

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

RAFAEL HENRIQUE BARROS DA SILVA

ANÁLISE DE REGRESSÃO E SIMULAÇÃO NO GERENCIAMENTO DE INDICADORES
DE DESEMPENHO INDUSTRIAL

DOURADOS – MS

2016

RAFAEL HENRIQUE BARROS DA SILVA

ANÁLISE DE REGRESSÃO E SIMULAÇÃO NO GERENCIAMENTO DE INDICADORES
DE DESEMPENHO INDUSTRIAL

Trabalho de Conclusão de Curso de graduação
apresentado para obtenção do título de Bacharel em
Engenharia de Produção, pela Faculdade de
Engenharia da Universidade Federal da Grande
Dourados. Orientador: Prof. Dr. Walter Roberto
Hernandez Vergara.

DOURADOS – MS

2016

RAFAEL HENRIQUE BARROS DA SILVA

ANÁLISE DE REGRESSÃO E SIMULAÇÃO NO GERENCIAMENTO DE INDICADORES
DE DESEMPENHO INDUSTRIAL

Trabalho de Conclusão de Curso de graduação aprovado como requisito parcial para obtenção do título de Bacharel em Engenharia de Produção, pela Faculdade de Engenharia da Universidade Federal da Grande Dourados pela comissão formada por:

Orientador Prof. Dr. Walter Roberto Hernandez Vergara
FAEN - UFGD

Prof. Dr. Fábio Alves Barbosa
FAEN – UFGD

Prof. Dr. Fabiana Raupp
FAEN - UFGD

Dourados, 28 de abril de 2016.

RESUMO

Os métodos estocásticos podem ser utilizados na solução de problemas e explicação de fenômenos naturais através de procedimentos estatísticos. Este trabalho busca associar dois dos métodos estocásticos mais influentes, análise de regressão e simulação de sistemas, com o objetivo de facilitar o entendimento prático da análise de dados. Os algoritmos são desenvolvidos em Excel, utilizando várias técnicas estatísticas como a teoria da regressão, ANOVA, método de Monte Carlo e fatoração de Cholesky para criar modelos de análises de sistemas simples e múltiplos com até cinco variáveis independentes. Neste estudo, os modelos são testados através da execução de análises de indicadores de desempenho industrial. Entre os resultados das análises, estão índices numéricos criados com o objetivo de melhorar o gerenciamento das metas para indicadores de conformidade, através da identificação de instabilidade, correlação e anomalias do sistema. Apesar de simples e acessíveis, os modelos de análise apresentados nesta pesquisa entregaram resultados satisfatórios com inúmeras possibilidades de aplicações empresariais e acadêmicas, além do potencial para desdobramento em novas técnicas de análise.

Palavras-chave: Análise de Regressão; Simulação de Sistemas; Indicadores de Desempenho.

ABSTRACT

Stochastic methods can be used in problem solving and explanation of natural phenomena through the application of statistical procedures. This research aims to associate two of the most influential stochastic models, regression analysis and systems simulation, in order to facilitate the practical understanding of data analysis. The algorithms are designed on Excel, by using several statistical techniques such as regression theory, ANOVA, Monte Carlo method, and Cholesky Factorization to create models to analyze either simple or multiple systems with up to five independent variables. In this study, the models are tested through analysis performed on industrial performance measures. Among the results, indexes were created to improve the measurement system management, by identifying instability, correlation, and anomalies. Even though the models are simple and accessible, they provided very satisfactory results with numerous possibilities of industrial and academic applications, besides the potential for new analysis techniques.

Keywords: Regression Analysis; Systems Simulation; Performance Measures.

Dados Internacionais de Catalogação na Publicação (CIP).

S586a ,

Análise de regressão e simulação no gerenciamento de indicadores de desempenho industrial / -- Dourados: UFGD, 2016.

48f. : il. ; 30 cm.

Orientador: Walter Roberto Hernandez Vergara

TCC (graduação em Engenharia de Produção) - Faculdade de Engenharia, Universidade Federal da Grande Dourados.

Inclui bibliografia

1. Análise de regressão. 2. Simulação. 3. Indicadores de desempenho. I. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

©Direitos reservados. Permitido a reprodução parcial desde que citada a fonte.

