.4 são apresentados os valores de coeficiente de variação para cada uma das 25 variáveis eliminadas nesta etapa, bem como o IS gerado. Percebe-se um incremento na qualidade de clusterização à medida que as variáveis com menor variabilidade são eliminadas. Quando a variável 31 é retirada, porém, o IS apresenta decréscimo significativo, indicando que deve-se iniciar o processo de eliminação “omita uma variável por vez”.

Variável Candidata	13	12	22	6	9	5	8	20	26	25	11	28	30
Coeficiente	15%	19%	24%	25%	26%	27%	27%	28%	31%	33%	37%	40%	40%
IS Agrupamento gerado	0,7736	0,7785	0,7787	0,7836	0,7882	0,7950	0,8002	0,8002	0,8008	0,8014	0,8015	0,8027	0,8035
Sequência de eliminação	1	2	3	4	5	6	7	8	9	10	11	12	13
Variável Candidata	10	7	4	24	23	21	27	2	15	16	3	35	31
Coeficiente	40%	42%	45%	45%	47%	49%	49%	57%	57%	61%	62%	72%	73%
IS Agrupamento gerado	0,8036	0,8040	0,8044	0,8054	0,8069	0,8069	0,8093	0,8094	0,8096	0,8113	0,8114	0,8129	0,6353
Sequência de eliminação	14	15	16	17	18	19	20	21	22	23	24	25	26

Tabela 2.4 – Eliminação por coeficiente de variação

O procedimento iterativo “omita uma variável por vez” foi inicializado com as variáveis restantes: 31, 18, 14, 17, 29, 19, 32, 1, 34 e 33, sendo os valores médios de IS para cada iteração apresentados na Tabela 2.5. O maior valor de IS médio é alcançado quando três variáveis (31, 18 e 14) são retidas, apesar de os resultados em geral terem sido satisfatórios. É

possível identificar que há certa complementariedade e talvez interação entre as variáveis, pois a exclusão de uma pode gerar resultados com desempenhos diferentes em outras. Exemplo disto é a variável 1, que na primeira iteração é a segunda que mais contribui para a qualidade do agrupamento (a qualidade do agrupamento cai a IS=0,7868 quando essa variável é omitida da clusterização). Por outro lado, na terceira iteração, esta mesma variável torna-se a mais dispensável (IS=0,9711 quando omitida).

Número Variável	IS 10 Var	IS 9 Var	IS 8 Var	IS 7 Var	IS 6 Var	IS 5 Var	IS 4 Var	IS 3 Var	IS 2 Var	IS 1 Var
31	0,6353	0,5597	0,4372	0,5292	0,5868	0,69	0,8141	0,9818	0,4935	0,9456
18	0,8133	0,9566	0,967	0,9724	0,9761	0,9786	0,9809	0,9818	0,9456	
14	0,8138	0,9571	0,9675	0,9729	0,9767	0,9791	0,9814	0,9818		
17	0,8145	0,8201	0,8382	0,9733	0,9771	0,9795	0,9818			
29	0,8139	0,9575	0,9679	0,9733	0,9771	0,9795				
19	0,8141	0,9576	0,9681	0,9734	0,9772					
32	0,8153	0,9589	0,9695	0,9748						
1	0,7868	0,9606	0,9711							
34	0,8202	0,9657								
33	0,9554									
Is Médio	0,8083	0,8993	0,8858	0,9099	0,9118	0,9213	0,9396	0,9818	0,7196	0,9456

Tabela 2.5 – IS’s gerados pelo procedimento “omita uma variável por vez”

A Figura 2.2 apresenta a evolução da qualidade das clusterizações em razão da eliminação da variáveis nos três passos. Tal redução facilita o processo de caracterização e apropriação dos produtos, inclusive novos, às famílias. A redução de variáveis chegou a mais de 93%, passando das 45 originais para 3 (variáveis 31, 18 e 14, em ordem decrescente de importância).

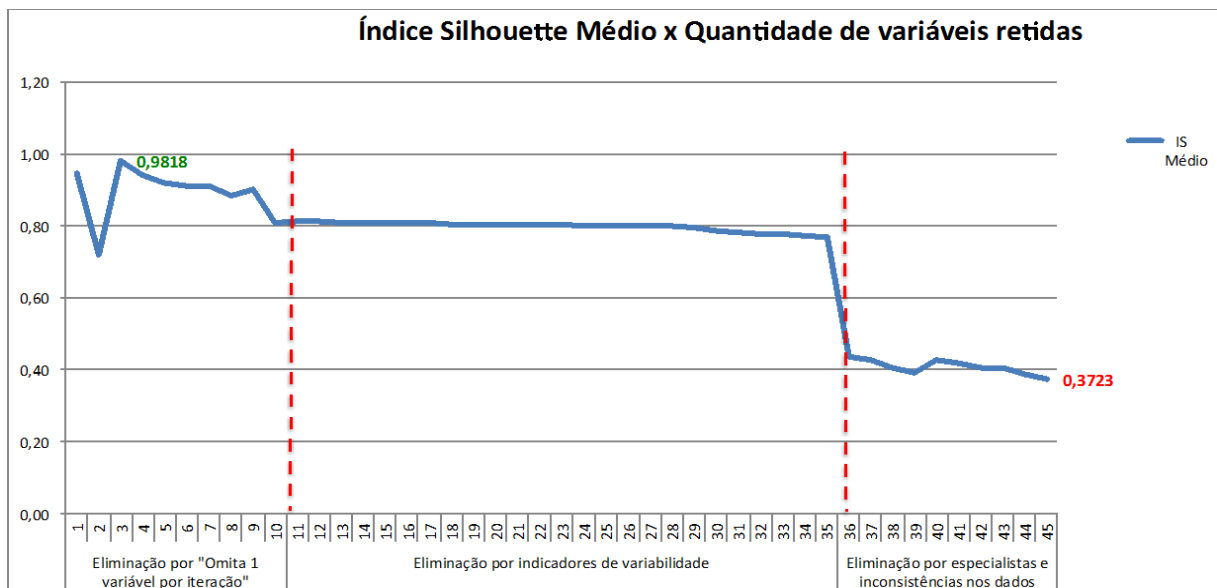


Figura 2.3 – Evolução da qualidade de clusterização medida pelo IS

Na Tabela 2.6 são apresentados os clusters e os produtos que os compõem; percebe-se que a

distribuição dos produtos aos grupos ficou mais equilibrada em relação aos agrupamentos formados no processo hierárquico. Também são apresentados os percentuais de participação em volume de cada produto em seu respectivo cluster.

Cluster	Quant. Produtos	Alocação de produtos aos clusters e composição clusters													
		Produtos	1	2	3	7	8	72	19	20	6	5	68	64	
Bebidas/Doces Elaborados	19	% Part.	42%	24%	11%	6%	6%	2%	2%	1%	1%	1%	1%	1%	
		Produtos	69	67	65	71	73	63	66						
		% Part.	0,6%	0,5%	0,4%	0,2%	0,2%	0,2%	0,2%						
Salgados	22	Produtos	24	70	74	46	30	28	25	29	47	44	21	23	
		% Part.	8%	7%	7%	7%	7%	6%	5%	5%	5%	5%	4%	4%	
		Produtos	45	27	35	22	13	26	32	31	33	34			
		% Part.	4%	4%	4%	3%	3%	3%	3%	2%	2%	2%			
Sanduíches	12	Produtos	75	4	62	38	37	76	39	40	43	36	42	41	
		% Part.	26%	22%	18%	10%	9%	6%	2%	2%	2%	2%	1%	0%	
Bebidas/Doces Simples	23	Produtos	10	61	51	59	12	48	50	15	18	14	11	60	
		% Part.	32%	14%	12%	9%	9%	7%	3%	2%	2%	2%	2%	1%	
		Produtos	9	17	52	16	55	54	58	56	53	49	57		
		% Part.	1,1%	1,0%	1,0%	1,0%	0,8%	0,5%	0,4%	0,4%	0,4%	0,1%	0,1%		

Tabela 2.6 – Alocação final produtos aos clusters

Os agrupamentos gerados foram avaliados por especialistas, tendo sido considerados satisfatórios em termos de similaridades dos produtos inseridos em cada grupo. Em termos práticos, os clusters 2 - Salgados - e 3 - Sanduíches - são formados somente por alimentos, enquanto que os grupos 1 - Bebidas e Doces Elaborados - e 4 - Bebidas e Doces de simples elaboração - são formados por alimentos e bebidas. A empresa agrega produtos em famílias pela natureza destes, por exemplo, há famílias de bebidas quentes e geladas, industrializadas e naturais, alimentos doces e salgados, quentes e frios. No total, são nove as famílias utilizadas habitualmente pela empresa para agrupar os produtos. Através da sistemática utilizada neste estudo, nenhuma destas famílias foi segregada, isto é, produtos pertencentes à mesma família foram alocados ao mesmo cluster. Isto indica que a utilização da sistemática pôde captar similaridades não aparentes sem conflitar com as conforme conhecimento técnico da empresa. Este ponto se torna importante, pois em caso de oscilação de demanda específica de um produto, os demais produtos da mesma família (utilizada pela empresa) podem absorver a variação de demanda, minimizando o desvio entre valor estimado e executado para o cluster em questão.

