

FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA” CENTRO
UNIVERSITÁRIO EURÍPIDES DE MARÍLIA – UNIVEM CURSO DE
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

FELIPE MARQUES BESSA DEANIN

APLICANDO A LINGUAGEM R PARA O MODELO S&OP

MARÍLIA
2015

FELIPE MARQUES BESSA DEANIN

APLICANDO A LINGUAGEM R PARA O MODELO S&OP

Trabalho de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Fundação de Ensino “Eurípides Soares da Rocha”, mantenedora do Centro Universitário Eurípides de Marília – UNIVEM, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador:
Prof. Geraldo Pereira Junior

MARÍLIA
2015

Marques Bessa Deanin, Felipe

Aplicando a Linguagem R para o modelo S&OP / Felipe Marques Bessa Deanin; orientador: Prof. Geraldo Pereira Junior. Marília, SP: [s.n.], 2015.

50 folhas

Monografia (Bacharelado em Sistemas de Informação): Centro Universitário Eurípides de Marília.

CDD: 658.4038



CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA - UNIVEM
MANTIDO PELA FUNDAÇÃO DE ENSINO "EURÍPIDES SOARES DA ROCHA"

BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Felipe Marques Bessa Deanin

Aplicando a Linguagem R para o modelo S&OP.

Banca examinadora da monografia apresentada ao Curso de Bacharelado em Sistemas de Informação do UNIVEM/F.E.E.S.R., para obtenção do Título de Bacharel em Sistemas de Informação.

Nota:

8,5 (oitoe meio)

Orientador:Geraldo Pereira Junior

1º.Examinador:Giulianna Marega Marques

2º.Examinador:Jorge Luiz Barbosa Maciel Junior

Marília, 02 de dezembro de 2015.

*Dedico este trabalho ao meu pai, Claudenir Deanin, e à
minha mãe, Solange Deanin.*

AGRADECIMENTOS

Agradeço aos meus pais, Claudenir e Solange, e à minha namorada, Bárbara, por toda a dedicação e esforço para que eu consiga atingir meus objetivos, lutar por meus sonhos e, principalmente, por não me deixar desistir diante das dificuldades.

Ao Professor Geraldo, por aceitar a orientação deste trabalho e realizá-la com excelência; por toda paciência, dedicação e por acreditar em meu potencial para a realização desta tarefa.

À Professora Giulianna, por todos os ensinamentos e pelas sugestões que, sem dúvida, muito contribuíram para esta pesquisa.

Agradeço, enfim, a todos os amigos de sala; em especial: Everton, Fernando, Guilherme Plaza, Carol, pelo encorajamento, pela ajuda e pelo compartilhamento de experiências em cada momento de dificuldade e, também, nos inúmeros instantes de alegria.

*“Tudo posso naquele que me fortalece.” (Filipenses
4,13)*

MARQUES BESSA DEANIN, Felipe. Aplicando a Linguagem R para o Modelo S&OP. 2015. 49f. Trabalho de Curso (Bacharelado em Sistemas de Informação) – Centro Universitário Eurípides de Marília, Fundação de Ensino “Eurípides Soares da Rocha”, Marília, 2015.

RESUMO

Em tempos modernos, que exigem, cada vez mais, objetividade e resultados efetivos, torna-se fundamental que as organizações tenham um bom planejamento de vendas e operações. Para isso, a organização deve entender e visualizar sua demanda de forma precisa, para que, futuramente, em suas decisões, na realização do seu planejamento de vendas, consiga utilizar tal informação aliada às outras informações, conquistando, assim, o melhor balanceamento entre a sua demanda e sua oferta. Devido ao grande volume de dados gerados e o tempo que tende a ser mais escasso na rotina das organizações, nota-se a inviabilidade de se analisar referidos dados apenas com uma análise humana. O presente trabalho, portanto, tem como o objetivo principal processar esses dados gerados, de uma forma rápida e de fácil entendimento, atendendo a uma resposta eficiente e eficaz, de tal maneira que a organização não precise disponibilizar um grande círculo de funcionários, nem precise buscar um auxílio especializado fora dela para atender à tarefa. Logo, será utilizada a linguagem de programação R Shiny na criação de uma interface de fácil utilização, em que serão executados os métodos estatísticos no software R Language, em especial a correlação de Pearson, sendo plotadas as informações lapidadas em um gráfico, a fim de que seja possível uma análise visual mais facilitada.

Palavras-chave: R language, Sales and Operations Planning, Planejamento, S&OP, Shiny, Pearson.

MARQUES BESSA DEANIN, Felipe. Aplicando a Linguagem R para o Modelo S&OP. 2015. 49f. Trabalho de Curso (Bacharelado em Sistemas de Informação) – Centro Universitário Eurípides de Marília, Fundação de Ensino “Eurípides Soares da Rocha”, Marília, 2015.

ABSTRACT

In modern times requiring increasingly objectivity and effective results, it becomes critical as organizations who have good sales and planning operations. For this, one organization must understand and view your demand form of need, for what future in your decision paragraph the your sales and planning director, can use this information allied at additional information to achieve the best balance between your demand is his offer. The large volume of data generated and the due time every time more scarce organizations, notes an impracticality of analyzing these data only with a human analysis. This work has as the main objective, processes data generated from a quick way and understanding easy and that has a response efficient and effective, where the organization not provide accurate a large circle of officials and not get accurate hum aid specialized para forums undertake that work. For it, will used language programming R Shiny paragraph make an creating an easy-to-use interface, where this, perform statistical methods in R Language software, especially the Pearson Correlation and we will plot as faceted information in a chart, paragraph a closer look easy.

Keywords: R Language, Sales and Operations Planning, Planning, S&OP, Shiny, Pearson.

LISTA DE ILUSTRAÇÕES

Figura 1 – Período de congelamento (time fence).....	20
Figura 2 – Processo Mensal do <i>Sales & Operations Planning</i>	22
Figura 3 – Fluxograma do papel da estatística na tomada de decisão gerencial.....	29
Figura 4 – Exemplo Peso x Altura.....	30
Figura 5 – Diferença de Escalas	31
Figura 6 – Fórmula do Coeficiente de Correlação de Pearson.....	31
Figura 7 – Classificação do Resultado (r) do Coeficiente de Correlação de Pearson.....	32
Figura 8 – Equação de Regressão.....	33
Figura 9 – Coeficiente Angular	33
Figura 10 – Ponto de Interceptação.....	33
Figura 11– Regressão Múltipla.....	34
Figura 12 – Etapas do KDD.	35
Figura 13 – Parte dos dados das Vendas mensais dos produtos	36
Figura 14 – Código user-interface.....	38
Figura 15 – Código server R	39
Figura 16 – Galeria de Aplicações do Shiny.....	40
Figura 17 – Dez maiores correlações entre os produtos	42
Figura 18 – Regressão Linear Múltipla no R	42
Figura 19 – Cálculo Regressão Linear produto X6.	43
Figura 20 – Cálculo Regressão Linear produto X8.	44
Figura 21 – Cálculo Regressão Linear produto X10.	44
Figura 22 – Cálculo Regressão Linear produto X14.	44
Figura 23 – Cálculo Regressão Linear produto X18.	45
Figura 24 – Gráfico com as Previsões de alguns produtos	47

LISTA DE TABELAS

TABELA 1	Participantes sugeridos do processo de <i>S&OP</i>
TABELA 2	Classificação do Coeficiente de Correlação
TABELA 3	Servidores Shiny
TABELA 4	Resultados Obtidos

LISTA DE ABREVIATURAS E SIGLAS

S&OP	Sales and Operations Planning
GNU	General Public License
UI	User Interface

SUMÁRIO

INTRODUÇÃO	14
1 PLANEJAMENTO DE VENDAS E OPERAÇÕES.....	16
1.1 Processos do Planejamento de Vendas e Operações.....	22
2 LINGUAGEM R.....	26
3 ESTATÍSTICA.....	28
3.1 PROBABILIDADE	30
3.1.1 Conceitos Fundamentais	30
3.1.2 Método Correlação	31
3.1.3 Diagramas de Dispersão	31
3.1.4 Regressão Linear Múltipla	34
4 KNOWLEDGE DISCOVERY IN DATABASES (KDD).....	36
4.1 Mineração de Dados.....	37
5 RSTUDIO Shiny	39
6 DESENVOLVIMENTO DA APLICAÇÃO	42
7 RESULTADOS OBTIDOS	46
CONCLUSÃO	47
REFERÊNCIAS.....	48

INTRODUÇÃO

Durante muito tempo, as empresas, em razão da grande concorrência do mercado, vêm buscando melhoria no seu planejamento, visando, sobretudo, ao menor custo de produção aliado à confecção de um produto bom, barato e acessível. Em razão disso, portanto, a maioria das organizações, a fim de conseguir informações necessárias para seu planejamento, reúnem seus colaboradores de mais alto nível e indagam a esses sua opinião sobre as projeções da empresa para os próximos meses; essa tática, contudo, nem sempre ajuda muito o andamento do planejamento, pois, em geral, cada colaborador se pauta em um referencial e isso, ao invés de favorecer os passos seguintes, no fundo, prejudica as representações futuras, visto, após a rodada de opiniões, duas perguntas sempre ficam no ar: 1) Em qual número se deve acreditar? 2) Quais métodos, ou cálculos, serão utilizados para se chegar a um número mais aproximado possível da realidade? Observando essa dificuldade das organizações, o intuito deste trabalho, então, é automatizar o processo para se atingir, tanto quanto mais adequadamente, a esse número estimado de vendas. Para tanto, será empregado o modelo de negócio S&OP, que realiza a coleta de informações pela gerência junto à alta diretoria a organização. Durante tal processo, deve-se introduzir essas informações no ambiente de programação estatística R Language, a fim de que realizem os cálculos, a partir do método de Correlação, e, por fim, seja gerado um gráfico para a visualização, análise e entendimento mais facilitados, por meio de uma interface de clara utilização. Assim, espera-se reduzir o tempo para que os diretores, junto com a alta gerência, tomem suas decisões em relação ao planejamento da organização de uma forma geral.

Motivação e Justificativa

Com a finalização do projeto, espera-se divulgação de um protótipo em que um usuário consiga inserir os seus dados e gerar um gráfico para a tomada de decisão, sem necessidade de um alto conhecimento de programação, bem como de conhecimento estatístico, diminuindo, pois, o tempo e o dinheiro gasto com profissionais externos para a realização desse processo. Justifica, ainda, tal motivação o tempo gasto, fora de sua função habitual, que o funcionário empregará para realizar a coleta e mineração dos dados para os transformar em informações relevantes; afinal, tempo é dinheiro.

O público-alvo deste trabalho são empresas de pequeno e médio porte que necessitem de informação rápida e confiável para realizar seu planejamento e que não tenham recursos e mão-de-obra qualificada para realizar, manualmente, essa tarefa.