Sumário

1	INTRODUÇÃO	6
1.1	CARACTERIZAÇÃO DO TEMA	6
1.2	PROBLEMA	8
1.3	JUSTIFICATIVA	9
1.4	OBJETIVOS	10
1.4.1	OBJETIVO GERAL	10
1.4.2	OBJETIVOS ESPECÍFICOS	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	INDICADORES DE DESEMPENHO	12
2.2	ANÁLISE DE REGRESSÃO	14
2.2.1	REGRESSÃO LINEAR SIMPLES	16
2.2.2	REGRESSÃO LINEAR MÚLTIPLA	17
2.3	SIMULAÇÃO DE SISTEMAS	19
2.3.1	MÉTODO DE MONTE CARLO	22
3	METODOLOGIA	24
4	SELEÇÃO DE DADOS	26
4.1	INDICADOR DE RENDIMENTO	27
4.2	INDICADOR DE CONFORMIDADE	28
5	DESENVOLVIMENTO DOS MODELOS DE ANÁLISE ... Erro! Indicador não definido.	
5.1	MODELO DE REGRESSÃO	29
5.2	MODELO DE SIMULAÇÃO	29
6	RESULTADOS E DISCUSSÕES	32
6.1	ANÁLISE SIMPLES DE INDICADORES	32
6.2	ANÁLISE MÚLTIPLA DE RENDIMENTO	38
6.3	ANÁLISE MÚLTIPLA DE CONFORMIDADE	38
7	CONSIDERAÇÕES FINAIS	44
8	REFERÊNCIAS BIBLIOGRÁFICAS	46

1 INTRODUÇÃO

1.1 CARACTERIZAÇÃO DO TEMA

Modelos matemáticos são usados para solucionar problemas na ciência desde a primeira definição do princípio científico da descoberta. O sucesso na utilização desses modelos requer entendimento das propriedades implícitas do fenômeno, as características matemáticas e estatísticas do modelo e os problemas práticos que podem ser encontrados ao utilizar os modelos em situações reais. Existem basicamente duas abordagens ao utilizar modelos matemáticos para explicar fenômenos naturais. A primeira busca usar modelos complexos para explicar um fenômeno em sua totalidade, mas mesmo casos mais simples podem apresentar problemas que são resolvidos apenas por ferramentas matemáticas mais sofisticadas. Um modelo que explica completamente, quando possível, a ação de uma resposta a um problema é chamado determinístico. A segunda abordagem na análise de problemas naturais utiliza modelos mais simples para obter uma solução que se aproxima da solução exata do problema. Esse método é chamado estocástico, que geralmente possui soluções que podem ser avaliadas por distribuição de probabilidade. Isto é, soluções de modelos estocásticos são apresentadas em um intervalo de confiança e apoiadas pelo resultado de um teste de hipóteses (FREUND; WILSON, 2006).

Modelos estocásticos constam da coleta, síntese e análise de dados para estabelecer afirmações sobre o mundo real. Um dos modelos estatísticos mais utilizados para analisar dados é análise de regressão, cujo diferencial é a habilidade de inferir sobre as características de variáveis depois de ter valores controlados de variáveis explicativas. Vários métodos estocásticos são importantes, mas a análise de regressão tem sido a ferramenta mais influente, tão difundida nos negócios modernos que muitas vezes se subestima o fato de ter mais de cem anos (FREES, 2010).

Análise de regressão consiste de técnicas estatísticas para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. Na regressão, a variável dependente é projetada como uma função das variáveis independentes e seus correspondentes coeficientes. Se um modelo de regressão representa adequadamente a relação entre as variáveis, ele pode ser usado para prever a variável de resposta, identificar variáveis independentes importantes e estabelecer a relação causal desejada (YAN, 2009).

Várias ferramentas são utilizadas para testar a eficiência de um modelo de regressão, uma delas é a análise de variâncias. A análise de variâncias é um procedimento de cálculos que fornece

informações sobre os níveis de variação dentro de um modelo de regressão e estabelece uma base para testar sua significância. Quando a análise de regressão não satisfaz completamente os objetivos da pesquisa, outras ferramentas se fazem necessárias. Outro método estocástico amplamente utilizado na análise de dados é a simulação de sistemas (FREES, 2010).