2.4.3 Simulação

Para realização da simulação de demanda dos produtos agrupados em clusters, foram utilizados os dados históricos de 90 dias de demanda individual de cada produto. Inicialmente, foram levantados preço e custo unitário de obtenção de cada produto. Em um segundo

momento, foram calculados média e desvio padrão de demanda de cada produto. Estas informações foram utilizadas para obter os parâmetros de preço, custo unitário de obtenção, média e desvio padrão de cada cluster através de ponderação pela participação de cada produto nas vendas totais de seu respectivo cluster, conforme composições e percentuais apresentados na Tabela 2.6 da seção 2.4.2. Na Tabela 2.7 são apresentados os dados de média e desvio padrão gerados pela simulação de cada grupo. Os dados de preço e custo unitário ponderados de cada cluster, também apresentados na Tabela 2.7, foram utilizados na Equação (3) de forma a possibilitar a comparação de cenários produtivos diferentes. Na Tabela 2.7 também são apresentados os percentuais de participação de cada cluster no histórico de vendas em unidades. Para obtenção dos dados de média e desvio padrão de demanda diária, foram simulados 10.000 valores seguindo uma distribuição normal para cada cluster, conforme estrutura apresentada na Tabela 2.1.

Clusters	Dados simulados SMC		Dados histórico de vendas		
	Médias (un)	Desvios (un)	% Participação	Custo médio ponderado (R\$)	Preço médio ponderado (R\$)
1. Bebidas/Doces Elaborados	101	22,39	37,4%	0,66	3,71
2. Salgados	29	6,05	10,8%	1,87	5,98
3. Sanduíches	17	5,19	6,0%	1,14	3,31
4. Bebidas/Doces Simples	123	23,06	45,9%	1,10	3,58

Tabela 2.7 – Simulação de demanda clusters

Estes dados foram inseridos na Equação (3) em três cenários distintos de lotes de produção diária: conservador, agressivo e misto. Os cenários são os mesmos para os quatro clusters. O cenário conservador visa minimizar perdas, mesmo que implique em aumento de escassez de produtos. Já no cenário agressivo, o objetivo é realizar todo potencial de demanda maximizando a receita, sem considerar o risco de incremento das perdas. Por último, com o cenário misto, buscou-se encontrar o limite entre maximização das vendas, sem incorrer em maiores perdas. Em termos quantitativos, no cenário conservador foi considerado um lote de produção diária referente a 70% dos valores de média apresentados na Tabela 2.7, no cenário misto os lotes de produção diária foram as médias da Tabela 2.7, e no agressivo foram considerados valores 30% superiores a estes. Tais cenários foram definidos em acordo com os especialistas.

Na Tabela 2.8 é apresentado um resumo com os valores de lote de produção em cada cenário, bem como média e desvio padrão de $L(c)$. Percebe-se que uma redução no lote de produção diária ocasiona perdas consideráveis em $L(c)$, já que este recua mais de 66% do cenário conservador em relação ao cenário misto ($L(c)$ médio reduz de R\$ 593 para R\$ 200),

introduzindo ainda maior variabilidade (desvio padrão de R\$ 112). Em contrapartida, o cenário mais agressivo incrementa $L(c)$ médio em menos de 1%, também aumentando a variabilidade (desvio padrão de R\$ 120).

Como segundo plano de análise, a sistemática auxilia em uma melhor compreensão sobre os custos de perdas e de escassez inerentes a cada cenário. Nas colunas em vermelho na Tabela 2.8 são apresentados tais valores. Percebe-se que, apesar do custo de perda quase zero no cenário conservador, o custo de escassez é equivalente a $L(c)$, lembrando que o custo de escassez representa a receita adicional não realizada devido à falta de produtos. Nota-se que mesmo em um cenário mais agressivo, o custo de escassez é inevitável, porém este não gera desembolso financeiro real à empresa. Tal comportamento é oposto ao custo de perda, que representa prejuízo referente aos produtos disponíveis não vendidos e consequentemente descartados.

Cluster	Cenário Conservador					Cenário Misto					Cenário Agressivo				
	Lote de produção diária (un)	Média de $L(c)$	Desvio Padrão de $L(c)$	Custo de Perdas Médio	Custo de Escassez Médio	Lote de produção diária (un)	Média de $L(c)$	Desvio Padrão de $L(c)$	Custo de Perdas Médio	Custo de Escassez Médio	Lote de produção diária (un)	Média de $L(c)$	Desvio Padrão de $L(c)$	Custo de Perdas Médio	Custo de Escassez Médio
1. Bebidas/Doces Elaborados	70	R\$ 95	R\$ 73	R\$ (1)	R\$ (119)	100	R\$ 243	R\$ 50	R\$ (6)	R\$ (34)	130	R\$ 263	R\$ 76	R\$ (20)	R\$ (4)
2. Salgados	20	R\$ 24	R\$ 33	R\$ (0)	R\$ (58)	29	R\$ 90	R\$ 22	R\$ (4)	R\$ (15)	38	R\$ 87	R\$ 36	R\$ (16)	R\$ (1)
3. Sanduíches	11	R\$ 4	R\$ 14	R\$ (0)	R\$ (19)	16	R\$ 21	R\$ 10	R\$ (2)	R\$ (7)	21	R\$ 22	R\$ 15	R\$ (6)	R\$ (2)
4. Bebidas/Doces Simples	86	R\$ 76	R\$ 76	R\$ (1)	R\$ (4)	123	R\$ 239	R\$ 49	R\$ (10)	R\$ (4)	160	R\$ 223	R\$ 83	R\$ (40)	R\$ (4)
Genral	188	R\$ 200	R\$ 112	R\$ (2)	R\$ (200)	268	R\$ 593	R\$ 74	R\$ (22)	R\$ (60)	349	R\$ 596	R\$ 120	R\$ (83)	R\$ (10)

Tabela 2.8 – Comparação dos cenários produtivos simulados

2.4.4 Análise e verificação de resultados

Para avaliar os benefícios da simulação de demanda com base nos clusters gerados foi realizada uma simulação de Monte Carlo para demanda individual dos produtos. Para cada produto foi rodada a Simulação de Monte Carlo, com base em sua distribuições de probabilidade individuais. Em ambos os cenários – produtos agrupados em clusters e individualmente - foi utilizada a distribuição normal e os resultados finais comparados. Na Tabela 2.9 são apresentados os dados de demanda simulada para os produtos individualmente e para os grupos formados, lembrando que o objetivo principal de agrupar os produtos é facilitar e agilizar o processo de programação da produção sem, no entanto, gerar resultados financeiros inconsistentes àqueles obtidos pela avaliação individual dos produtos. Os produtos inseridos nos clusters 2 e 3 são caracterizados por significativas oscilações nos seus históricos de vendas (o que é amortecido pela simulação com base nos grupos, os quais apoiam-se em

médias de demanda e desvios para cálculo); isso explica a divergência entre dados de demanda simulados agrupada e individual. Por outro lado, os resultados dos clusters 1 e 4 são satisfatórios, pois estes representam as maiores vendas em volume e retorno financeiro, além de apresentar maior regularidade nas vendas.

Clusters	SMC Agrupados		SMC Individual		VARIACÕES	
	Médias (un)	Desvios (un)	Médias (un)	Desvios (un)	Média	Desvios
1. Bebidas/Doces Elaborados	101	22,39	106	14,2	-5%	58%
2. Salgados	29	6,05	37	7,4	-19%	-19%
3. Sanduíches	17	5,19	19	6,1	-14%	-15%
4. Bebidas/Doces Simples	123	23,06	130	23,6	-5%	-2%
Geral	270	33,12	292	29,14	-8%	14%

Tabela 2.9 – Comparação simulação agrupada x individual

São dois os aspectos que interferem no desempenho da sistemática ao simular-se cenários produtivos para itens clusterizados: (i) representatividade dos clusters nos volumes totais de vendas, e (ii) representatividade dos produtos nas vendas totais dos clusters. Oscilações de demanda pontuais e específicas de determinados produtos, como sazonalidades, promoções ou ações específicas afetam diretamente estes dois parâmetros. Caso tais alterações sejam verificadas, é preciso ajustar os percentuais de participação dos produtos nos seus respectivos clusters (e destes no total) antes de proceder à SMC. A Tabela 2.10 compara os dados simulados no cenário misto com os dados reais de vendas do mês de setembro de 2012. O erro absoluto médio foi de 2,89 unidades, que representa um desvio de 1%. Neste mesmo mês, a matriz de composição das vendas foi similar à utilizada para simulação, logo o desempenho dos dados simulados foi satisfatório. Porém, se a matriz apresentasse alterações significativas, mesmo que a quantidade geral de itens fosse rigorosamente a mesma que a simulada, as quantidades projetadas para os clusters poderiam apresentar erros consideráveis.

Clusters	SMC Mista		Dados reais set/12		Desvio - Erro %		Erro abs. (sim.-real)
	% vendas (un)	Quant. Sim. (un)	% Vendas (un)	Quant. (un)	% Vendas	Quant.	
1. Bebidas/Doces Elaborados	37,4%	108,55	36,59%	105	2,21%	3,38%	3,55
2. Salgados	10,8%	31,23	10,80%	31	-0,02%	0,73%	0,23
3. Sanduíches	6,0%	17,88	6,62%	19	-10,12%	-5,89%	-1,12
4. Bebidas/Doces Simples	45,9%	132,23	45,99%	132	-0,30%	0,18%	0,23
Geral	100%	289,89	100%	287	0,00%	1,01%	2,89

Tabela 2.10 – Comparação simulação agrupada x individual

2.5 Conclusão

Este artigo apresentou uma sistemática combinando método de seleção de variáveis para

clusterização e Simulação de Monte Carlo (SMC) com vistas ao aprimoramento do processo produtivo. Como primeiro passo tem-se a coleta e estruturação de dados de produtos, variáveis e demanda. A segunda etapa inicia-se pela identificação de quantidade adequada de clusters através de procedimento hierárquico de clusterização e dendograma. Na sequência, reduz-se a quantidade de variáveis para clusterização, em um primeiro momento através de indicador de dispersão e finalmente por procedimento iterativo do tipo “omite uma variável por iteração”. A qualidade das clusterizações durante a etapa de seleção de variáveis é medida através do Índice Silhouette. Uma vez definidas as variáveis de clusterização, os produtos são agrupados e a SMC é estruturada valendo-se dos grupos gerados. A simulação tem como objetivo avaliar cenários distintos de produção com vistas à maximização de lucros.