Objetivo Geral

O objetivo geral deste trabalho é o estudo do método de mineração de dados KDD (*Knowledge Discovery in Databases*), para transformar os dados obtidos da empresa em informações. Após a transformação do dado, este será estudado e utilizado a partir do método de Correlação de Pearson. Tal método implica desenvolver análises estatísticas sobre a informação já lapidada e, por fim, propor um protótipo que promova o cálculo estatístico e retorne uma previsão de vendas para os meses futuros, desde o método de regressão múltipla, plotado em gráficos, para que, assim, subsista uma forma mais facilitada que permita aos responsáveis entenderem e repassarem as informações ali contidas, apresentando a seus diretores e/ou gerentes, permitindo nesses o entendimento estatístico para extração da informação repassada.

Objetivos Específicos

Os objetivos específicos envolvem a diminuição do tempo que utilizado para fazer a previsão de vendas de um grupo que possua, em média, 100 vendedores. Também se busca auxiliar a organização, com uma redução de seus produtos em estoque. Todo esse processo tende a ser amparado por técnicas de Correlação e Regressão Múltipla.

Organização do Trabalho

A pesquisa em tela tem como suporte revisão bibliográfica que fundamenta, teoricamente, bem como aprofunda o tema abordado.

Conjuntamente, o trabalho está distribuído em 7 capítulos organizados da seguinte forma: o Primeiro Capítulo (Planejamento de Vendas e Operações) contém a abordagem geral em torno do tema, os objetivos da pesquisa e alguns pilares do processo analisado.

Já o Segundo Capítulo (Linguagem R) conduz a um breve conceito sobre a Linguagem R, quais são suas modalidades estatísticas, entre outras características da linguagem.

O Terceiro Capítulo (Estatística) volta-se ao conceito sobre a Estatística. Após serem enfocados termos básicos, será abordado o conceito de Probabilidade e, também, seus termos básicos; por último, virá a Correlação de Pearson.

Quanto ao Quarto Capítulo (*Knowledge Discovery in Databases*), o interesse estará no método de mineração de dados.

O Quinto Capítulo (RStudio Shiny) constitui a demonstração do funcionamento do framework Shiny, produzido pela RStudio; há, ainda, uma breve introdução à criadora deste.

No penúltimo capítulo, o Sexto (Desenvolvimento da Aplicação), ocorre o desenvolvimento do protótipo e a explicação do funcionamento do mesmo, com prints das telas principais do recurso.

E, finalmente, o Sétimo Capítulo (Resultados Obtidos) responderá pelos resultados

obtidos ao final da pesquisa.

1 PLANEJAMENTO DE VENDAS E OPERAÇÕES

Planejamento de Vendas e Operações é um termo que vem do inglês *Sales and Operations Planning* (S&OP); trata-se de um processo que auxilia nas tomadas de decisões da empresa para que a mesma consiga balancear a sua demanda e oferta.

Para uma definição mais resumida, segundo Wallace (2001), o Planejamento de Vendas e Operações é um processo aprimorado de tomada de decisões que ajuda as pessoas, nas empresas, a proporcionarem excelente atendimento aos clientes e a conduzirem melhor os negócios (WALLACE, 2001).

E, sob uma definição mais completa, ainda o autor oferece a seguinte explicação:

O Planejamento de Vendas e Operações (*S&OP*) é um processo empresarial que ajuda as empresas a manterem a demanda e a oferta balanceadas. Isto é feito através do enfoque nos volumes agregados (famílias e grupos de produtos) de modo que os problemas de mix (produtos individuais e pedidos de clientes) possam ser controlados mais prontamente. Ele ocorre num ciclo mensal e apresenta informações tanto em unidades como em reais. O S&OP é multidisciplinar, envolvendo a Gerência Geral, Vendas, Operações, Finanças, e Desenvolvimento de Produtos. Ele envolve múltiplos níveis dentro da empresa até o executivo responsável pela unidade de negócios inclusive, por exemplo, o presidente da divisão, gerente-geral da unidade de negócios, ou o diretor de uma pequena empresa. O S&OP interliga os planos estratégicos e o plano de negócios da empresa aos processos - a entrada do pedido, a programação e as ferramentas de compras que ele utiliza para a condução dos negócios em uma base semanal, diária e horária. Utilizado adequadamente, o S&OP capacita os gerentes das empresas a visualizarem os negócios holisticamente e dar-lhes uma janela para o futuro. (WALLACE, 2001)

E, complementando, observa que a missão do S&OP é o balanceamento da demanda com a oferta em nível de volume, de tal modo que esse volume se refere aos índices globais de vendas, índices de produção, inventários agregados e pedidos pendentes ou em atraso (WALLACE, 2001).

Objetivos do S&OP

Em Corrêa, Gianesi & Caon (2001), a execução eficaz do processo de S&OP só é possível se os objetivos a seguir forem alcançados. São eles:

- **Suportar o planejamento estratégico do negócio**

É a garantia, por meio de análises e revisões periódicas, de que o planejamento feito é viável e está sendo cumprido. Normalmente, esse planejamento estratégico

é expresso em moeda, o que torna mais complicado de outras áreas da empresa conseguirem entender o que esse planejamento está querendo passar. O autor pondera que o S&OP deve fornecer a ligação entre o plano estratégico de negócio e as diversas áreas da empresa, garantindo que o plano de cada área esteja em sincronia com o plano estratégico (CORRÊA, GIANESI & CAON, 2001).

- **Garantir que os planos sejam realísticos**

Como se sabe, algumas áreas da empresa dependem de outras áreas; assim, somente pode-se garantir a viabilidade plano de uma determinada área se as demais áreas da empresa validarem este plano por meio da participação na tomada de decisões (CORRÊA, GIANESI & CAON, 2001).

- **Gerenciar as mudanças de forma eficaz**

Deve-se tornar ativo diante das mudanças futuras (exemplo: mudança de foco no mercado, alteração do volume da produção); ou seja, deve-se analisar o impacto de cada mudança, nas demais áreas da empresa, para que estas aconteçam dentro do prazo esperado (CORRÊA, GIANESI & CAON, 2001).

- **Gerenciar os estoques de produtos finais e/ou carteira de pedidos, de forma a garantir bom desempenho de entregas**

É fundamental que a gestão dos níveis de estoque e da carteira de pedidos seja assumida pelo diretor Geral. O objetivo do S&OP é agir sobre as vendas e a produção para que haja um equilíbrio, a fim de evitar grandes volumes em estoque, tampouco para que não haja falta de produtos, deixando de atender ao cliente (CORRÊA, GIANESI & CAON, 2001).

- **Avaliar o desempenho**

Deve-se utilizar o S&OP para identificar quanto o desempenho real de cada área se desviou do plano. Assim, consegue-se distinguir as áreas que estão dentro do controle e as áreas que saíram deste controle, visando uma atenção da área gerencial para que tenha ações para corrigir a área que está fora do plano (CORRÊA, GIANESI & CAON, 2001).

- **Desenvolver trabalho em equipe**

O S&OP deve criar condições para que todas as áreas da empresa participem da criação do planejamento total da empresa. O autor avalia que o caráter de participação e negociação para a tomada de decisões é um dos pontos-chave do S&OP (CORRÊA, GIANESI & CAON, 2001).

Definições de uma política de S&OP

A política de S&OP é um documento que irá definir todos os objetivos, as características e os principais parâmetros do planejamento da empresa. Importante considerar ser de extrema importância que haja consenso e comprometimento de todos com a política do S&OP. Dentro desta política, Corrêa, Giansesi & Caon (2001) separam alguns tópicos relevantes que deverão estar contidos neste documento. A saber:

- **Objetivo**

O objetivo específico a ser atingido pela empresa com a implantação do processo de S&OP. Aqui, sempre necessário recordar que o processo deve ser avaliado e aprimorado continuamente, para que a empresa possa avaliar o cumprimento de seus objetivos e, em caso de necessidade, recorrer ao direcionamento correto para a aprimoração do processo (CORRÊA, GIANESSEI & CAON, 2001).

- **Processo**

O processo do S&OP é um composto de atividades preparatórias que deverão ser executadas por cada departamento, a fim de sinalizar para a alta direção, avaliar a capacidade de cada um e criar linhas mestras para esse processo (CORRÊA, GIANESSEI & CAON, 2001).

- **Cronograma**

Devem-se definir datas limites para as atividades do ciclo mensal, criar a programação das reuniões para a alta direção. Tais práticas evitam que, futuramente, haja problemas com superposição a outros compromissos (CORRÊA, GIANESSEI & CAON, 2001).

- **Participantes**

Deverão ser definidos aqueles que participarão da reunião de S&OP; note-se que cada participante deve ser assíduo. Caso haja motivos que inviabilizem a presença de um membro, deverá ser passada sua autonomia a outra pessoa substituta. Corrêa, Giansesi & Caon (2001), ainda, salientam tais informações, a partir de uma tabela dos participantes sugeridos para o

processo de S&OP:

Tabela 1: Participantes sugeridos do processo e S&OP

ÁREA	PARTICIPANTES	POTENCIAIS PARTICIPANTES
Alta Administração	Diretor Geral	–
Vendas	Diretor de Vendas	Gerente de Vendas Gerente de Atendimento ao Cliente Gerente de Distribuição Gerente de Assistência Técnica
Marketing	Diretor de Marketing	Gerente de Produto Gerente de Demanda
Manufatura	Diretor de Manufatura	Gerente Industrial, de Produção ou de Fábrica Gerente de Suprimentos Gerente de Garantia da Qualidade
Engenharia ou P&D	Diretor de Engenharia ou P&D	Gerente de Desenvolvimento Gerente de Métodos e Processos
Finanças	Diretor de Finanças	Gerente de orçamento Gerente de Custos
Recursos Humanos	Diretor de Recursos Humanos	Gerente de Pessoal

Fonte: CORRÊA, GIANESI & CAON, 2001.

- **Definições das famílias de produtos**

Deve-se definir o grupo ou as famílias dos produtos para o planejamento no nível S&OP, o que tornará mais fácil para a alta direção analisar a família de produtos e não cada produto por si (CORRÊA, GIANESI & CAON, 2001).

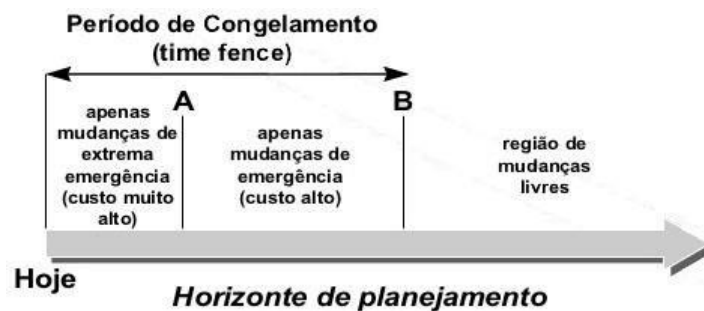
- **Horizonte de Planejamento**

Este será definido em função do recurso crítico, considerando a variável tempo. O horizonte de planejamento deve ser bastante longo para que consiga analisar perguntas como “O que aconteceria se...?”. Normalmente, é realizado para o decurso de um ano (CORRÊA, GIANESI & CAON, 2001).