A análise e modelagem de simulação é o processo de criação e experimentação de modelos matemáticos computadorizados de um sistema físico, buscando imitar suas operações ao longo do tempo. Nesse sentido, a simulação se assemelha à análise de regressão porque ambos fornecem um conjunto de informações ao invés de um resultado específico, permitindo ao analista interpretar a solução do problema. Em um cenário de predominância de linguagens de programação, rápido crescimento do poder de computação, e ampla disponibilidade de novas tecnologias, aumenta exponencialmente a necessidade de soluções computacionais mais complexas e avançadas como o método de Monte Carlo (BANKS et al, 2004).

Simulação de Monte Carlo é o método computacional iterativo usado para examinar e investigar o comportamento de sistemas físicos e matemáticos utilizando técnicas estocásticas. Trata-se de um método amplamente utilizado e uma ferramenta estatística bem sucedida na pesquisa de uma variedade de problemas, áreas e casos nos quais é impossível ou inviável calcular resultados exatos usando algoritmos determinísticos (MARK; MORDECHAI, 2011).

O método de Monte Carlos tem provado ser uma técnica estatística de amostragem computacional muito útil para alcançar soluções numéricas aproximadas para problemas quantitativos de sistemas complexos e parâmetros incertos. Quando se trata de metodologia de solução de problemas, a abordagem de Monte Carlo é conhecida por possuir a maior acurácia. Contudo, essas técnicas eram tradicionalmente consideradas muito dispendiosas por causa da quantidade necessária de tempo de operações computacionais. Atualmente, como resultado da constante evolução da tecnologia de computação e crescente disponibilidade de recursos, essas operações computacionais podem ser significativamente aceleradas (MARK; MORDECHAI, 2011).

O uso de modelos de regressão e simulação para explicar fenômenos naturais tem aplicações em todas as áreas de pesquisa, incluindo engenharia, negócios, economia, ciências sociais, saúde e ciências biológicas, entre outros. Especialmente nas engenharias, os métodos se desenvolveram em procedimentos mais complexos e úteis na análise de situações reais. As aplicações nas áreas de engenharia de produção e gestão empresarial são ilimitadas, como pesquisa

operacional, planejamento de manutenção, análise de riscos, estudo de métodos e processos, entre outros. Para realização dessas análises matemáticas, as operações e resultados dos processos são convertidos em parâmetros numéricos, conhecidos como indicadores de desempenho (FREUND; WILSON, 2006).

A medição de desempenho é um processo que consiste nas seguintes atividades: coleta de dados, análise e relato de informações referentes ao desempenho de um sistema que permite estabelecer se os processos estão operando como deveriam. Não é possível haver melhorias em um processo se não se sabe onde ele está falhando, os indicadores de desempenho fornecem esse tipo de *feedback*. Os indicadores de desempenho podem ser relacionados aos resultados financeiros da organização ou focar nas variáveis determinantes do processo, como flexibilidade, qualidade, eficiência e produtividade. A experiência na utilização das medidas de desempenho para avaliação e tomada de decisões mostrou que os indicadores internos, diretamente relacionados às operações da organização são ainda mais determinantes na melhoria do resultado final.

1.2 PROBLEMA

Em uma empresa com um sistema de medição de desempenho, os indicadores de desempenho operacional são referências para o atingimento de metas da produção. Gestores executam suas ações com o objetivo de melhorar os indicadores atuantes sobre suas respectivas áreas. Contudo, indicadores podem ser obscuros ou não refletir corretamente o estado de uma atividade visto que nem sempre existe um claro entendimento sobre sua composição e cálculo. Esse desconhecimento de fatores incertos na estruturação do sistema de medição de desempenho leva ao erro, uma vez que muitas das ações das quais se espera serem eficazes na melhoria dos resultados são baseadas em um fundamento errôneo e não afetam diretamente nos indicadores.

Quando não há uma completa compreensão do sistema de medição, a organização se deixa guiar por resultados que podem não representar o que se acredita. Não se reconhece, por exemplo, uma anomalia na tendência dos resultados, um ponto fora dos limites de controle. Assim, surgiu na empresa estudada uma teoria de que não se podiam alcançar simultaneamente indicadores relacionados à produtividade e qualidade. Sem a realização de testes, não era possível confirmar a veracidade desta hipótese, mas essa situação já tinha impacto direto nas decisões gerenciais.