A sistemática proposta apresentou uma relação esforço-desempenho/resultado satisfatória, quando bem ajustada à situação que se deseja avaliar. A necessidade de seleção de variáveis para clusterização ficou evidente para redução de esforços e recursos computacionais e aumento da qualidade dos agrupamentos formados. Quanto à SMC, ressalta-se que falhas de ajuste nos percentuais de composição dos clusters podem causar erros consideráveis, porém é fácil ajustá-lo quando cada cluster é simulado individualmente. Por fim, a sistemática proposta tem como vantagens a facilidade e flexibilidade de geração de cenários alternativos de análise, evidenciando seu caráter prático.

Desdobramentos futuros incluem a análise da possível sinergia ou interferência entre as variáveis na etapa de seleção de variáveis, assim como o efeito que a escala e sua cardinalidade impõem à clusterização. Outro ponto consiste na investigação de como ajustar os parâmetros de participação dos produtos nos clusters e destes no total, quando em cenários de alta oscilação de demanda.

2.6 Referências

AMANIFARD, N.; RAHBAR, B.; HESAN, M. Numerical Simulation of the Mitral Valve Opening Using Smoothed Particles Hydrodynamics. Proceedings of the World Congress on Engineering 2011, Vol. III. WCE, July 6-8, 2011, London, UK.

ANZANELLO, M. J. Seleção de variáveis com vistas à classificação de bateladas de produção em duas classes. *Gestão e Produção*, São Carlos, v. 16, n. 4, p. 526-533, out.-dez. 2009.

ANZANELLO, M. J.; FOGLIATTO, F. S. Selecting the best variables for grouping mass-customized products involving worker's learning. *Int. J. Production Economics* 130 (2011) 268–276, 2011.

CABRER, B; OLMOS, J.; PAVIA, J. M.; SALA, R. Actualización de matrices de origen-destino. Un análisis de alternativas a través de MonteCarlo. XV Jornadas de ASEPUMA y III Encuentro Internaciona, 2007.

- CAI, Y.; SUN, Y.** Spirit - Tree: hierarquical clustering analysis of millions of 16s rRNA pyrosequences in quasilinear computaional time. *Nucleic Acids Research*, 2011, Vol. 39, No. 14e95 doi:10.1093/nar/gkr349
- CATELLI, A.** Controladoria: Uma abordagem da Gestão Econômica – GECON. São Paulo. Atlas, 2010.
- CHEZNIAN, V. U.; SUBASH, T.;** Hierarchical sequence clustering algorithm for data mining. *Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011*, July 6 - 8, 2011, London, U.K.
- CHOPRA, S.; MEINDL, P.** Gerenciamento da cadeia de suprimentos : estratégia, planejamento e operação. São Paulo. Prentice Hall, 2003
- COSTA FILHO, P. A.; POPPI, R. J.** Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio. *Determinação de glicose, maltose e frutose. Quim. Nova*, Vol. 25, No. 1, 46-52, 2002.
- COSTA, F. J.** A influência do valor percebido pelo cliente sobre comportamentos de reclamação e boca a boca. Tese (Doutorado em Administração de Empresas). Fundação Getúlio Vargas - EAESP. 2007
- GIL, A. C.** Como elaborar projetos de pesquisa. 3.ed. São Paulo: Atlas, 1991.
- GUYON, I.; ELISSEEFF, A.** An introduction to variable and feature selection. *Journal of Machine Learning Research* 3 (2003) 1157-1182, 2003.
- HAIR JR., J. F.; ANDERSON, R. E.; TATHAM, R.L.; BLACK, W.C.** Análise multivariada de dados. Prentice-Hall. São Paulo. 2003.
- HORTA, R. A. M.; ALVES, F. J. C.** Aplicação de técnicas de data mining para o entendimento da política de financiamento das empresas brasileiras. IV Congresso ANPCONT, 2012. Disponível em: <http://www.anpcont.com.br/site/docs/congressoIV/04/MFC161.pdf>
- JAIN, A. K.** Data clustering: 50 years beyond the K-means. *Pattern Recognition* 31 (2010) 651-666. Disponível em: www.elsevier.com/locate/patrec
- KASHEF, R.; KAMEL, M. S.** Cooperative clustering. *Pattern Recognition* 43 (2010) 2315–2329. Disponível em: www.elsevier.com/locate/pr
- KIM, W. C. MAUBORGNE, R.** A estratégia do Ocano Azul. São Paulo. Campus, 2005.
- LIU, M.; JIANG, X.; KOT, A. C.** A multi-prototype clustering algorithm. *Pattern Recognition* 42 (2009) 689-698. Disponível em: www.elsevier.com/locate/pr
- MIMAROGLU, S.; ERDIL, E.** Combining multiple clusterings using similarity graph. *Pattern Recognition* 44 (2011) 694–703. Disponível em: www.elsevier.com/locate/pr
- MOHAMMAD, N. T.** A fuzzy clustering aproach to filter spam E-mail. *Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011*, July 6 - 8, 2011, London, U.K.
- NAGATANI, T.; OZAWA, S.; ABE, S.** Fast variable selection by block addition and block deletion. *Journal of Intelligent Learning Systems and Applications*, 2010, 2, 200-211.
- NAVEIRO, R. M.; PEREIRA FILHO, I. C.** A análise de grupamentos: uma contribuição à padronização do projeto. *Produção*, v. 2, n. II2, p. 157, março 1992.

- PAMPLONA, E. O. ; SILVA, W. F.** Contribuição da Simulação de Monte Carlo na Projeção de Cenários para Gestão de Custos na Área de Laticínios. In: IX Congresso Internacional de Custos - Florianópolis, SC, Brasil - 28 a 30 de novembro de 2005.
- SARAIVA JÚNIOR, A. F.; RODRIGUES, M. V.; COSTA, R. P.** Simulação de Monte Carlo aplicada à decisão de mix de produtos. *Produto e Produção*, v. 11, n. 2, p. 26-54, jun 2010.
- VILLANUEVA, W. J. P.** Comitê de máquinas em predição de séries temporais. Dissertação (Mestrado em Engenharia Elétrica e de Computação). Universidade Estadual de Campinas, 2006.
- RAFAELI, L.** Análise envoltória de dados como ferramenta para avaliação de desempenho relativo. Dissertação (Mestrado em Engenharia). Universidade Federal do Rio Grande do Sul. 2009.
- RODRIGUES, D. M.; SELITTO, M. A.** Análise do desempenho de fornecedores de uma empresa de manufatura apoiada em análise de aglomerados. *Produção*, v. 19, n. 1, p. 055-069, 2009
- SANTHISREE, K; DAMODARAM, A.** SSM-DBSCAN and SSM-OPTICS: Incorporating a new similarity measure for density for density based clustering of web usage data. *International Journal on Computer Science and Engineering (IJCSSE) ISSN:0975-3397 Vol. 3 No. 9 september 2011.*
- SARAIVA JÚNIOR, A. F.; TABOSA, C. M.; COSTA, R. P.** Simulação de Monte Carlo aplicada à análise econômica de pedido. *Produção*, v. 21, n. 1, p. 149-164, jan./mar. 2011 doi: 10.1590/S0103-65132011005000016.
- SENRA, L. F. A. C.; NANJI, L. C.; MELLO, J. C. C. B. S.; MEZA, L. A.** Estudo sobre métodos de seleção de variáveis em DEA. *Pesquisa Operacional*, v.27, n.2, p.191-207, Maio a Agosto de 2007.
- STEINLEY, D.; BRUSCO, M.** A new variable weighting and selection procedure for K-means cluster analysis. *Multivariate Behavioral Research* 43 (1), 77–108, 2008.
- STEINER, M. T. A; CHAVES NETO, A.; BRAULIO, S. N.; ALVES, V.** Métodos estatísticos multivariados aplicados à engenharia de avaliações. *Gestão e Produção*, São Carlos, v. 15, n. 1, p. 23-32, jan.-abr. 2008.
- YIN, R. K.** Estudo de Caso: planejamento e métodos. 2.ed. Porto Alegre: Bookman, 2001.
- ZAPATA, CARLOS J.; PIÑEROS, LUIS C.; CASTAÑO, DIEGO A.** El método de simulación de Montecarlo en estudios de confiabilidad de sistemas de distribución eléctrica . *Scientia Et Technica*, Vol. X, Núm. 24, mayo-sin mes, 2004, pp. 55-60 Universidad Tecnológica de Pereira Colombia.