- **Time Fences**

São os intervalos de tempo dentro do horizonte de planejamento, que será explicado pela seguinte figura:

Figura 1: Período de congelamento (time fence)



Fonte: CORRÊA, GIANESI & CAON, 2001.

Após essa breve introdução aos objetivos do S&OP e aos componentes de uma política para o processo de S&OP, serão indicadas novas informações igualmente essenciais.

Logo, Corrêa, Gianesi & Caon (2001) afirmam que, para que o processo ocorra e tenha um bom desempenho, são fundamentais as informações disponíveis para a tomada de decisão; no caso do S&OP, é imprescindível ressaltar itens que merecem uma atenção especial. São eles:

- **Desempenho Passado**

O autor estabelece que um dos aspectos mais importantes para o S&OP é a análise do desempenho passado. A redução dos desvios em relação ao plano depende de uma análise cuidadosa sobre o desempenho passado, identificando as causas dos desvios e avaliando as alternativas para que esses desvios sejam diminuídos. Nessas condições, um planejamento com olhos apenas para o futuro conduz ao risco de que os erros de desvios continuem (CORRÊA, GIANESI & CAON, 2001).

- **Estado atual**

É o levantamento do estado atual em relação à produção, às vendas, ao estoque e à carteira; este pode representar um problema para empresas com pouca estrutura, apresentando informações desencontradas sobre o mesmo tema ou dados diferentes sobre o mesmo evento. É eficaz que as informações sejam analisadas com o maior cuidado possível (CORRÊA, GIANESI & CAON, 2001).

- **Parâmetros**

São os dados que permitem transformações; por exemplo: a quantidade que deverá produzir em relação à necessidade ou capacidade. Observar que a

dificuldade destes parâmetros são o fato de que há uma variação em função do mix de produtos. Assim, devem-se analisar, periodicamente, esses parâmetros para que estejam andando em paralelo com a realidade (CORRÊA, GIANESI & CAON, 2001).

- **Previsões**

São a alma de qualquer processo de planejamento. No caso do S&OP, as previsões são necessárias de médio e longo prazo e, mesmo que ainda se utilize de agrupamento dos produtos em famílias de produtos, poucos ficam confortáveis para fazer essas previsões. O autor atenta, ainda, para o fato de que alguns aspectos sustentam a mudança da visão de pensamentos do tipo “é impossível fazer previsão de longo prazo”; ei-los:

- Considerando-se que existem recursos que necessitam de um longo tempo para que possam ser disponibilizados (por exemplo, mão-de-obra capacitada para tal tarefa, implantar novas linhas de produção, etc.), não tem como não se pensar na necessidade deste recurso a longo prazo, para que se possa tomar uma decisão hoje quanto a adquirir (ou não) este recurso. Adiar a decisão sobre este recurso, por causa de incertezas do futuro, significa não adquiri-lo no momento. O autor reflete, além disso, que, se por um lado o futuro da empresa depende das decisões tomadas hoje, por outro, tais decisões dependem da melhor visão do futuro que se tem hoje (CORRÊA, GIANESI & CAON, 2001).
- Qual nível seria bom em termos de acerto da previsão: 50%, 70% ou 90%? Sabe-se que não se tem como definir qual nível de acerto é o melhor; o autor sopesa que a previsão necessita ser tão boa ou melhor que a de seu concorrente, dependendo de seu processo de planejamento e tomada de decisão. Sabe-se, também, que fazer a previsão, independentemente do nível de acerto, é melhor do que fazer planejamento e tomar decisões no famoso “eu acho” (CORRÊA, GIANESI & CAON, 2001).
- É comum se ouvir que previsões a longo prazo são previsões inúteis, pois um grande fator dessas previsões é o mercado, que, muitas vezes, é inconstante. Muitas empresas que buscam melhorar suas previsões sabem que os fatores externos têm uma grande porcentagem de influência em uma previsão de longo prazo e, por isso, descobriram que a outra parte, também expressiva, deriva de sua organização interna, onde falta competência para

fazer estas previsões (CORRÊA, GIANESI & CAON, 2001).

Resultados do S&OP

O plano resultante do *Sales and Operations Planning*, quando transformado para as diversas áreas da empresa, mostra os objetivos específicos de cada uma: a manufatura deve atingir o plano de produção, a área financeira deve prover s recursos do orçamento e a área de vendas, atingir o plano de vendas. Assim definidos, os objetivos estarão, por consequência, homogêneos e integrados aos objetivos da empresa; ou seja, providenciam medidas para que o desempenho de cada área da empresa seja avaliado, com a observação de que esta atividade de avaliação deve estar ligada ao processo de planejamento (CORRÊA, GIANESI & CAON, 2001).

1.1 Processos do Planejamento de Vendas e Operações

Corrêa, Giansesi & Caon (2001) analisam que o processo do S&OP consiste em cinco etapas sucessivas; a saber:



Fonte: CORRÊA, GIANESI & CAON, 2001.

Resumidamente, a Figura 3 demonstra os cinco passos citados pelos autores. O primeiro deles (*levantamento de dados*) compreende o levantamento dos dados históricos, que suportam os dados atuais da empresa em relação a vendas, produção, estoque e, também, indicam o desempenho passado em relação a esses e a outros aspectos. Já o segundo,

planejamento de demanda, inclui a gestão das previsões e a elaboração do plano de vendas. Continuando, o *planejamento de produção* (materiais e capacidades) deriva uma *reunião preliminar de S&OP*, onde estarão configurados todos os outros setores da empresa para analisar os planos e identificar os problemas e suas alternativas. Por fim, a *reunião executiva de S&OP*, em que os planos definidos na reunião preliminar serão validados junto com a alta direção (CORRÊA, GIANESI & CAON, 2001).

Vendas e marketing são responsáveis pela comparação da demanda real passada, com o plano de vendas, verificando o potencial do mercado e projetando a demanda futura. Após este passo, o plano atualizado de demanda é enviado à manufatura, que, por sua vez, irá elaborar o plano de produção, verificando as necessidades de capacidade; observe-se que todas as dificuldades de atender à demanda são neutralizadas ou, então, o plano de vendas deverá ser alterado. O resultado será um plano atualizado de operações que atenderá à demanda; assim, este plano será desagregado para dar origem ao plano mestre de produção (CORRÊA, GIANESI & CAON, 2001).

Quanto mais rápido o ciclo de planejamento for executado, mais rápida também será a resposta da empresa em relação às mudanças do mercado. Contudo, essa rapidez não poderá comprometer que cada passo do processo seja executado adequadamente, pois a decisão final sairá deste ciclo. Corrêa, Giansesi & Caon (2001) ponderam que a redução de tempo do ciclo vem com o aprimoramento do processo, à medida em que estes forem repetidos mensalmente. E concluem que o importante é conseguir que o processo (e sua duração) sejam firmes, mês a mês, e que todas as áreas da empresa, estejam comprometidas na aprimoração destes processos, recolhendo e analisando informações, assim, elaborando suas recomendações o mais rápido possível, para que a reunião executiva seja realizada, também, tanto quanto mais rapidamente.

Levantamento de Dados / Preparação dos Dados Históricos

Após a definição da família de produtos e dos aspectos que se deseja analisar no processo de S&OP, é necessário levantar mecanismos de preparação destes dados que serão utilizados no planejamento. Na maioria das vezes, tais elementos não serão encontrados no formato que se deseja para o processo de decisão, pois os produtos estão desagregados, bem como alguns deles saem de produção ou sofrem alterações. Os parâmetros de tempo devem ser revistos periodicamente, não somente pelo fato da mudança nas famílias de produtos, mas, também, por conta da variação do mix de produção e vendas (CORRÊA, GIANESI & CAON, 2001). Essa etapa é fundamental para o processo de S&OP, pois é a partir destes itens que serão

realizadas as análises para a tomada de decisão ao longo de todo o processo. Toda área da empresa deverá ter um responsável, especificamente, para a preparação dessas informações, sendo esta tarefa tomada como atividade do seu dia-a-dia; haverá, ainda, datas mensais agendadas para a entrega desses resultados (CORRÊA, GIANESI & CAON, 2001).

Planejamento de Demanda

O Planejamento de Demanda, conforme mencionado, em geral, é de responsabilidade do marketing e da área de vendas e consiste na elaboração de um plano de vendas bastante chamativo, capaz de indicar o que a empresa estará disposta a vender e oferecer ao mercado, para cada família de produto, mensalmente, ao longo do planejamento.

CHOPRA (2011), separa a demanda de produtos em duas partes: ou seja, produtos com demanda estável e produtos cuja demanda muda constantemente.

Ainda segundo referido autor, para produtos com a demanda mais estável, necessário desenvolver um plano de oferta, relativamente mais simplificado. Isto é: a organização providencia sua capacidade para atender à determinada demanda esperada, produzindo a quantidade que corresponde a esta demanda. Nesse caso, os produtos são elaborados próximos à época em que serão comercializados e a cadeia de suprimentos, a seu turno, mantém o estoque reduzido. Já quanto aos produtos com uma demanda mais instável, o estudioso classifica esta mudança como constante, em razão de mudança previsível (CHOPRA, 2011).

Sob tal razão de mudança previsível, ainda Chopra afirma que é causada por fatores sazonais (por exemplo, jaquetas de esqui) e fatores não sazonais (por exemplo, promoções), acrescentando que tal dinâmica pode causar aumentos e declínios previsíveis nas vendas (CHOPRA, 2011). Para o autor, portanto:

Variabilidade previsível é a mudança na demanda que pode ser prevista. (CHOPRA, 2011)

Continuando com SUNIL & CHOPRA (2011), o objetivo do planejamento de vendas e operações é de que a empresa combine, o mais corretamente possível, duas opções gerais para trabalhar com a variável previsível; são elas:

1. Gerir a oferta, usando capacidade, estoque, subcontratação e pedidos em atraso; e
2. Gerir a demanda, usando descontos a curto prazo e promoções comerciais.

Para CHOPRA (2011), o uso dessas duas ferramentas permite que a cadeia aumente sua lucratividade, pois a oferta e a demanda estão bem equilibradas. E prossegue observando

que as empresas, normalmente, dividem a gestão de oferta e demanda em funções distintas, de tal forma que: o marketing, habitualmente, gere a demanda e a operação responde pela oferta. Essa distinção, contudo, pode causar problemas, pois, como mencionado, o objetivo é gerir a oferta e a demanda do modo mais equilibrado possível (CHOPRA, 2011).