Para melhorar o gerenciamento de indicadores de desempenho, várias ferramentas estatísticas podem ser aplicadas na análise dos resultados. Então, surge outro problema, as

ferramentas disponíveis geralmente exigem elevado conhecimento técnico para sua aplicação. Softwares estatísticos especializados na análise de dados fornecem resultados e informações satisfatórias, mas na maioria das vezes não se compreende os procedimentos utilizados em seu desenvolvimento. Além disso, nem toda empresa tem acesso a ferramentas estatísticas mais complexas, visto seu elevado custo de investimento. Dessa forma, a análise dos indicadores é baseada nos conhecimentos técnicos e práticos dos analistas e engenheiros.

Engenheiros são confrontados com o tratamento de indicadores de desempenho rotineiramente e podem interpretar as informações fornecidas por um programa computacional, mas nem sempre são hábeis para executar uma análise completa partindo somente dos indicadores do processo. Na formação acadêmica, esses profissionais são fundamentados em teorias e procedimentos estatísticos, mas procedimentos como análises de regressão e simulação ainda são utilizados em situações reais com o risco que apresentam. Apesar de ser uma área científica essencial nas atividades de engenharia, acadêmicos encontram dificuldade em realizar uma análise completa de dados para extrair informações importantes e propor melhorias práticas e funcionais ao processo. Como utilizar procedimentos estatísticos e ferramentas mais simples para avaliar indicadores de desempenho, de modo a melhorar o entendimento teórico e prático de analistas?

1.3 JUSTIFICATIVA

A complexidade dos problemas da ciência e dos negócios intriga os pesquisadores a entender a relação entre as variáveis de um sistema, seja para definir a causa de um problema ou até prever o comportamento futuro de um fenômeno natural. Comumente, são encontradas situações onde se questiona a influência que um fator exerce sobre os outros que compõem um sistema. Entre as várias técnicas disponíveis para responder esse tipo de questionamento, se destacam a teoria da análise de regressão e o uso de sistemas computacionais para simulação com instrumentos e mecanismos cada vez mais precisos. Neste sentido, a simulação de uma análise de regressão pode se tornar uma ferramenta ainda mais completa e útil na resolução de problemas reais.

A simulação é intuitivamente atraente para um cliente porque reproduz o que acontece em um sistema real ou o que se espera de um sistema que está na fase de concepção, permitindo que um fenômeno da natureza seja representado em termos matemáticos e estatísticos, baseado nas mais diversas formas de distribuição de probabilidades e correlação entre as variáveis. Os

resultados apresentam soluções práticas e próximas da realidade para problemas de diferentes níveis de complexidade, correspondendo diretamente aos resultados reais do sistema. Além disso, é possível desenvolver o modelo de um sistema sem hipóteses duvidosas e tendenciosas, tradicionalmente utilizadas em outros modelos matemáticos. Por estas e outras razões, a simulação é frequentemente a melhor opção na resolução de problemas.

O alinhamento entre esses dois procedimentos estocásticos permite abranger tópicos mais avançados no meio acadêmico e industrial que poderiam ser ignorados no ensino tradicional ou limitados pelo tempo. Nesse sentido, a simulação da análise de regressão pode ajudar a compreender as propriedades estatísticas dos estimadores de uma regressão analiticamente derivados de uma distribuição amostral. A pesquisa tem o potencial de desenvolver uma aprendizagem profunda dos conhecimentos estatísticos, sentido oposto da aprendizagem tradicional focada em memorização.

Os modelos de análise de regressão e simulação são desenvolvidos em Microsoft Office Excel, software acessível e popular nos meios acadêmicos e empresariais. A utilização desses modelos oferece uma alternativa aos programas especializados de análise, que requerem custo elevado e complexo conhecimento técnico prévio. Apesar de simples, o Excel disponibiliza recursos estatísticos relevantes que facilitam a compreensão e desenvolvimento da análise de regressão e simulação. Nos modelos, o usuário pode manipular dados da maneira mais conveniente e até mesmo, a estrutura do modelo para obtenção de novos resultados.