3 SEGUNDO ARTIGO

SIMPLIFICAÇÃO DA GESTÃO DE ESTOQUES COM BASE EM SELEÇÃO DE VARIÁVEIS PARA CLUSTERIZAÇÃO

Resumo

A identificação das variáveis mais relevantes para o agrupamento de produtos de acordo com suas demandas produtivas e características físicas aparece como fundamental para aprimorar políticas de gestão de estoques, visto que decisões passam a ser tomadas para grupos de produtos, e não para itens individuais. Este artigo apresenta uma sistemática de seleção de variáveis para clusterização com objetivo de facilitar decisões acerca da gestão de inventários. A sistemática é composta por três etapas: (i) levantamento de dados, (ii) seleção de variáveis para clusterização, e (iii) comparação de políticas ótima e simultânea de gestão de estoques para os grupos de produtos gerados. A sistemática proposta foi aplicada em uma empresa do segmento alimentício caracterizada por grande variedade de produtos e insumos, permitindo simplificar a gestão de estoques de 76 produtos para 3 clusters genéricos, reduzindo o volume de pedidos individuais em aproximadamente 86% e elevando custos de gestão de estoques em apenas 6,2% quando comparado à política ótima.

Palavras-chaves: Seleção de Variáveis, Clusterização, Gestão de Estoques

Abstract

The identification of relevant variables for grouping products according to their features appears enables improving inventory management policies, since decisions are taken to product groups and not for individual items. This paper presents an approach for selecting clustering variables aimed at facilitating decisions about inventory management. The system consists of three steps: (i) data collection, (ii) variable selection for clustering products with similar features into groups, and (iii) comparison of simultaneous and optimal policies of inventory management for groups of products generated. The methodology proposed was applied in a food company characterized by a large variety of products, grouping 76 generic products into 3 clusters, reducing the volume of product orders by approximately 86% and rising inventory costs only 6.2%

when compared to optimal policy.

Keywords: *variable selection, clustering, inventory management*

3.1 Introdução

A variedade de produtos necessária para atender o espectro de consumidores traz à tona o problema relativo à gestão de estoques de matérias-primas e insumos. Quanto maior a necessidade de itens, mais complexas e custosas serão estas atividades. Por outro ângulo, tal gestão oferece oportunidades para criar diferencial competitivo frente à concorrência (WANKE, 2005). Uma forma de viabilizar a gestão de estoques consiste no agrupamento de produtos com características semelhantes, permitindo reduzir a complexidade e volume de informações para análise (GUYON e ELISSEEFF, 2003, HAIR JR. et al., 2003). A geração de tais grupos ou famílias usualmente se vale de variáveis que descrevem características físicas e necessidades de processamento dos diversos modelos de produtos.

Neste cenário, a seleção das variáveis mais relevantes para a alocação dos produtos em famílias de acordo com suas demandas produtivas e características físicas surge como oportuna para melhorar e facilitar políticas de gestão de estoques, visto que decisões passam a ser tomadas para grupos de produtos, e não para itens em separado. De acordo com Hair Jr. et al. (2003), a utilização da totalidade de variáveis em procedimentos de clusterização tende a reduzir a qualidade dos grupos formados, visto que variáveis ruidosas e irrelevantes comprometem a eficiência dos algoritmos de clusterização, conduzindo a alocações equivocadas. Os agrupamentos, por sua vez, podem ser utilizados em paralelo a outras ferramentas (por exemplo, sistemáticas de gestão de reposição de estoque), viabilizando tomadas de decisão mais precisas. Em cenários de altos custos associados ao processamento e recebimento de pedidos ou procedimento operacional complexo, a simplificação e redução de carga de trabalho geradas pelo agrupamento de produtos vem a ser oportuna.

Este artigo propõe uma sistemática de seleção de variáveis com vistas à formação de famílias (grupos) de produtos com demandas produtivas e necessidades de insumos similares. Tais grupos são utilizados para facilitar a gestão de estoques de famílias de produtos (e não itens individualizados), permitindo a visualização e conciliação de objetivos às vezes conflitantes, simplificando o processo de gestão sem acarretar elevação significativa dos custos de guarda e colocação de pedidos frente aos gerados pela política ótima (obtida através do Lote Econômico de Compra - LEC) (RITZMAN e KRAJEWSKI, 2005; WANKE, 2010). A primeira etapa da sistemática proposta apoia-se em um procedimento de seleção do tipo

“omita uma variável por iteração” em conjunto com um índice para avaliação da qualidade do agrupamento gerado, o Índice Silhouette (IS). À cada variável omitida, a qualidade do agrupamento é avaliada; a variável que, quando omitida, conduzir ao maior Índice Silhouette, é eliminada (visto que tal variável é a que menos contribui na formação de agrupamentos precisos). O número de clusters é inicialmente estimado através de clusterização hierárquica (apoiada em dendograma), e a efetiva inserção dos produtos às famílias é realizada através do algoritmo k-means (ferramenta de clusterização do tipo não-hierárquica) (HAIR JR et al., 2003). Os grupos gerados pelas variáveis selecionadas são, então, analisados em cenários de gestão de estoques, comparando-se política ótima de colocação de pedidos (gerada pelo LEC) versus política simultânea (na qual pedidos para itens de um mesmo grupo são feitos de forma conjunta, reduzindo custos associados a solicitações e logística).

Além da presente introdução, o artigo é composto por um referencial teórico acerca dos fundamentos de clusterização, seleção de variáveis e de gestão de estoques na seção 3.2. O método é apresentado na seção 3.3, sendo seus resultados descritos na seção 3.4. A seção 3.5 traz as conclusões do estudo e sugestões de trabalhos futuros.

3.2 Referencial Teórico

Essa seção traz os fundamentos da seleção de variáveis em clusterização e políticas de gestão de estoques.

3.2.1 Seleção de variáveis com vistas à clusterização

Há dois tipos básicos de procedimentos de clusterização: os hierárquicos e os não-hierárquicos, cada qual apresentando uma lógica de agrupamento distinta e apropriada a diferentes aplicações (HAIR JR. et al., 2003).

Nos procedimentos hierárquicos, as observações são agrupadas em clusters de forma progressiva; as inserções são monitoradas através do dendograma, que trata-se de uma árvore hierárquica e que ilustra a forma como as relações hierárquicas são formadas (CARVALHO et al., 2009). Em linhas gerais, os agrupamentos gerados em um estágio genérico N estarão contidos no estágio N+1. Santhisree e Damodaran (2011) sinalizam que há duas mecânicas de funcionamento dos procedimentos hierárquicos de acordo com a sequência de agregação: os aglomerativos e os divisivos. A clusterização hierárquica aglomerativa parte do pressuposto que, inicialmente, cada observação representa um cluster e estas vão sendo agrupadas sequencialmente por critérios de semelhança até que todas observações formem um único cluster. Na clusterização divisiva acontece o oposto: inicialmente todas as observações fazem

parte de um único cluster, que é dividido por questões de diferença entre observações até que os clusters sejam formados por uma única observação (CARVALHO et al., 2009).

Por sua vez, os procedimentos não-hierárquicos partem do pressuposto do conhecimento do número de clusters a ser formado (KASHEF e KAMEL, 2009, FONTANA e NALDI, 2009); tal informação nem sempre é trivial, podendo comprometer a qualidade final da clusterização (JAIN et al., 1999; JAIN, 2010). Uma vez definida a quantidade de clusters, os algoritmos lançam “sementes” randômicas, as quais são caracterizadas como centroides para a medição das distâncias das observações. Um procedimento não-hierárquico bastante utilizado por conta de sua simplicidade e bons resultados é o k-means (FONTANA e NALDI, 2009), onde k é o número de clusters. O método busca minimizar a soma das distâncias das observações aos centroides. As observações são alocadas aos k clusters de forma a minimizar a soma das distâncias entre as observações e os centróides (LIU et al., 2008). Em termos operacionais, Hair Jr. et al. (2003) sugerem o uso de procedimento hierárquico para verificação e definição do número adequado de clusters e, na sequência, a utilização de procedimento não-hierárquico, como o k-means, para efetiva realização dos agrupamentos.

Com a crescente variedade de produtos demandados pelo consumidor, uma dificuldade adicional emerge em procedimentos de clusterização: a quantidade de variáveis que os caracterizam tais produtos (SENRA et al., 2007). Em um passado não muito distante, a quantidade de variáveis necessárias para descrever um produto era reduzida, visto que a gama de modelos ofertada era significativamente menor (KIM e MAUBORGNE, 2005). Neste contexto surge a seleção de variáveis que, segundo Nagatani et al. (2010), incrementa a qualidade de métodos multivariados de agrupamento de observações como a clusterização.

Seleção de variáveis é a identificação de um conjunto reduzido de variáveis que permite definir padrões em grandes bases de dados (VILLANUEVA, 2006). Sistemáticas para seleção de variáveis normalmente baseiam-se em três formas de operacionalização: (i) Backward, na qual modelos iniciam com todas as variáveis; tais variáveis são progressivamente eliminadas através de testes estatísticos e a precisão do modelo continuamente avaliada; (ii) Forward, na qual as variáveis são progressivamente inseridas através de testes estatísticos e a precisão do modelo continuamente avaliada; e (iii) Stepwise, na qual as variáveis inseridas anteriormente podem ser eliminadas após a inserção de uma variável mais relevante. Tais sistemáticas têm sido utilizadas regularmente em procedimentos de clusterização com vistas à formação de agrupamentos mais precisos (GUYON e ELISSEEFF, 2003; STROIEKE et al., 2010).