Planejamento de Produção

O Planejamento de Produção é de responsabilidade da área de manufatura e apoiado pela produção e suprimentos. O objetivo do planejamento de produção é preparar (um ou mais) planos alternativos de produção para cada família de produtos, procurando atender à demanda, que, por sua vez, está representada pelo plano de vendas. Também busca gerar níveis desejados de estoque, esses expressos pela política de cada família de produtos, que sejam viáveis em termos de capacidade e, também, em termos de materiais críticos. Para Corrêa, Giansi & Caon (2001), a ideia de se criar planos alternativos ocorre porque, raramente, se consegue estabelecer um plano que atenda a esses três objetivos sem que haja contratempos para serem resolvidos.

2 LINGUAGEM R

A linguagem R (R Language) é uma linguagem e um ambiente voltado para computação estatística e gráfica, bem como uma plataforma de programação open source (plataforma aberta) para análise estatísticas e ambiente gráfico, desenvolvida no Bell Laboratories (antiga AT & T), por John Chambers (R Language, 2015).

“R” fornece várias modalidades estatísticas, que são: linear, modelagem não-linear, testes estatísticos clássicos, análises de séries temporais, classificação, clustering, etc.; e suas técnicas gráficas são altamente extensíveis (R Language, 2015).

Um dos pontos fortes da linguagem R é a facilidade com que se produz uma publicação de qualidade, incluindo símbolos matemáticos e fórmulas, quando necessárias; além disso, também o cuidado com que foram escolhidos os padrões de design menores em gráficos (sob os quais o usuário tem total controle) constitui um aspecto a ser destacado (R Language, 2015).

R, portanto, é um software livre sob os termos da GNU (General Public License, <http://www.gnu.org/>), que compila e roda em uma grande variedade de plataformas UNIX e sistemas semelhantes (incluindo FreeBSD e Linux), Windows e MacOS (R Language, 2015).

Ambiente R

R é um conjunto integrado de instalações de software para manipulação de dados, cálculos e exibição em gráficos. São constituição inclui:

- Tratamento eficaz dos dados e armazenamento fácil;
- Um conjunto de operadores para cálculo de uma estrutura de dados, matriz;
- Uma grande coleção de ferramentas intermediárias integradas e coerentes para análise de dados;
- Exibições gráficas em tela ou papel para análise de dados;
- Programação simples e eficaz (R LANGUAGE, 2015).

O termo “ambiente” é utilizado para caracterizá-lo como um sistema

totalmente planejado e coerente, sem a necessidade de incrementar ferramentas específicas e inflexíveis.

R destina-se ao redor de uma verdadeira linguagem computacional, permitindo que os usuários definam novas funções, de acordo com suas necessidades, tornando facilitada sua utilização. Para tarefas mais intensivas, pode-se integrar códigos, C, C++ e Fortran, e chamá-los em tempo de execução. Levantamento de Requisito (R Language, 2015).

3 ESTATÍSTICA

Para Stigler:

Estatística é o ramo da matemática aplicada cujo princípio deriva da teoria da probabilidade, que tem por objeto o agrupamento metódico assim como o estudo de séries de fatos ou de dados numéricos. (apud PEREIRA, 2014, p. 46)

Já em McClave (2009), a Estatística é a ciência dos dados, o que envolve coletar, classificar, resumir, organizar, analisar e interpretar a informação numérica.

Enquanto em Rao (apud BAYER et al, 2004), a estatística é uma ciência que estuda e pesquisa sobre: a) o levantamento de dados, com a máxima quantidade de informação possível para um dado custo; b) o processamento de dados para a quantificação da quantidade de incerteza existente na resposta para um determinado problema; c) tomada de decisões sob condições de incerteza, sob o menor risco possível (BAYER et al, 2009).

A estatística fornece diversas técnicas para extrair informações de dados, permitindo obter uma melhor compreensão das situações apresentadas.

A estatística, desse modo, é uma ciência que se preocupa com: a) Organização; b) Descrição; c) Análises; e d) Interpretações.

Detalhadamente, a organização e a descrição estão associadas à estatística descritiva; o terceiro e o quarto item voltam-se à estatística indutiva ou inferencial.

A estatística descritiva utiliza métodos numéricos e gráficos para detectar padrões em um conjunto de dados; ou seja, ela resume a informação revelada deste conjunto de dados, para apresentá-los de forma conveniente (MCCLAVE, 2009).

Por outro lado, a estatística inferencial ou indutiva utiliza-se de uma amostra de dados para fazer previsões, estimativas, acerca de um conjunto maior de dados (MCCLAVE, 2009).

Na sequência, são dispostos conceitos associados ao universo da Estatística:

População: É o conjunto de unidades (pessoas, objetos, transações ou eventos) que a ser estudado (MCCLAVE, 2004).

Amostra: É um subconjunto de unidades de uma população (MCCLAVE, 2009). Em outras palavras, é o agrupamento dos elementos que serão examinados para a realização

do estudo estatístico desejado.

Variável: É a característica dos elementos contidos na amostra que se deseja estudar estatisticamente (MCCLAVE, 2009). Pode-se classificar a variável em duas formas diferentes; a saber: a) **qualitativa:** quando a variável é classificada por tipos ou atributos; neste caso, não pode ser classificada em uma escala numérica (MCCLAVE, 2009); e **quantitativa:** quando a variável pode ser classificada em uma escala numérica de ocorrência natural (MCCLAVE, 2009).

Esta pesquisa empregará a estatística inferencial e dados quantitativos para realizar as previsões e estimativas da quantidade de vendas que a organização poderá ter nos seus próximos meses, tendo como base os dados de meses anteriores.

Após decidir qual o tipo de dado (quantitativo ou qualitativo), apropriados para este trabalho, é necessário que estes sejam coletados. Para isso, geralmente, existe 4 formas de coleta de dados: 1) por fonte publicada; 2) por estudo controlado; 3) a partir de pesquisa; ou 4) por meio de observação (MCCLAVE, 2009).

Dados de Fonte Publicada

São dados em que pesquisador já os coletou e disponibilizou em uma revista, jornal, livro, sites na Internet (MCCLAVE, 2009).

Dados de Estudo Controlado

São dados em que o pesquisador exerce um rígido controle sob a unidade de estudo (pessoas, objetos, eventos). Esse tipo de coleta permite que se extraia uma quantidade maior de informações sobre os dados, se comparado a um estudo não controlado (MCCLAVE, 2009).

Dados de Pesquisa

O pesquisador seleciona uma *amostra* de um grupo de pessoas e submete-os às perguntas, registrando suas repostas; a partir daí, pode retirar as informações de sua pesquisa. Um exemplo clássico deste tipo de coleta de dados é a pesquisa eleitoral, utilizada para “prever” a corrida eleitoral (MCCLAVE, 2009).

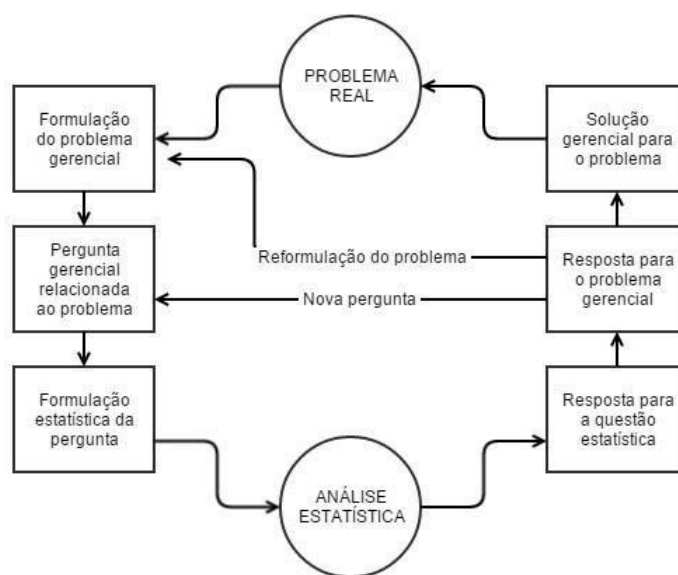
Dados coletados por meio de Observação

Aqui, o pesquisador observa unidades do experimento em seu ambiente natural e registra o(s) comportamento(s) que o interessa; é diverso do estudo comportamental, em que há o controle da amostra (MCCLAVE, 2009).

O papel da estatística no gerenciamento da tomada de decisões

Permanecendo com McClave (2009), o papel da estatística no gerenciamento da tomada de decisão encontra-se demonstrado pelo fluxograma a seguir:

Figura 3: Fluxograma do papel da estatística na tomada de decisão gerencial



Fonte: MCCLAVE, 2009.

Como se observa, o problema inicia-se no mundo real; tal problema, então, é formulado nos termos gerenciais como uma pergunta gerencial. Seguindo o fluxograma no sentido anti-horário, identifica-se o papel da estatística neste processo. Para responder à questão estatística, devem-se traduzir os problemas gerenciais para problemas estatísticos, coletar os dados da amostra e analisá-los, assim a questão será respondida. A resposta obtida servirá para resolver os problemas gerenciais ou poderá sugerir a reformulação do problema original.

3.1 PROBABILIDADE

3.1.1 Conceitos Fundamentais

Segundo TAVARES (2012), os conceitos de probabilidade são utilizados para estudar fenômenos aleatórios; ou seja: situações em que os resultados desses fenômenos

são conhecidos, porém não se sabe qual deles ocorrerá. Como exemplo, pode-se mencionar a ação de se jogar uma moeda para o alto. Sabe-se que cairá cara ou coroa, porém não é possível saber qual será, efetivamente, o resultado (TAVARES, 2012).

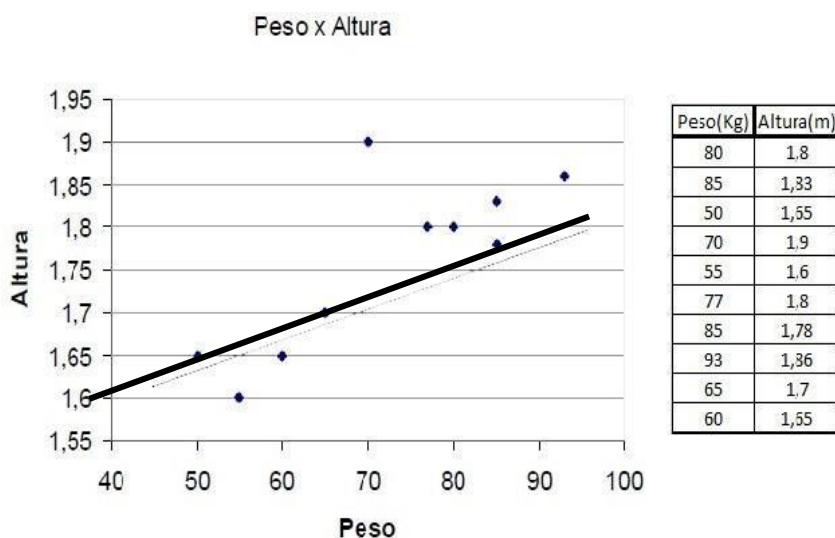
3.1.2 Método Correlação

Esta pesquisa utilizará o Método de Correlação, que emprega duas variáveis que estão relacionadas: ou seja, se ocorre uma mudança em uma variável (independente), também ocorrerá na outra (dependente). Por exemplo, a situação conhecida velocidade x consumo de combustível (TAVARES, 2012).

3.1.3 Diagramas de Dispersão

O Diagrama de Dispersão corresponde ao campo em que é mostrada a relação entre as variáveis que serão medidas sobre o mesmo indivíduo. Uma variável será posta sobre o eixo horizontal (eixo x) e a outra, conseqüentemente, será posta sobre o eixo vertical (eixo y). Para se aplicar o diagrama de dispersão, utilizam-se os valores das variáveis do indivíduo, colocando-as em um gráfico (TAVARES). Por exemplo:

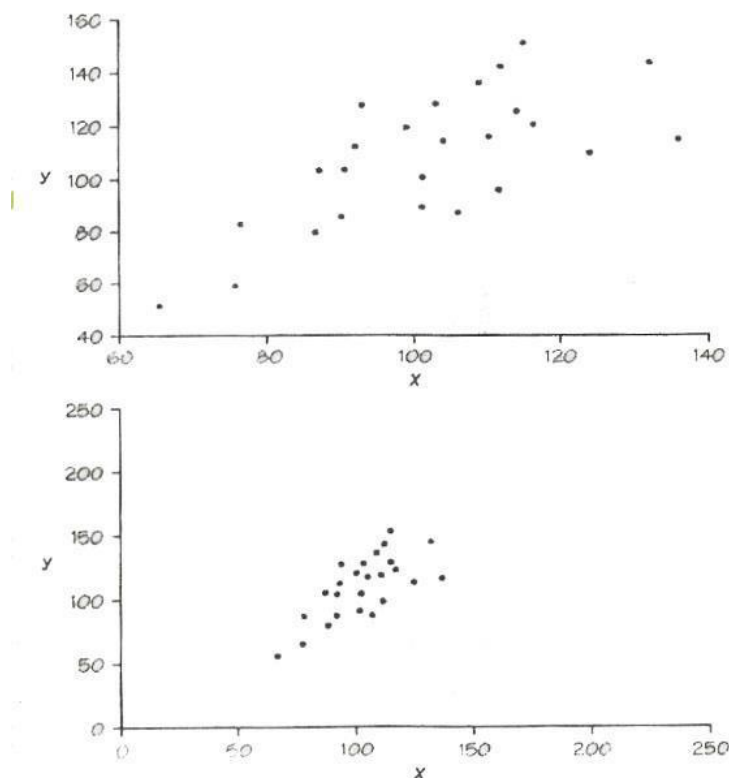
Figura 4: Peso x Altura



Fonte: PORTNOI, 2006.

A confiabilidade quanto à expressão do gráfico requer cuidados redobrados, pois a escala utilizada pode ser diferente. Nesse caso, é preciso muita atenção com as conclusões.

Figura 5: Diferença de Escalas



Fonte: PORTNOI, 2006.

Para que não ocorra “confusões”, será utilizada medida numérica para auxiliar na análise: o Coeficiente de Correlação Linear (r) (PORTNOI, 2006).

O Coeficiente de Correlação Linear, também conhecido como Coeficiente de Correlação de Pearson, mede o grau de relacionamento linear entre os valores de x e y de uma amostra; mede a intensidade e a direção da relação, utilizando a fórmula:

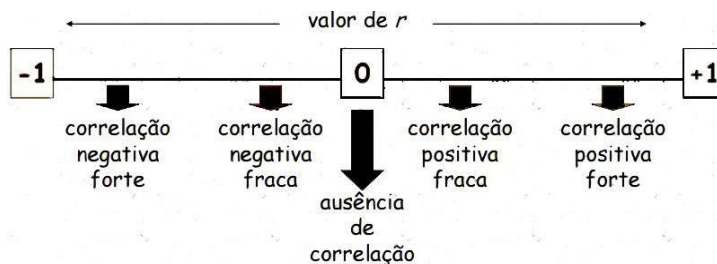
Figura 6: Fórmula do Coeficiente de Correlação de Pearson

$$r = \frac{n \sum(x_i \cdot y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Fonte: PORTNOI, 2006.

O resultado esperado 'r' sempre será um valor entre: $-1 \leq r \leq 1$. Para melhor compreensão desse resultado, segue ilustração:

Figura 7: Classificação do Resultado (r) do Coeficiente de Correlação de Pearson



Fonte: PORTNOI, 2006.

Ainda para a análise do resultado, deve-se levar em consideração algumas propriedades do Coeficiente de Correlação de Pearson:

- O valor de 'r' não varia se todos os valores de qualquer uma das variáveis forem convertidos para escalas diferentes;
- O valor de 'r' não é afetado pela escolha de 'x' ou 'y', permutando 'x' ou 'y', 'r' permanecerá inalterado;
- 'R' só mede a intensidade ou grau de relacionamentos *lineares*; não é, portanto, adequado para medição de intensidade de relacionamentos não-lineares (PORTNOI, 2006).

Para que se tenha uma noção mais exata de onde o valor do coeficiente de correlação se encontra, especificando a Figura 7, será utilizada a tabela a seguir:

Tabela 2: Classificação do Coeficiente de Correlação

COEFICIENTE DE CORRELAÇÃO	CORRELAÇÃO
$r = 1$	Perfeita positiva
$0,8 < r < 1$	Forte positiva
$0,5 \leq r < 0,8$	Moderada positiva
$0,1 \leq r < 0,5$	Fraca positiva
$0 < r < 0,1$	Ínfima positiva
0	Nula
$-0,1 < r < 0$	Ínfima negativa
$-0,5 < r \leq -0,1$	Fraca negativa
$-0,8 < r \leq -0,5$	Moderada negativa
$-1 < r \leq -0,8$	Forte negativa
$r = -1$	Perfeita negativa

Fonte: SOUSA, 2008.

Caso o diagrama sugira uma relação *linear*, seria interessante representar este padrão em uma reta no gráfico.

Para definir como ajustar essa reta dentro desse padrão, será utilizado o método dos mínimos quadrados (PORTNOI, 2006).

Para isso, a seguinte equação de regressão simples descreve a relação entre as duas variáveis da coleção dos dados amostrais, pela fórmula:

Figura 8: Equação de Regressão

$$\hat{y} = \alpha + \beta x$$

Fonte: PORTNOI, 2006.

Em que, β : é o coeficiente angular, calculado por:

Figura 9: Coeficiente

$$\beta = \frac{n(\sum x_1 y_1) - (\sum x_1)(\sum y_1)}{n(\sum x_i^2) - (\sum x_i)^2}$$

Fonte: PORTNOI, 2006.

Angular. α : é o ponto em que a reta intercepta o eixo y, calculado por:

Figura 10: Ponto de Interceptação

$$\alpha = \frac{\sum y_i - \beta \sum x_i}{n}$$

Fonte: PORTNOI, 2006.

3.1.4 Regressão Linear Múltipla

Nesta pesquisa, será utilizada a regressão linear múltipla, que é calculado por:

Figura 11: Regressão Múltipla

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_n X_n$$

Fonte: CONTI, 2012.

A Regressão Linear Múltipla, por sua vez, vem para calcular um dado dependente que se deseja prever (y); porém, este dado depende de N variáveis independentes, conforme a fórmula anterior, sendo representado por X_2, X_3, X_4, X_n .

4 KNOWLEDGE DISCOVERY IN DATABASES (KDD)

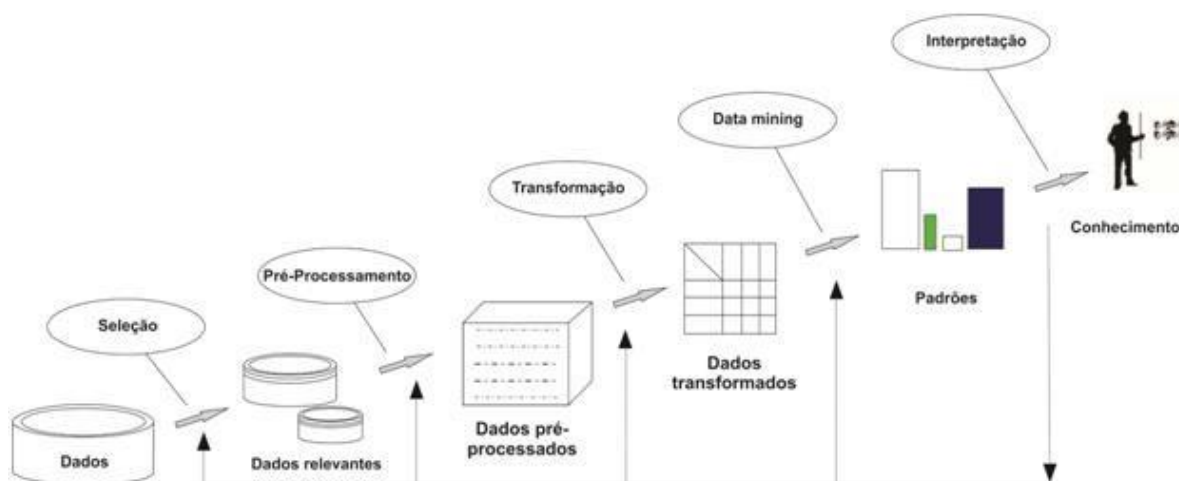
KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. (FAYYAD, 1996)

Sabe-se que entender os processos de sua empresa e o perfil do seu cliente é de grande valia para realizar uma estratégia de negócio, bem como melhorar o serviço e a qualidade do atendimento. Porém, as organizações buscam essas informações de várias formas diferentes, não raro, parecem se “esquecer” de que uma importante base para buscar essas informações está, justamente, nas suas próprias bases de dados (BRITO, 2012).

Contudo, é preciso transformar as informações conhecidas em dados. Logo, visando a este aprimoramento, no final da década de 80, surge um processo chamado *Knowledge Discovery in Databases* (KDD), em português: Descoberta de Conhecimento em Banco de Dados, cujo objetivo é descobrir conhecimento útil, válido, relevante e novo sobre uma determinada base de dados.

Segundo Fayyad (1996), o processo de KDD é constituído por 5 fases como mostra a imagem a seguir:

Figura 12: Etapas do KDD



Fonte: FAYYAD et al, 1996.