A avaliação do desempenho de sistemáticas de seleção em clusterização pode ser realizada através do Índice Silhouette (IS) (STROEIKE et al., 2010). Este índice varia de -1 a 1, sendo que -1 indica alocação equivocada e 1 alocação correta de uma observação ao cluster de destino. O IS é calculado para cada observação clusterizada, e uma média geral de todas as observações permite avaliar o desempenho geral da clusterização (STROEIKE et al., 2010; MIMAROGLU e ERDIL, 2010). Matematicamente, o IS parte da distância média entre a observação e as demais alocadas no seu respectivo cluster ($a(j)$) e a distância média entre a observação e as observações do cluster vizinho mais próximo ($b(j)$), conforme a Equação (1). A distância Euclidiana é normalmente utilizada neste procedimento.

$$IS(j) = \frac{b(j) - a(j)}{\text{Max}\{b(j), a(j)\}} \quad (1)$$

3.2.2 Gestão de Estoques

Segundo Christopher (2011), quando um cliente depara com a falta de um produto desejado, somente em 19% das ocorrências ele substitui por outro da mesma marca. De tal forma, Belfiore et al. (2006) salientam que deve-se encontrar maneiras de não incorrer em cenários de escassez de disponibilidade de produto, sem no entanto incrementar custos de estoques e/ou ao preço final ao consumidor. Wanke (2005) sinaliza que a manutenção de estoques pode acarretar um custo anual de oportunidade na ordem 20% a 40% do valor imobilizado; o mesmo autor enfatiza que apenas a administração de estoques pode representar 1% do valor do produto. Neste contexto, onde a penalização pode vir de ambas as extremidades da cadeia (fornecedor e cliente), o ajuste correto de assertividade e nível de estoque, agilidade e informação, torna-se fundamental (CHOPRA e MEINDL, 2005). Szajubok et al. (2006) mencionam ainda a dificuldade relativa à classificação dos itens de estoque por conta da elevada variedade de tais itens e da necessidade de tratamento adequado a cada categoria.

Diversos custos tornam-se relevantes na gestão de estoques: (i) custo da mercadoria propriamente dita, (ii) custo do pedido, (iii) custo de armazenagem, (iv) custo de oportunidade, e (v) custo de escassez (WANKE, 2005; RITZMAN e KRAJEWSKI, 2008; CHRISTOPHER, 2011). O custo da mercadoria é o mais explícito, pois trata-se do desembolso para obtê-la. O custo de pedido, também chamado de setup ou preparação (CHOPRA e MEINDL, 2005), está relacionado à emissão ou geração do pedido e aborda custos com preparação de máquinas e insumos, emissão de documentação do pedido e transporte, entre outros. Em relação à armazenagem, consideram-se custos com local,

equipamentos e pessoal necessários para manusear e armazenar o material. Já o custo de oportunidade está ligado aos recursos financeiros necessários para administrar e armazenar os estoques, e que poderiam ser investidos em outras frentes para geração de retorno financeiro. O custo de escassez está relacionado à indisponibilidade do produto e não realização da venda, o que pode ainda acarretar danos à imagem do produto faltante.

Uma abordagem bastante difundida para gestão de estoques é o Economic Order Quantity (EOQ) ou Lote Econômico de Compra - LEC (RITZMAN e KRAJEWSKI, 2008; WANKE, 2010). É uma política que visa manter os estoques suficientemente reduzidos para que não haja desabastecimento, minimizando custos anuais de pedidos e manutenção de estoques e frequência de pedidos. A lógica consiste em definir o tamanho ideal de lote que minimize custos e frequência de pedidos. O tamanho do lote pode ser calculado pela Equação (2), a qual apoia-se nas seguintes premissas: (i) demanda constante e conhecida com precisão, (ii) inexistência de limites acerca do tamanho do lote, (iii) somente custos de pedido e manutenção de estoques são considerados, (iv) decisões acerca de um produto não afetam outros itens, e (v) tempo de suprimento conhecido e constante.

$$Q_{eoq}^* = \sqrt{\frac{2 \cdot D \cdot S}{H}} \quad (2)$$

Na Equação (2), D representa a demanda anual do produto em unidades, S o componente referente ao custo de colocação de um pedido (\$/pedido), e H o custo referente à guarda ou armazenagem de uma unidade do produto em um ano (normalmente representado em percentual do custo do produto). Os custos de pedido $C(P)$ e de guarda $C(G)$ são estimados através das Equações (3) e (4), respectivamente.

$$C(P) = \frac{D \cdot S}{Q} \quad (3)$$

Na Equação (4), do custo de guarda, são explicitados três componentes. I é componente relativo ao estoque médio entre pedidos, sendo igual à metade do lote pedido Q ; o componente C representa o custo unitário do produto em análise e i é relativo ao custo de armazenagem, normalmente expresso em percentual de C .

$$C(G) = I \times C \times i$$

(4)

O custo total é obtido pela soma de $C(P)$ e $C(G)$. Assim, o LEC permite a minimização dos custos associados à colocação de pedidos e guarda, desde que a colocação das ordens de reposição siga rigorosamente o tamanho calculado para o lote.

Uma sistemática de gestão de estoques alternativa ao LEC é a Política Simultânea de manutenção de estoques, na qual os produtos são agrupados de forma a reduzir os esforços com processamento de pedidos. Para tanto, são redefinidos novos tamanhos de lotes (Q) e frequências de reposição, gerando políticas específicas para cada agrupamento. Um forma de agrupar os produtos é curva ABC (CHOPRA e MEINDL, 2005; CRHISTOPHER, 2011), na qual são priorizados itens de maior representatividade financeira. Nas proposições deste artigo, os grupos são formados com base em suas similaridades estruturais e financeiras através de clusterização com variáveis selecionadas, facilitando a adoção de políticas de gestão de estoques para cada grupo formado.

3.3 Método de trabalho

Por incorporar o uso de técnicas de análises multivariadas de dados e cálculos relativos a políticas de gestão de estoques, o presente artigo pode ser caracterizado como quantitativo. Em relação à natureza do estudo, pode-se afirmar que trata-se de uma pesquisa aplicada, uma vez que a sistemática visa fornecer embasamento para decisões sobre um problema específico (GIL, 1991). Do ponto de vista do procedimento, é classificado como uma pesquisa experimental, já que o objetivo maior é a busca de meio para viabilizar agrupamentos precisos e decisões sobre gestão de estoques. Quanto aos objetivos, o método é descritivo, pois são levantados e descritos dados e características de produtos com vistas a melhorias na gestão de gestão de estoques (YIN, 2001).

A metodologia proposta é operacionalizada em três etapas: (i) coleta de dados, (ii) seleção de variáveis para clusterização de observações em grupos, (iii) geração de cenários de gestão de estoques e análise dos resultados obtidos. Tais etapas são agora detalhadas.

3.3.1 Coleta de dados

Os dados necessários para execução da sistemática proposta baseiam-se em variáveis que caracterizam os produtos sob diferentes perspectivas. As variáveis podem ser de natureza física (peso, volume ou composição), de ordem financeira (custo, preço, margem bruta), e relativas aos recursos produtivos e de armazenagem utilizados por cada produto, bem como a forma de consumo de cada produto.

Os dados relativos a características físicas e necessidade de recursos produtivos são coletados em procedimentos operacionais padrão e receitas de produtos. Aqueles associados a aspectos financeiros podem ser coletados em tabelas de custo padrão, notas fiscais de entrada, relatórios gerenciais e própria relação de preços de produtos. Dados sobre forma de consumo

podem ser obtidos por observação.

3.3.2 Seleção de variáveis para clusterização

O ponto de partida para seleção de variáveis para clusterização não-hierárquica é a definição da quantidade de clusters necessários. Esta definição é realizada através da execução de procedimento hierárquico através de dendograma (ferramenta visual de procedimentos hierárquicos). O dendograma possibilita visualizar a forma com os agrupamentos são gerados progressivamente e as distâncias entre estes. Como maiores distâncias entre centros dos agrupamentos indicam melhor qualidade de separação entre observações não-similares, é possível estimar o número k de clusters a ser utilizado no procedimento não-hierárquico k -means. Tal combinação de procedimento hierárquico com não-hierárquico conduz a clusterizações de maior qualidade (SANTHISREE e DAMODARAM, 2011)

Para selecionar as variáveis mais relevantes, propõe-se um procedimento iterativo do tipo “omita uma variável por iteração” associado ao procedimento não-hierárquico k -means. A lógica do procedimento baseia-se na verificação da qualidade da clusterização gerada quando cada variável é omitida. À cada omissão de variável, executa-se o algoritmo k -means e verifica-se a qualidade da clusterização resultante através do Índice Silhouette (IS) (STROIEKE et al., 2010). A variável que gerar melhor qualidade quando omitida é eliminada do grupo de variáveis candidatas, visto que quando omitida a qualidade é incrementada. Este procedimento é realizado até que reste somente uma variável. O subconjunto de variáveis com o maior IS médio é selecionado para clusterização. Os grupos formados são então analisados em termos de políticas de estoques, como segue.

3.3.3 Política Ótima x Simultânea Clusterizada

Como primeira etapa para análise de políticas de gestão de estoques, investiga-se o comportamento de cada produto através da utilização do lote econômico (EOQ ou LEC) (RITZMAN e KRAJEWSKI, 2005, WANKE; 2010). Para tanto, são calculados tamanhos de lote, frequência de reposição e custos associados à colocação de pedido e guarda. Conforme descrito em 2.2, o modelo EOQ apresenta premissas que podem não ser atendidas completamente em um cenário real, porém fornece aproximações adequadas para uso em aplicações práticas.