A primeira fase é a Seleção de Dados. É nesta fase que serão selecionados os dados apropriados para a análise, de acordo com o critério estabelecido. Exemplo: Pessoas que possuem carro (FORMIGA, 2012).

A segunda fase é o Pré-Processamento, é o ponto em que será efetuada a limpeza dos dados selecionados na fase anterior. Exemplo: Sexo = “F” ou “M” ou Sexo = “M” ou “H” (FORMIGA, 2012).

A terceira fase é a Transformação; nela, será executada a padronização dos dados para suprir possíveis limitações das técnicas de mineração, disponibilizando os dados de maneira usável e navegável. Exemplo: atributos de data e hora (FORMIGA, 2012).

A quarta fase é a Mineração; aqui, aplicam-se técnicas e algoritmos para descoberta de conhecimentos (FORMIGA, 2012).

A quinta fase é a Análise e Assimilação de Resultados; utiliza ferramentas estatísticas e de visualização para melhor entendimento do conhecimento extraído (FORMIGA, 2012).

4.1 Mineração de Dados

Para a realização deste trabalho, foram utilizadas tarefas da quarta fase do KDD, a fase da Mineração de Dados. Segundo PINHEIRO et al. (2015), as ações da Mineração de Dados compreendem: a) Classificação; b) Estimativa; c) Associação; d) Segmentação/Clustering; e) Sumarização. E as técnicas utilizadas são: a) Memory Based Reasoning (MRB); b) Algoritmo Genético; c) Regras de Associação; e d) Árvores de Decisão.

Após o estudo dessas tarefas da Mineração de Dados e dos processos do KDD, foram aplicadas algumas delas para a realização da Mineração de Dados, atendendo aos objetivos desta pesquisa.

Note-se que todos os dados foram descaracterizados para que não ocorra vazamento de informações da empresa que cedeu os dados para a realização deste estudo.

Foi aplicada a terceira fase do KDD, a fase da transformação desses dados para que eles sejam usáveis e navegáveis; isto é: os dados foram recebidos de forma física (impressos em folha A4) e não havia um total das vendas mensais, apenas a quantidade de vendas no decurso do mês, separado por clientes. Para realizar a transformação, foi somado, produto a produto vendido para os clientes dentro daquele mês, chegando a um número total, mensalmente, de cada um dos 20 produtos.

Após realizar a soma e conseguir o número total de vendas dos produtos,

mensalmente, durante o período de 01/01/2012 a 30/11/2014, estes dados foram transportados para uma planilha do Excel, como demonstrado em recorte, sob a forma da figura seguinte:

Figura 13: Parte dos dados das Vendas mensais dos produtos

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	MESES	1 - Produt	2 - Produt	3 - Produt	4 - Produt	5 - Produt	6 - Produt	7 - Produt	8 - Produt	9 - Produt	10 - Produ	11 - Produ	12 - Produ	13 - Produ	14 - Produ	15 - Produ
2	jan/12	12814	24932	14776	45791	51317	17843	79251	71216	383551	21831	40004	119241	20774	181131	28103
3	fev/12	20839	27578	24091	50444	63635	37194	100081	85841	381491	40404	67371	126621	36335	221661	40199
4	mar/12	21349	35147	31135	62749	84436	40508	113771	98331	664808	45717	62877	161021	41638	300601	45739
5	abr/12	18709	29561	15802	51846	48504	30558	102111	91021	419643	34295	46280	131351	38832	228231	26299
6	mai/12	19306	24697	17833	52899	67107	22323	83171	85536	422342	30745	57722	116261	30785	217721	27627
7	jun/12	26902	43967	20995	70705	100904	28664	114851	125211	889147	35641	67299	176741	38561	389141	41808
8	jul/12	27715	41531	21205	64922	52939	29927	116631	115751	615571	41556	77936	186111	37975	340071	45231
9	ago/12	72715	59649	22510	92251	86775	33476	183241	207536	909693	35922	77767	264851	46701	428551	52489
10	set/12	26065	27494	22282	57974	43731	35528	107191	102576	392515	48331	58929	102291	47766	218071	38039
11	out/12	33436	42959	26953	79129	86153	35557	144151	172646	336723	49429	78599	159521	51806	273461	43644
12	nov/12	22849	41304	17842	71399	51678	28052	119271	150611	396356	36055	73706	187641	42272	309791	50653
13	dez/12	18193	33559	11449	71621	38372	18083	108601	154476	257379	17753	62441	165161	23584	323201	37261
14	jan/13	22234	46419	15223	71014	56427	28155	136711	145391	452375	32988	73860	191931	40012	329781	51042
15	fev/13	32353	43673	28663	81202	81809	32396	140331	110926	557939	68286	89216	186081	42156	301091	63440
16	mar/13	26149	38003	16855	44454	52968	39464	51711	19731	463130	43367	59691	156401	47047	217871	29615
17	abr/13	22633	57726	15730	78557	73007	28268	128811	78066	405899	35346	62702	139111	34042	316871	44702
18	mai/13	31501	54668	22459	61816	54133	26166	114591	33801	363572	32196	55023	147871	33070	262331	41635
19	jun/13	50557	35298	30793	34176	14301	52098	59621	54581	455896	60395	63902	145811	53929	245571	31721
20	jul/13	26464	57129	20419	60655	57479	27726	107351	85851	460627	33840	43235	187151	27998	315941	48968
21	ago/13	27688	47629	29356	63857	110980	29487	111541	80136	350833	37181	33513	184771	50445	356771	45955
22	set/13	26206	46949	26956	77253	49907	31345	124651	82156	573039	22126	41077	160081	53352	294411	48990
23	out/13	29143	52745	27640	6388	75883	32247	95991	83871	519261	28813	44923	248641	41174	465661	40199
24	nov/13	22422	44022	26272	15342	37077	31571	80541	64005	330477	31767	33055	176041	37112	267301	46022

Fonte: Autoria própria

Após, foram aplicadas as tarefas de Classificação e Associação.

A tarefa de Classificação foi utilizada para a construção de um modelo dos dados, a fim de que se pudesse classificar estes dados, que foram recebidos de forma física, conforme mencionado. Deu-se a classificação dos produtos que tinham uma grande quantidade de vendas e, também, daqueles que tiveram saída todos os meses durante o período de 01/01/2012 a 30/11/2014.

Quanto à tarefa de Associação, esta foi utilizada para associar itens de consumo da mesma máquina, cujos itens também mantinham correlação uns com os outros. Exemplo: a cada 1.000 impressões realizadas, é comum trocar a peça “x”.

5 RSTUDIO SHINY

SHINY é um pacote open source que fornece um ótimo e poderoso framework que permite criar aplicações web, utilizando a linguagem R.

Shiny foi criado pela equipe do R Studio, que, tempos atrás, visualizou uma grande popularidade da linguagem R e resolveu especializar-se na mesma. O R Studio tem como inspiração a inovação das pessoas nesta ciência e começou a desenvolver ferramentas open source para a comunidade, bem como produtos para profissionais.

O objetivo do R Studio é capacitar os usuários para que sejam produtivos nesta linguagem (SHINY, 2014).

O framework Shiny oferece, para os usuários que têm interesse em começar a trabalhar com o mesmo, um tutorial (em sua página, com 7 (sete) lições).

Cada uma dessas lições explica o funcionamento deste framework, que é estruturado por dois componentes que, por sua vez, contêm o código fonte; são eles: ui.R e o server.R.

O primeiro (ui.R) é o arquivo que abrange o código fonte, de todo o user-interface (ui); ou seja, este controla os layouts e a aparência da sua aplicação. Ressalte-se que, para poder utilizar o Shiny, dentro do R Studio, é preciso instalar o pacote Shiny, por meio do comando: `install.packages("shiny")`.

A imagem a seguir representa um caso do código fonte do ui:

Figura 14: Código user-interface

```
library(shiny)

# Define UI for application that draws a histogram
shinyUI(fluidPage(

  # Application title
  titlePanel("Hello Shiny!"),

  # Sidebar with a slider input for the number of bins
  sidebarLayout(
    sidebarPanel(
      sliderInput("bins",
                  "Number of bins:",
                  min = 1,
                  max = 50,
                  value = 30)
    ),

    # Show a plot of the generated distribution
    mainPanel(
      plotOutput("distPlot")
    )
  )
))
```

Fonte: SHINY, 2015.

Já o segundo server R é onde se encontram todas as instruções que o computador precisa para criar uma aplicação. A próxima ilustração demonstra um caso do código fonte do server R.

Figura 15: Código server R

```
library(shiny)

# Define server logic required to draw a histogram
shinyServer(function(input, output) {

  # Expression that generates a histogram. The expression is
  # wrapped in a call to renderPlot to indicate that:
  #
  # 1) It is "reactive" and therefore should re-execute automatically
  #    when inputs change
  # 2) Its output type is a plot

  output$distPlot <- renderPlot({
    x <- faithful[, 2] # Old Faithful Geyser data
    bins <- seq(min(x), max(x), length.out = input$bins + 1)

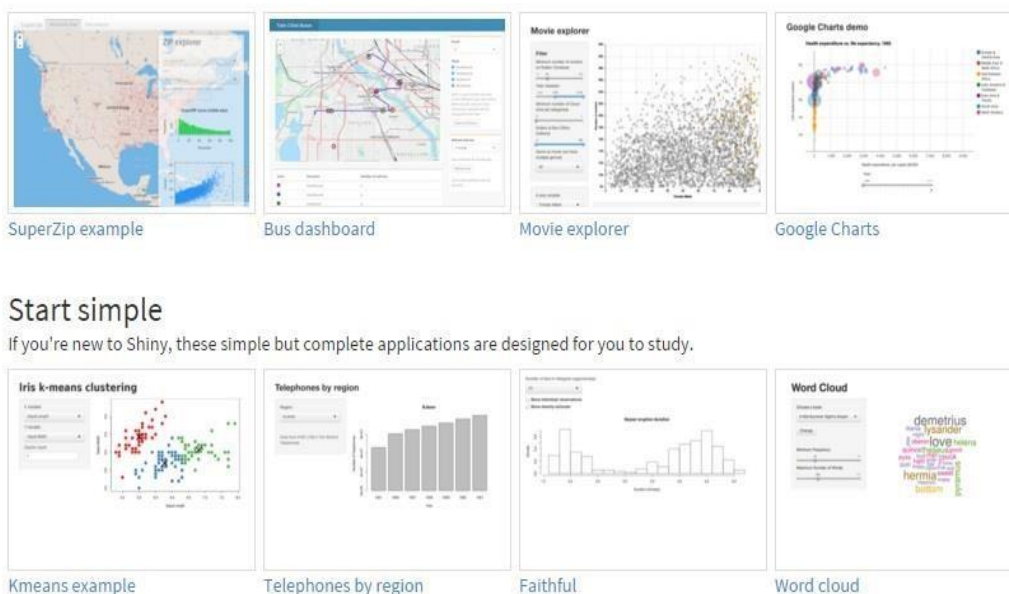
    # draw the histogram with the specified number of bins
    hist(x, breaks = bins, col = 'darkgray', border = 'white')
  })
})
```

Fonte: SHINY, 2015.

As duas imagens anteriores (Figura 14 e Figura 15) exemplificam como é definida a estrutura do Shiny.

O Shiny também conta com uma galeria, com vários tipos de implementações, como se pode observar na próxima imagem:

Figura 16: Galeria de Aplicações do Shiny



Fonte: SHINY, 2015.

Outro diferencial do Shiny está na possibilidade de dispor o aplicativo ou os documentos interativos on-line; para isso, deve-se utilizar o servidor Shiny, que oferece dois tipos de servidores: a) Open Source Edition; e b) Professional Edition.

Tabela 3: Servidores Shiny

SERVIDORES SHINY	OPEN SOURCE EDITION	PROFESSIONAL EDITION
Visão Geral	<ul style="list-style-type: none"> • Ótimo para hospedagem de aplicações públicas leves; • Não suporta autenticação SSL; • Único processo R por aplicação 	<ul style="list-style-type: none"> • Todos recursos de fonte aberta; • Perfeito para grupos de trabalhos e empresas; • Suporta autenticação SSL; • Possui painel administrativo com tempo real e dados históricos de desempenho; • Pode usar vários processos R
Documentação	https://github.com/rstudio/shiny-server	http://rstudio.github.io/shiny-server/latest/
Apoio	<ul style="list-style-type: none"> • Apenas fóruns da comunidade 	<ul style="list-style-type: none"> • E-mail de Suporte • Resposta até 8 horas
Licença	<ul style="list-style-type: none"> • Open Source AGPL v3 	<ul style="list-style-type: none"> • Contrato de Licença RStudio
Preço	Grátis	<ul style="list-style-type: none"> • 20 usuários simultâneos \$9.995 /servidor /ano • 20 usuários adicionais simultâneos \$4.995 /servidor /ano • 150 usuários adicionais simultâneos \$14.995/servidor /ano

Fonte: Shiny, 2015.

6 DESENVOLVIMENT DA APLICAÇÃO

O desenvolvimento da aplicação foi realizado por meio do software R Studio, para os cálculos do Coeficiente de Correlação de Pearson, da Regressão Múltipla, bem como da predição para os próximos meses.

Na interface, foi utilizado o framework, para aplicações web, criado pelo R Studio, chamado Shiny.

Iniciando os trabalhos no software R Studio, foi selecionado o diretório em que está a base de dados pelo comando:

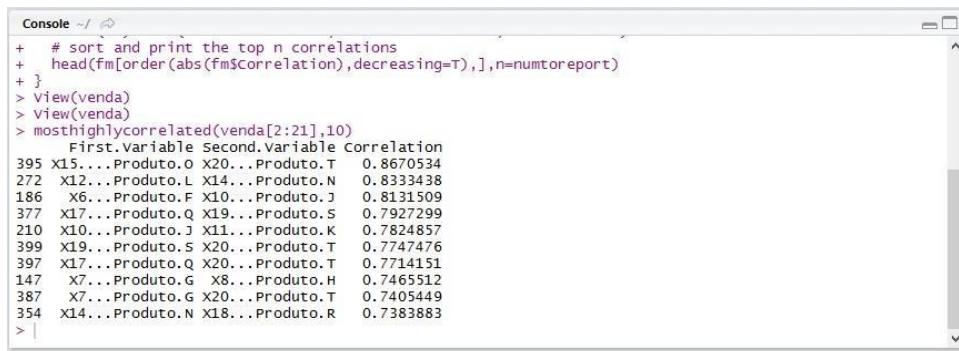
```
setwd ("C:/Users/Felipe Marques/Desktop/APP - CORRELAÇÃO")
```

Após a seleção do diretório, foi colocado os dados na variável VENDAS, pelo comando:

```
vendas <- read.table ("C:\\Users\\Felipe Marques\\Desktop\\APP - CORRELAÇÃO\\vendas tcc.csv" sep = ";")
```

A imagem seguinte indica as 10 (dez) maiores correlações entre os 20 produtos:

Figura 17: Dez maiores correlações entre os produtos

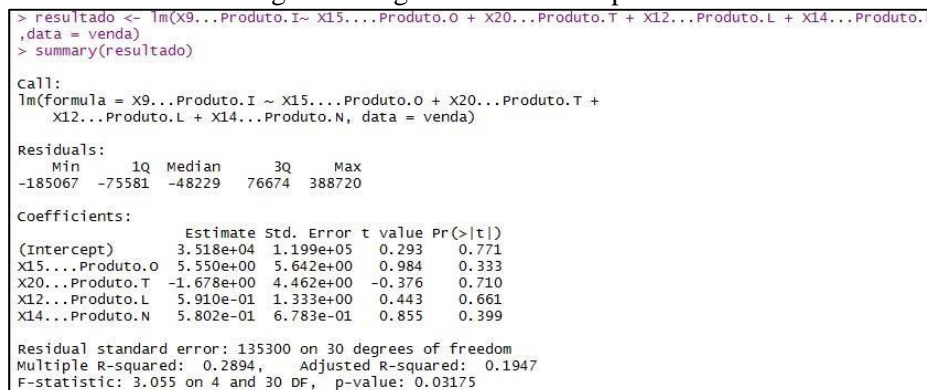


```
Console ~/
+ # sort and print the top n correlations
+ head(fm$order(abs(fm$correlation),decreasing=T),n=numtoreport)
+ }
> View(venda)
> View(venda)
> mosthighlycorrelated(venda[2:21],10)
First.variable Second.variable Correlation
395 x15...Produto.O x20...Produto.T 0.8670534
272 x12...Produto.L x14...Produto.N 0.8333438
186 X6...Produto.F x10...Produto.J 0.8131509
377 x17...Produto.Q x19...Produto.S 0.7927299
210 x10...Produto.J x11...Produto.K 0.7824857
399 x19...Produto.S x20...Produto.T 0.7747476
397 x17...Produto.Q x20...Produto.T 0.7714151
147 X7...Produto.G X8...Produto.H 0.7465512
387 X7...Produto.G x20...Produto.T 0.7405449
354 x14...Produto.N x18...Produto.R 0.7383883
> |
```

Fonte: Autoria própria

Após o cálculo das maiores correlações entre os produtos, será utilizada a Regressão Linear Múltipla, a fim de que seja reconhecida a relação entre as variáveis.

Figura 18: Regressão Linear Múltipla no R



```
> resultado <- lm(X9...Produto.I ~ X15...Produto.O + X20...Produto.T + X12...Produto.L + X14...Produto.N
,data = venda)
> summary(resultado)

Call:
lm(formula = X9...Produto.I ~ X15...Produto.O + X20...Produto.T +
  X12...Produto.L + X14...Produto.N, data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
-185067  -75581  -48229   76674  388720

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.518e+04  1.199e+05   0.293   0.771
x15...Produto.O  5.550e+00  5.642e+00   0.984   0.333
x20...Produto.T -1.678e+00  4.462e+00  -0.376   0.710
x12...Produto.L  5.910e-01  1.333e+00   0.443   0.661
x14...Produto.N  5.802e-01  6.783e-01   0.855   0.399

Residual standard error: 135300 on 30 degrees of freedom
Multiple R-squared:  0.2894,    Adjusted R-squared:  0.1947
F-statistic: 3.055 on 4 and 30 DF,  p-value: 0.03175
```

Fonte: Autoria própria

A figura anterior traduz, no software R, a Equação da Regressão Linear Múltipla, $\gamma = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_n X_n$. Neste caso, as informações faltantes são as de X_2 até X_5 , que foram indicadas pelos vendedores da empresa (apenas se recorde de que os dados foram descaracterizados, devido ao sigilo da informação da empresa).

Foram desenvolvidos 5 (cinco) cálculos para as previsões dos produtos X6, X8, X10, X14 e X18. A seguir, constam os cálculos no R e nas fórmulas:

$$X6 = 14.490 + 28103x(0,8001) + 32509x(-0,3938) = 24.173,16$$

Figura 19: Cálculo Regressão Linear produto X6

```
> resultado <- lm(X6...Produto.F ~ X15...Produto.O + X20...Produto.T, data = venda)
> summary(resultado)

Call:
lm(formula = X6...Produto.F ~ X15...Produto.O + X20...Produto.T,
    data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
-17210.2  -4246.6   -826.6   3586.0  20420.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14490.9358  5207.6943   2.783  0.00897 **
X15...Produto.O    0.8001    0.2612   3.063  0.00441 **
X20...Produto.T   -0.3938    0.1965  -2.004  0.05355 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7299 on 32 degrees of freedom
Multiple R-squared:  0.2575,    Adjusted R-squared:  0.2111
F-statistic: 5.548 on 2 and 32 DF,  p-value: 0.008539
```

Fonte: Autoria própria

Para cada produto X6 vendido, será vendido 0,80 produtos X15 a mais e recuará -0,39 do produto X20.

$$X8 = 20.547 + 28103x(-0,99) + 32509x(2,431) = 71.754,40$$

Figura 20: Cálculo Regressão Linear produto X8

```
> resultado <- lm(X8...Produto.H ~ X15...Produto.O + X20...Produto.T, data = venda)
> summary(resultado)

Call:
lm(formula = X8...Produto.H ~ X15...Produto.O + X20...Produto.T,
    data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
 -71271  -27184    815   17112  74610

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20547.7560  22978.4818   0.894  0.37788
X15...Produto.O   -0.9976    1.1524  -0.866  0.39312
X20...Produto.T    2.4316    0.8670   2.805  0.00849 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32200 on 32 degrees of freedom
Multiple R-squared:  0.3568,    Adjusted R-squared:  0.3166
F-statistic: 8.875 on 2 and 32 DF,  p-value: 0.0008586
```

Fonte: Autoria própria

$$X10 = 6.813 + 28103x(1,3523) + 32509x(-0,616) = 24.791,14$$

Figura 21: Cálculo Regressão Linear produto X10

```
> resultado <- lm(X10...Produto.J~ X15...Produto.O + X20...Produto.T,data = venda)
> summary(resultado)

Call:
lm(formula = X10...Produto.J ~ X15...Produto.O + X20...Produto.T,
    data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
-21818.9 -6909.1  -874.3   7513.6 23514.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6813.9633   8144.7316    0.837  0.40902
X15...Produto.O    1.3523     0.4085    3.311  0.00231 **
X20...Produto.T   -0.6167     0.3073   -2.007  0.05329 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11410 on 32 degrees of freedom
Multiple R-squared:  0.3038,    Adjusted R-squared:  0.2603
F-statistic: 6.982 on 2 and 32 DF,  p-value: 0.003045
```