Em um segundo momento, investiga-se o uso de política simultânea de gestão de estoques, isto é, com produtos agrupados através das variáveis selecionadas (e não através do LEC). Tais grupos também são avaliados em termos de tamanhos de lote, frequência de reposição e

custos associados à colocação de pedido e guarda. Christopher (20011) e Chopra e Meindl (2005) sugerem o uso da curva ABC para agrupamento e priorização de itens em função da representatividade financeira. Na presente sistemática, os clusters formados a partir da seleção e verificação de variáveis mais importantes para clusterização são implementados na sistemática de gestão simultânea. São geradas políticas simultâneas distintas e independentes para cada cluster, definindo frequências de reposição de acordo com a demanda individual de seus respectivos produtos. Os produtos são então agrupados e geridos de acordo com suas frequências de reposição. Por exemplo, para o cluster genérico A podem ser definidas três frequências de reposição (mensal, bimestral e trimestral) para os itens que o compõem; uma parcela de produtos daquele cluster será repostada mensalmente, outra bimestralmente e a última a cada três meses. Por fim, os custos da política ótima e simultânea são comparados.

3.4 Resultados e Discussão

O presente estudo foi aplicado em uma empresa de alimentação que oferece produtos de diversas naturezas, como doces e salgados, bebidas e alimentos, industrializados e naturais. Como características destes salienta-se o curto prazo de validade, o grande volume de itens, os procedimentos específicos de acondicionamento e armazenagem, que afetam diretamente os objetivos da sistemática.

3.4.1 Coleta de dados

Os dados coletados foram organizados em uma matriz descritiva dos produtos, a qual apresenta variáveis que caracterizam os produtos em análise com vistas ao seu agrupamento. Foram levantadas 20 variáveis para 76 produtos, conforme ilustrado na Tabela 3.1. Foi coletado um histórico de vendas de 90 dias e calculada a quantidade diária média de vendas de cada produto.

3.4.2 Seleção de variáveis para clusterização

A etapa de seleção de variáveis para clusterização foi iniciada através de procedimento hierárquico, identificando-se três como número adequado de clusters, conforme o dendograma apresentado na Figura 3.1. O dendograma posiciona os produtos no eixo vertical de acordo com os agrupamentos gerados; o eixo horizontal traz as distâncias entre os centros dos clusters resultantes. À medida que os produtos são agrupados, formam-se os ramos representados horizontalmente, que indicam a sequência de agrupamento dos produtos, resultando na redução do número de clusters e aumento das distâncias entre seus centros.

Produtos	Var 1	Var2	Var 3	Var ...	Var ...	Var ...	Var 18	Var 19	Var 20
1	1,9	0,6	16,0	20,7	0,8	2,7	180	120	31,9
2	4,8	4,4	40,0	44,0	5,2	10,0	90	10	1,6
3	3,0	0,7	14,9	17,6	0,9	3,9	180	40	7,1
...	0,9	0,6	25,2	42,0	0,6	1,5	3	36	4,2
...	2,0	0,5	13,5	18,5	0,7	2,7	180	40	1,4
...	3,1	0,6	12,2	15,3	0,9	4,0	180	20	1,9
74	1,4	0,9	34,0	38,3	1,0	2,4	360	60	35,4
75	2,6	2,3	42,1	46,3	2,4	5,0	360	38	9,7
76	2,5	3,2	42,0	48,5	4,0	6,5	90	10	2,1

Tabela 3.1 – Matriz descritiva de produtos e variáveis (parcial)

A etapa de clusterização *k*-means foi realizada em software Matlab 7.8 através do procedimento iterativo “omite uma variável por vez”. O IS médio inicial valendo-se das 20 variáveis originais foi de IS=0.7869. O maior valor de IS médio é alcançado quando sete variáveis (5, 17, 15, 13, 10, 18 e 6) são retidas, IS= 0.8767. Na Figura 3.2 é apresentada a evolução do IS médio à medida que as variáveis são excluídas. Quanto mais próximo de 1 o IS melhor a qualidade do agrupamento gerado, porém a marca de 0.8767 indica uma boa qualidade e suficiente para o desenvolvimento do estudo.

Produtos	
Cluster 1	9, 10, 11, 12, 14, 15, 16, 17, 18, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 e 61
Cluster 2	2, 3, 5, 6, 7, 8, 19, 20, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72 e 73
Cluster 3	13, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 62, 70, 74, 75 e 76

Tabela 3.2 – Alocação dos produtos aos clusters

Os clusters resultantes são apresentados na Tabela 3.2; percebe-se uma distribuição de produtos equilibrada entre os clusters, sendo que no cluster 1 foram reunidos 23 produtos, 19 no cluster 2, e 34 no cluster 3 (representando 30%, 25% e 45% do total de produtos, respectivamente).

Os agrupamentos gerados foram avaliados por especialistas e considerados satisfatórios em termos de similaridades dos produtos inseridos em cada grupo. Salienta-se ainda que, no cluster 1, foram alocados bebidas e alimentos de elaboração mais simples, no cluster 2 bebidas mais elaboradas e de maior valor agregado e no cluster 3 alimentos mais elaborados e de maior valor agregado (observações estas feitas por especialistas: proprietária, coordenadora e duas colaboradoras).

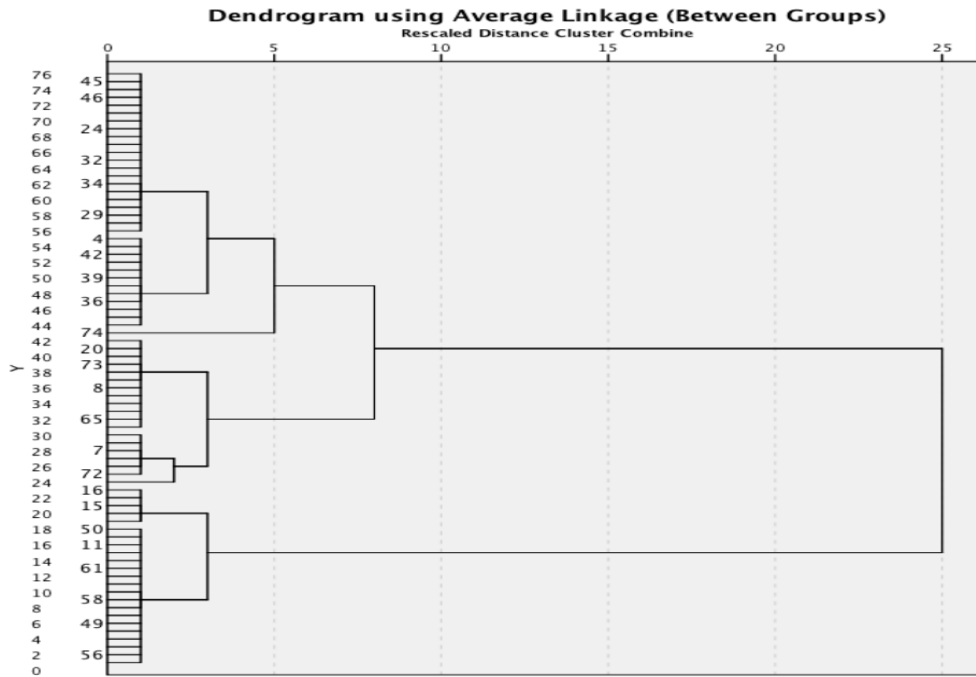


Figura 3.1 - Processo Hierárquico de Clusterização: identificação de nº de clusters

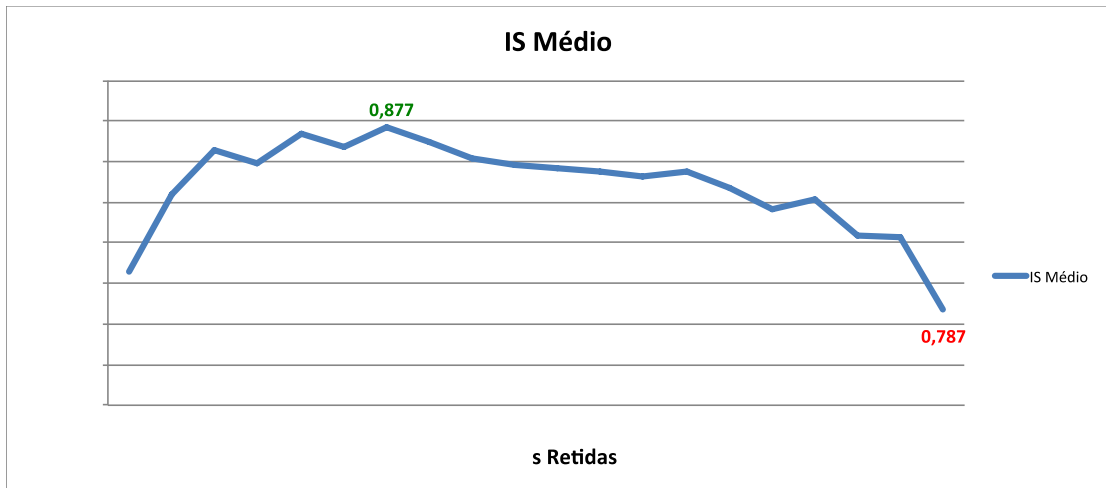


Figura 3.2 – IS’s gerados pelo procedimento “omita uma variável por vez”

3.4.3 Política Ótima x Simultânea Clusterizada

Inicialmente, foram calculados os tamanhos de lotes econômicos, número de pedidos de reposição por ano e intervalos entre pedidos em semanas para cada produto de forma individual. Também calculados os custos de processamento de pedidos e de guarda; repare que, por se tratar do lote econômico, custos de guarda e de colocação de pedido são idênticos, conforme a Tabela 3.3. O valor total anual de custo de guarda e pedido ficou em torno de R\$ 8.700,00, o que equivale a 4,1% do custo de produto no mesmo período. Os lotes econômicos variaram de 1 a 28 unidades, a quantidade de pedidos de 1 a 34 por ano, e intervalos entre pedidos de 2 a 37 semanas. Ao gerir-se cada produto de forma individual, em um ano serão

feitos 507 pedidos. Para fins de comparação com a política simultânea, se os produtos de mesma frequência ótima fossem agrupados, ainda assim seriam necessários 214 pedidos ou reposições em um ano. Os lotes econômicos foram calculados utilizando custo anual de guarda (parte do termo H na equação 2) equivalente a 29,7% do valor do produto e custo de processamento (termo S na equação 2) de pedido equivalente a 12% do valor do produto. Tais parâmetros são calculados e utilizados pela empresa atualmente.

Cód. Prod.	Cluster	Demanda anual	Custo (R\$/un)	Lote Econômico (un)	Frequência anual	Intervalo entre pedidos (semanas)	Custo Produto (R\$/ano)	Custo de Pedido (R\$/ano)	Custo de Guarda (R\$/ano)
9	1	43	7,11	6	7	7	303	6,20	6,20
....	1	98	262,81	9	11	5	25.675	346,76	346,76
61	1	20	248,45	4	5	10	5.036	149,32	149,32
1	2	43	546,22	6	7	7	23.517	478,44	478,44
....	2	18	586,54	4	5	11	10.553	332,12	332,12
73	2	9	632,62	3	3	16	5.583	250,88	250,88
4	3	157	10,50	11	14	4	1.651	17,58	17,58
....	3	8	30,48	3	3	17	236	11,33	11,33
76	3	62	109,38	7	9	6	6.823	115,32	115,32
Total		4.399			507		212.283	4.339	4.339

Tabela 3.3 – Frequências e custos derivados dos lotes econômicos

As seguintes premissas balizaram a elaboração de políticas simultâneas de gestão de estoques: (i) frequência máxima de reposição mensal (o estabelecimento não poderia solicitar mais de dois pedidos no mesmo mês), (ii) frequência de reposição individual na política simultânea igual ou inferior àquela gerada através da utilização do lote econômico, e (iii) elaboração de, no máximo, 3 políticas distintas para os itens inseridos num mesmo cluster. A primeira premissa foi sugerida por especialistas devido a aspectos logísticos de recebimento dos pedidos, visto que reposições mais frequentes acarretam custos extras e demandam tempo considerável, já que é feita de forma manual, produto a produto, via telefone ou fax. Em relação ao recebimento, por se tratar de alimentos, há uma série de procedimentos impostos pela fiscalização de vigilância sanitária a serem cumpridos. Tais procedimentos tornam o recebimento mais demorado e burocrático. Em vistas destes fatos, um espaçamento adequado entre pedidos é fundamental. As premissas (ii) e (iii) foram definidas como direcionadoras e não restrições, uma vez que os valores obtidos através do cálculo do lote econômico são ótimos, não sendo possível superá-los em termos de custos no sistema.

Na Tabela 3.4 é apresentado o cenário de gestão simultânea proposto. Para os clusters 1 e 2, foram definidas políticas com frequência de reposição mensal, bimestral e semestral. Para mais da metade dos produtos alocados nestes clusters, a frequência de reposição seria semestral, com pedidos compostos por 16 produtos para o cluster 1 e 12 produtos para o cluster 2. Para 5 produtos do cluster 1 e 3 do cluster 2, a reposição seria realizada a cada 2

meses. Somente 2 produtos no primeiro cluster e 4 produtos no segundo necessitariam de reposição mensal. Para o cluster 3 não haveria necessidade de reposição mensal; para 7 produtos ela aconteceria 2 vezes a cada 3 meses. Para o último agrupamento também haveria reposições bimestrais (para 11 produtos) e quadrimestrais (16 produtos). A adoção de tais políticas para produtos clusterizados reduziria custos de colocação de pedidos em 33,4% e aumentaria o custo de guarda em 63,3%, resultando em um incremento total de 15% nos custos de gestão de estoques. Enfatiza-se, no entanto, o ganho não tangível associado à simplificação do processo de gestão do estoque, a qual é decorrente da adoção da política agrupada.

Cluster	Freq. Anual Reposições	Produtos (un)	Custo Produto (R\$)	Política Ótima			Política Simultânea			Variação		
				Custo Pedidos (R\$)	Custo Guarda (R\$)	Custos Totais (R\$)	Custo Pedidos (R\$)	Custo Guarda (R\$)	Custos Totais (R\$)	Custo Pedidos (%)	Custo Guarda (%)	Custos Totais (%)
1	12	2	22.803	113	113	226	46	282	329	-59,0%	149,5%	45,2%
	6	5	50.843	766	766	1.532	472	1.258	1.730	-38,4%	64,3%	12,9%
	2	16	19.576	682	682	1.364	344	1.453	1.797	-49,6%	113,2%	31,8%
2	12	4	10.344	100	100	201	81	128	209	-19,5%	27,4%	4,0%
	6	3	24.265	493	493	986	405	601	1.005	-17,9%	21,8%	2,0%
	2	12	22.942	873	873	1.746	471	1.703	2.174	-46,1%	95,1%	24,5%
3	9	7	21.506	305	305	610	270	355	625	-11,4%	16,4%	2,5%
	6	11	27.304	569	569	1.137	483	676	1.159	-15,0%	18,9%	1,9%
	3	16	12.700	438	438	876	320	629	949	-27,0%	43,5%	8,2%
Total	58	76	212.283	4.339	4.339	8.678	2.891	7.085	9.977	-33,4%	63,3%	15,0%

Tabela 3.4 – Políticas de manutenção de estoques

A partir do cenário inicial da Tabela 3.4, passou-se a uma revisão subjetiva na política simultânea (chamado de simultânea Revisada), com vistas à redução do custo guarda gerado. Na Tabela 3.5 são apresentadas as melhorias obtidas com a revisão do cenário anterior. O cluster 2 não apresentou alterações significativas, ao passo que os clusters 1 e 3 foram alterados em termos de suas frequências de reposição. No cluster 1, 4 produtos tiveram suas frequências de reposição alteradas de 12 para 9 meses, enquanto que um modelo de produto foi reduzido a 3 reposições anuais. No cluster 3, 3 produtos passaram a ser repostos mensalmente, 4 produtos uma vez a cada 6 semanas, e os outros 27 passaram a seguir uma lógica de reposição bimestral. Estas alterações possibilitaram uma redução nos custos de guarda, em relação ao cenário anterior, em 25% (de R\$ 7.085 para R\$ 5.279), incorrendo em um incremento relativo aos custos de pedidos de 36%. Esse novo cenário difere 6,2% dos resultados gerados pela política ótima. Salienta-se, entretanto, que tal incremento de 6,2% nos custos totais da política Simultânea Revisada, quando analisados em relação aos custos de produtos (cuja variação acontece apenas em função da variação demanda anual D) representam 4,3%. Em relação a política Ótima, o incremento de custos de guarda e de

pedidos frente aos custos de produtos de foi de apenas 0,2%, através da adoção da Simultânea Revisada.

Cluster	Freq. Anual Reposições	Produtos (un)	Política Ótima				Política Simultânea			Variação		
			Custo Produto (R\$)	Custo Pedidos (R\$)	Custo Guarda (R\$)	Custos Totais (R\$)	Custo Pedidos (R\$)	Custo Guarda (R\$)	Custos Totais (R\$)	Custo Pedidos (%)	Custo Guarda (%)	Custos Totais (%)
1	12	2	22.803	113	113	226	46	282	329	-59,0%	149,5%	45,2%
	9	4	50.540	760	760	1.520	700	834	1.534	-7,9%	9,8%	0,9%
	3	17	19.879	688	688	1.376	518	984	1.502	-24,7%	43,0%	9,1%
2	12	4	10.344	100	100	201	81	128	209	-19,5%	27,4%	4,0%
	6	5	35.666	849	849	1.698	855	883	1.738	0,7%	3,9%	2,3%
	2	10	11.541	517	517	1.034	321	857	1.177	-38,0%	65,8%	13,9%
3	12	3	8.226	94	94	188	88	102	189	-6,7%	8,4%	0,8%
	9	4	13.279	211	211	422	205	219	424	-3,0%	3,8%	0,4%
	6	27	40.004	1.007	1.007	2.013	1.123	990	2.113	11,6%	-1,7%	5,0%
Total	71	76	212.283	4.339	4.339	8.678	3.936	5.279	9.215	-9,3%	21,7%	6,2%

Tabela 3.5 – Políticas Simultâneas de manutenção de estoques revisadas

A Tabela 3.6 traz um quadro comparativo dos três cenários testados (Ótimo, Simultânea e Simultânea revisada) detalhando quantidades de pedidos anuais, custos de pedidos, de guarda e totais. Os dados também são desdobrados em clusters possibilitando a visualização de suas participações. Percebe-se que cluster 1 é o que representa a maior parcela do custo de guarda e o cluster 3 do custo de pedido, independente do cenário. Também verifica-se que, em ambas as proposições, a redução de quantidade de pedidos anuais aproxima-se de 90%.