Fonte: Autoria própria

$$X14 = 172550 + 28103x(-3,478) + 32509x(5,642) = 258.223,5$$

Figura 22: Cálculo Regressão Linear produto X14

```
> resultado <- lm(X14...Produto.N~ X15...Produto.O + X20...Produto.T,data = venda)
> summary(resultado)

Call:
lm(formula = X14...Produto.N ~ X15...Produto.O + X20...Produto.T,
    data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
-89841 -33441  -9924   32545 165261

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 172550.908  41687.459    4.139 0.000237 ***
X15...Produto.O   -3.478     2.091   -1.663 0.105988
X20...Produto.T    5.642     1.573    3.587 0.001098 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 58420 on 32 degrees of freedom
Multiple R-squared:  0.3997,    Adjusted R-squared:  0.3622
F-statistic: 10.65 on 2 and 32 DF,  p-value: 0.0002845
```

Fonte: Autoria própria

$$X18 = 48.634 + 28103x(-0,85) + 32509x(1,881) = 85.895,87$$

Figura 23: Cálculo Regressão Linear produto X18

```
> resultado <- lm(X18...Produto.R~ X15...Produto.O + X20...Produto.T,data = venda)
> summary(resultado)

Call:
lm(formula = X18...Produto.R ~ X15...Produto.O + X20...Produto.T,
    data = venda)

Residuals:
    Min       1Q   Median       3Q      Max
-48455 -22323  -4980   12493   66982

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48634.8637  21645.8631    2.247  0.0317 *
X15...Produto.O  -0.8521     1.0856   -0.785  0.4383
X20...Produto.T    1.8815     0.8167    2.304  0.0279 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30340 on 32 degrees of freedom
Multiple R-squared:  0.2598,    Adjusted R-squared:  0.2135
F-statistic: 5.616 on 2 and 32 DF,  p-value: 0.008121
```

Fonte: Autoria própria

Após a realização dos cálculos anteriores, as informações foram transpostas para um gráfico, a fim de facilitar o entendimento e tornar os dados mais legíveis. O gráfico contém todo o histórico das vendas dos produtos, em uma única cor, preto e branco, chegando até a linha vermelha, que, por sua vez, significa o fim do histórico (30/11/2014). As previsões feitas com os cálculos ora apontados estão após a linha vermelha e figuram em cores diferentes para que, assim, se possa definir qual previsão é de qual produto.

7 RESULTADOS OBTIDOS

Após a análise das técnicas de regressão múltipla e correlação desenvolvidas, as mesmas foram utilizadas nos 20 produtos dos dados ora examinados. Atendendo ao intuito desta pesquisa cujo interesse voltava-se a diminuir o tempo gasto para fazer a previsão confiável sobre as vendas de uma organização, os resultados atingidos foram:

Tabela 4: Resultados Obtidos

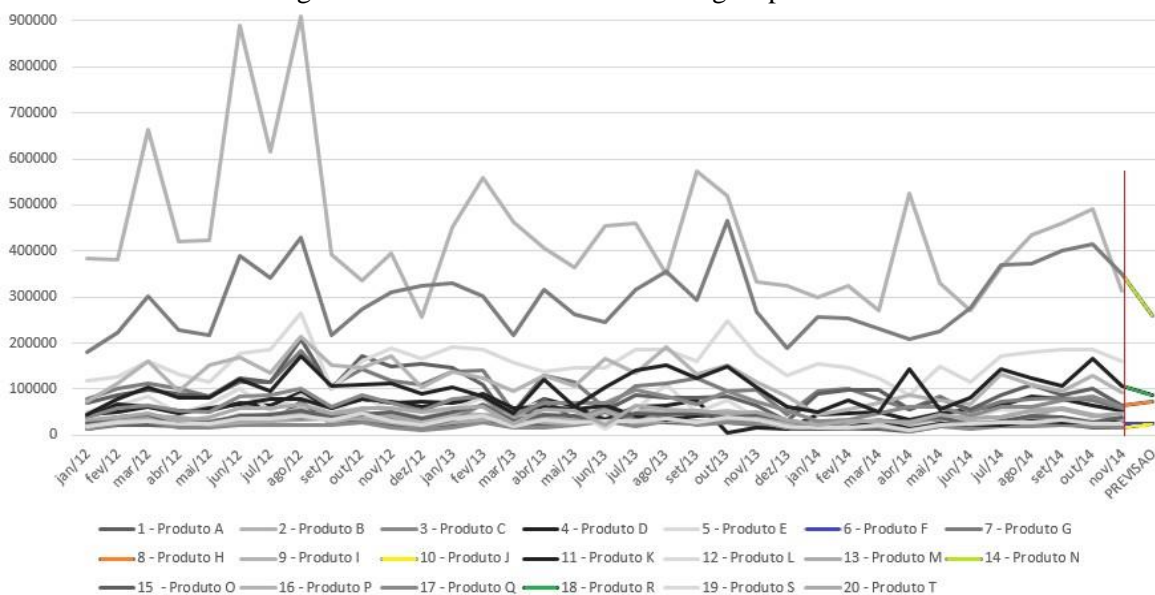
RESULTADOS	
Prever vendas: média de 100 vendedores (Antes)	8 horas
Prever vendas: média de 100 vendedores (Depois)	2 horas
Ganho	75%
Redução de Produtos em Estoque	16%

Fonte: Autoria própria

Para os cálculos efetuados no R, os resultados das previsões foram bastante satisfatórios, conforme exposto pelo histórico das vendas.

Da mesma forma, ampliando a visualização dos resultados, segue gráfico, com previsões apenas parciais, dada a confidencialidade das informações da empresa.

Figura 24: Gráfico com as Previsões de alguns produtos



Fonte: Autoria própria

CONCLUSÃO

Após revisão bibliográfica dos temas envolvidos, verificou-se a necessidade de uma ferramenta que possa auxiliar na eficácia de se agrupar dados históricos de vários anos, bem como emprega-los na realização de cálculos, em um curto espaço de tempo, a fim de apresentar tal dado lapidado e de uma forma de fácil entendimento.

Para atingir esse objetivo final, foi utilizado o processo de mineração de dados (KDD), pois os dados recebidos estavam impressos em papéis, desorganizados, constituindo dados inúteis para a análise desta pesquisa. Com a aplicação da mineração dos dados e a transformação desses em informações, foi aplicado o processo de planejamento de vendas e operações, buscando entender como as organizações podem fazer para alinhar e balancear sua demanda, preparando seu planejamento estratégico.

Compreendido o processo do S&OP, começaram os estudos em torno da Estatística e, igualmente, sobre o método estatístico Correlação de Pearson, para aplicar o aprendizado em forma de código, no software de linguagem estatística R Language.

Depois de realizar a verificação e o funcionamento do código, iniciou-se a implementação de uma interface com uma usabilidade mais facilitada, de tal modo que o responsável por esta área de estratégia da empresa não necessite de um entendimento de linguagem de programação e lógica excessivamente apurado, conseguindo utilizar a ferramenta em forma gráfica.

Enfim, conclui-se que o processo do planejamento de vendas e operações, junto com o método estatístico Correlação de Pearson e a regressão múltipla auxilia bastante a organização para que tenha uma previsão mais correta possível, balanceando, assim, sua oferta e demanda, de forma que a organização não fique com dinheiro parado em seu estoque, sendo que este poderia ser aplicado para várias outras opções dentro da organização.

REFERÊNCIAS

BRITO, Marcelo. **Aspectos teóricos da mineração de dados e aplicação das regras de classificação para apoiar o comércio.** Disponível em: <<http://www.devmedia.com.br/aspectos-teoricos-da-mineracao-de-dados-e-aplicacao-as-regras-de-classificacao-para-apoiar-o-comercio/25429#ixzz3pxx9or48>>. Acesso: 29 out. 2015.

CONTI, Fátima. **Regressão e Correlação.** Disponível em: <<http://www.cultura.ufpa.br/dicas/biome/biopdf/bioreg.pdf>>. Acesso: 11 nov. 2015.

CORRÊA, Henrique L.; GIANESI, Irineu, G. N.; CAON, Mauro. **Planejamento, Programação e Controle da Produção. MRP II / ERP.** Conceitos, Uso e Implantação. 4. Ed. São Paulo: Atlas, 2001.

FAYYAD, U. M.; Piatetsky-Shapiro, G.; Smyth, P. **From Data Mining to Knowledge Discovery in Databases.** Artificial Intelligence Magazine, v. 17, n. 3, p. 37-54, 1996.

FORMIGA, Andrei. **Knowledge Discovery Database (KDD) e Data Mining (DM).** Disponível em: <<http://pt.slideshare.net/ThiagoCorey/kdd-e-data-mining>>. Acesso: 29 out. 2015.

MCCLAVE J. T.; BENSON P. G.; SINCICH T. **Estatística para administração economia.** 10. ed. São Paulo: Pearson Prentice Hall, 2009.

PINHEIRO, Larissa; OLIVEIRA, Viliane; NEVONY, Wagner; SPINOLA, Rodrigo. **Processo de descoberta de conhecimento em bases de dados.** Revista SQL Magazine 108. Disponível em: <<http://www.devmedia.com.br/processo-de-descoberta-de-conhecimento-em-bases-de-dados-revista-sql-magazine-108/26998#ixzz3rBeZL8sq>> Acesso: 11 nov. 2015.

PORTNOI, Marcos. **Probabilidade e Estatística.** Disponível em: <http://www.eecis.udel.edu/~portnoi/classroom/prob_estatistica/2006_2/prob-estatistica2006.2.html>. Acesso: 26 mai. 2015.

R LANGUAGE. Disponível em: <<http://www.r-project.org/>>. Acesso: 02 jun. 2015.

ROCHA, J.; BAYER, A.; BITTENCOURT, H. R.; ECHEVESTE, S. **A Estatística e sua história.** Canoas. XII Simpósio Sul brasileiro de Ensino de Ciências, 2004.

R STUDIO. Disponível em: <<https://www.rstudio.com/>>. Acesso: 01 nov. 2015.

SHINY. Disponível em: <<http://shiny.rstudio.com/>>. Acesso: 01 nov. 2015.

SOUSA, Áurea. **Coefficiente de Correlação Linear de Pearson**, 2008. Disponível em: <http://www.aurea.uac.pt/pdf_MBA/coef_correl_Pearson.pdf>. Acesso: 16 nov. 2015.

TAVARES, Ricardo. **Probabilidade e avaliação de testes diagnósticos**. Disponível em: <<http://www.cead.ufop.br/jornal/index.php/estteste/article/download/36/>>. Acesso: 02 jun. 2015.

WALLACE, T. F. **Sales and operations planning: the how to handbook**. 2 ed. Ohio: T. F. Wallace & Company, 2004.