Políticas	Pedidos Anuais	Custo de pedido (R\$)	Custo de guarda (R\$)	Custos Totais (R\$)
Ótima	508	4.339	4.339	8.678
Simultânea	58	2.891	7.085	9.977
Cluster 1	20	862	2.994	3.856
Cluster 2	20	956	2.432	3.388
Cluster 3	18	1.074	1.659	2.733
Simultânea Revisada	71	3.936	5.279	9.215
Cluster 1	24	1.264	2.100	3.364
Cluster 2	20	1.256	1.868	3.124
Cluster 3	27	1.415	1.311	2.726
Varição (Ótima > Simultânea)	-88,6%	-33,4%	63,3%	15,0%
Varição (Ótima > Simultânea Revisada)	-86,0%	-9,3%	21,7%	6,2%

Tabela 3.6 – Comparativo das Políticas Ótima, Simultânea e Simultânea Revisada

3.5 Conclusão

Este artigo apresentou uma sistemática de seleção de variáveis para clusterização de modelos de produtos, facilitando a adoção de políticas de gestão de estoques para cada grupo formado. A sistemática é composta por três etapas: (i) coleta de dados, (ii) seleção de variáveis para clusterização através de uma sistemática “omita uma variável por vez”, e (iii) comparação de

políticas de gestão de estoques ótima e simultânea.

Ao ser aplicada em uma empresa do segmento alimentício caracterizada por grande variedade de produtos e insumos, a sistemática proposta permitiu simplificar a gestão de estoques de 76 produtos para 3 clusters, reduzindo o volume de pedidos individuais em aproximadamente 86% e elevando custos de gestão de estoques em apenas 6,2% quando comparado à política ótima. Os resultados são considerados satisfatórios, visto que a redução de esforços com manutenção de pedidos e reposições é proporcionalmente maior que o incremento de custos totais de gestão de estoques, levando-se em conta que na política ótima estes equivalem a 4,1% dos custos dos produtos e na simultânea passam a ser de 4,3%.

Desdobramentos futuros incluem a análise de sensibilidade dos custos de processamento de pedidos e de guarda. Também são oportunos detalhamentos acerca de sistemáticas que incluam estes itens à elaboração de políticas simultâneas de gestão de estoques. A incorporação de análises relativas à capacidade de armazenagem complementar a aplicação da sistemática, uma vez que o impacto do aumento no tamanho dos lotes recebidos (derivado da redução de quantidades de pedidos) não foi avaliado.

3.6 Referências

- ANZANELLO, M. J.; FOGLIATTO, F. S.** Selecting the best variables for grouping mass-customized products involving worker's learning. *Int. J. Production Economics* 130 (2011) 268–276, 2011.
- BELFIORE, P. P. ; COSTA, O. L. V.; FÁVERO, L. P. L.** Problema de estoque e roteirização: revisão bibliográfica. *Revista Produção*, v. 16, n.3, p. 442-454, Set/Dez 2006.
- CARVALHO, A. X. Y.; ALBUQUERQUE, P. H. M.; ALMEIDA JUNIOR, G. R.; GUIMARÃES, R. D.; LAURETO, C. R.** Clusterização hierárquica especial com atributos binários. Brasília, 2009. Disponível em: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=4738
- CHOPRA, S.; MEINDL, P.** Gerenciamento da cadeia de suprimentos: estratégia, planejamento e operação. São Paulo. Prentice Hall, 2003
- CHRISTOPHER, M.** Logística e gerenciamento da cadeia de suprimentos. São Paulo. Cengage Learning, 2011
- FONTANA, A.; NALDI, M. C.** Estudo de comparação de métodos para estimação de números de grupos em problemas de agrupamentos de dados. Universidade de São Paulo, 2009. ISSN – 0103-2569.
- GIL, A. C.** Como elaborar projetos de pesquisa. 3.ed. São Paulo: Atlas, 1991.
- GUYON, I.; ELISSEEFF, A.** An introduction to variable and feature selection. *Journal of Machine Learning Research* 3 (2003) 1157-1182, 2003.

- HAIR JR., J. F.; ANDERSON, R. E.; TATHAM, R. L.; & BLACK, W. C.** Análise multivariada de dados. Prentice-Hall. São Paulo. 2003.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J.** Data clustering: A Review. *ACM Computing Surveys*, v. 31, n. 3, September, 1999.
- JAIN, A. K.** Data clustering: 50 years beyond the K-means. *Pattern Recognition* 31 (2010) 651-666. Disponível em: www.elsevier.com/locate/patrec
- KASHEF, R.; KAMEL, M. S.** Cooperative clustering. *Pattern Recognition* 43 (2010) 2315–2329. Disponível em: www.elsevier.com/locate/pr
- KIM, W. C. MAUBORGNE, R.** A estratégia do Ocenô Azul. São Paulo. Campus, 2005.
- LIU, M.; JIANG, X.; KOT, A. C.** A multi-prototype clustering algorithm. *Pattern Recognition* 42 (2009) 689-698. Disponível em: www.elsevier.com/locate/pr
- MIMAROGLU, S.; ERDIL, E.** Combining multiple clusterings using similarity graph. *Pattern Recognition* 44 (2011) 694–703. Disponível em: www.elsevier.com/locate/pr
- NAGATANI, T.; OZAWA, S.; ABE, S.** Fast variable selection by block addition and block deletion. *Journal of Intelligent Learning Systems and Applications*, 2010, 2, 200-211.
- VILLANUEVA, W. J. P.** Comitê de máquinas em previsão de séries temporais. Dissertação (Mestrado em Engenharia Elétrica e de Computação). Universidade Estadual de Campinas, 2006.
- SANTHISREE, K; DAMODARAM, A. SSM-DBSCAN and SSM-OPTICS:** Incorporating a new similarity measure for density for density based clustering of web usage data. *International Journal on Computer Science and Engineering (IJCSE) ISSN:0975-3397 Vol. 3 No. 9 september 2011.*
- SENRA, L. F. A. C.; NANJI, L. C.; MELLO, J. C. C. B. S.; MEZA, L. A.** Estudo sobre métodos de seleção de variáveis em DEA. *Pesquisa Operacional*, v.27, n.2, p.191-207, Maio a Agosto de 2007.
- STROIEKE, R. E.; FOGLIATTO, F. S.; ANZANELLO, M. J.** Formação de agrupamentos homogêneos de trabalhadores através de curvas de aprendizado. Pré-anais XLII SPBO, Ubatuba, 2010.
- SZAJUBOK, N. K.; MOTA, C. M. M.; ALMEIDA, A. T.** Uso do método multicritério Electre Tri para classificação de estoques na construção civil. *Revista Pesquisa Operacional*, v. 26, n.3, p. 625-649, Set/Dez 2006.
- YIN, R. K.** Estudo de Caso: planejamento e métodos. 2.ed. Porto Alegre: Bookman, 2001.
- WANKE, P. F.** Metodologia para gestão de estoques de peças de reposição: um estudo de caso em empresa brasileira. *Revista Tecnológica. Centro de Estudos em Logística, Coppead, UFRJ. Dez., 2005.*
- WANKE, P. F.** The impact of different demand allocation rules on total stock levels. *Revista Pesquisa Operacional*, v. 30, n.1, p. 33-52, Jan/Abr 2010.

4 CONSIDERAÇÕES FINAIS

O aprimoramento de processos produtivos é uma preocupação antiga e constante. Com a crescente exigência do mercado consumidor e ampliação de acesso a tecnologias e metodologias de gestão, a concorrência empresarial torna-se cada dia mais acirrada. Detalhes antes negligenciados tornam-se diferenciais competitivos, demandando o desenvolvimento de abordagens que ofereçam suporte às decisões administrativas. Neste contexto, esta dissertação buscou propor uma sistemática que aborda elementos-chave dos processos produtivos: programação da produção e gestão de estoques.

No primeiro artigo foi apresentada uma sistemática que combina métodos de seleção de variáveis para clusterização integrados à Simulação de Monte Carlo (SMC), com vistas à simplificação na programação de produção de um cenário caracterizado por um elevado volume de produtos. Para tanto, foram criadas famílias (clusters) de produtos com características semelhantes apoiando-se em abordagens de seleção de variáveis. Os agrupamentos foram operacionalizados através do algoritmo não-hierárquico *k-means*, e a qualidade dos agrupamentos gerados auditada por especialistas e pelo Índice Silhouette. A partir das famílias geradas, foram simulados cenários produtivos por intermédio da SMC visando maximizar o lucro das vendas dos grupos de produtos. Dentre outras conclusões, a sistemática mostrou que, em situações de restrições computacionais, o uso de opinião de especialistas e indicadores de dispersão é válido para pré-seleção de variáveis. Também ficou evidente que a aplicação de SMC a produtos clusterizados pode levar a resultados satisfatórios, no que tange a decisões de programação diária.

No segundo artigo foi abordada a questão de gestão de estoques. Para tanto, inicialmente foi aplicada uma sistemática para seleção de variáveis de clusterização visando à formação de grupos de produtos com características similares; na sequência, tais grupos foram integrados a políticas simultâneas de gestão de estoques. Os custos de guarda e processamento de pedidos derivados da política simultânea foram comparados aos obtidos pelo Lote Econômico de Compras (LEC). A sistemática permitiu reduzir 76 produtos a três clusters. Para cada cluster foram geradas três políticas simultâneas de reposição de produtos diferentes, fazendo com que os custos de guarda e pedidos foram incrementados em 6,2%. Percebem-se, portanto, ganhos em termos de processamento e recebimento de pedidos, visto que a redução de quase 90% do volume de pedidos anuais compensa o incremento de custos.

Nesta dissertação, as sistemáticas abordaram a programação da produção e gestão de

estoques de forma independente. Fica clara a possibilidade de avaliar uma sistemática que trate de ambas de forma integrada em estudos futuros. Também parece oportuno incluir questões relativas a custo de processamento de pedidos e guarda de estoque, assim como incorporação de capacidades de armazenagem e produção. Em relação à etapa de seleção de variáveis, desdobramentos futuros incluem o desenvolvimento de um índice de importância das variáveis que guie a eliminação das mesmas.