Acredita-se que a união das análises de regressão e simulação apresenta uma evolução na solução de problemas, vinculando os melhores atributos das duas metodologias. A aplicação das duas análises aprimora as informações resultantes do modelo e fornece informações diferentes, que permitem ao usuário entender outros fatores que afetam o comportamento das variáveis. Essas vantagens diferenciais fazem dessa pesquisa necessária e útil, tanto no âmbito de pesquisa acadêmica quanto em aplicações empresariais.

1.4 OBJETIVOS

1.4.1 OBJETIVO GERAL

O objetivo geral desta pesquisa é apresentar um a criação e desenvolvimento de um modelo estatístico simples e facilitador de análise de regressão e simulação, aplicando indicadores de desempenho a fim de corroborar a eficiência dos mesmos.

1.4.2 OBJETIVOS ESPECÍFICOS

- Levantamento e pesquisa bibliográfica da análise de regressão simples e múltipla;
- Levantamento e pesquisa bibliográfica do método de simulação de Monte Carlo;
- Desenvolver e apresentar um modelo computacional em Excel composto para análise de regressão linear simples e múltipla;
- Desenvolver e apresentar modelo computacional composto de simulação de variáveis em Excel na análise de regressão linear simples e múltipla;
- Executar os modelos desenvolvidos com aplicação de indicadores de desempenho coletados de uma empresa real;
- Interpretar resultados da análise e propor melhorias ao sistema de gerenciamento dos indicadores de desempenho.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 INDICADORES DE DESEMPENHO

Organizações atingem seus objetivos ao satisfazer os desejos e necessidades de seus *stakeholders* com maior efetividade e eficiência que seus competidores. Ser efetivo é garantir que os resultados esperados sejam apropriadamente alcançados, enquanto eficiência significa o maior aproveitamento dos recursos disponíveis, sejam eles pessoas ou produtos. Essa é uma importante distinção porque não somente identifica duas dimensões fundamentais do desempenho, mas também destaca que existem fatores internos e externos influenciando a atuação da empresa. Assim, medir o desempenho de uma organização pode ser entendido como o processo de quantificação da eficiência e efetividade das ações executadas, as quais determinarão como a empresa está atuando no presente (NEELY et al, 2002).

Um sistema de medição de desempenho gera informações para estimar o nível de atendimento das metas, a fim de prover mais subsídios para as decisões a serem tomadas na organização. Essas informações são resultado da coleta e processamento de dados do processo, organizadas e apresentadas como indicadores de desempenho. Os indicadores são parâmetros qualitativos ou quantitativos sobre um processo examinado, que possibilitam analisar sua evolução e verificar se os objetivos estão sendo cumpridos. A definição desses indicadores deve ser bem aceita e claramente transmitida para gerentes e operadores, com resultados devidamente documentados sem distorções ou manipulações. A informação resultante do processo precisa ser exata, precisa, sensível a significantes mudanças, estável e replicável (HARBOUR, 2009).

Cada indicador representa um alvo estratégico específico, logo devem ser cuidadosamente selecionados, levando em conta as políticas de qualidade; áreas de interesse de estudo; variáveis de desempenho e saídas do processo. Um importante aspecto dos indicadores de desempenho é a rastreabilidade e verificação, para que a interpretação dos resultados seja confiável e direcione a organização à raiz do problema. . Um bom indicador deve cumprir alguns requisitos básicos, como representatividade; simples interpretação; capacidade de indicar tendências temporais; sensibilidade a mudanças dentro e fora da organização; facilidade na coleta e processamento de dados; e facilidade de atualização (FRANCESCHINI et al, 2007).

Uma das principais funções dos indicadores de desempenho é o controle. Processos que não são mensurados não podem ser controlados e, conseqüentemente, entendidos. Entender como um

processo funciona significa conhecer o que acontece no processo, os fatores que afetam seu desempenho, como o sistema se comporta diante de intervenções e os limites da capacidade do mesmo. Quando não se utiliza um sistema de medição de desempenho, há uma grande diferença entre como os gerentes acreditam que o processo funcione e como ele realmente funciona. Empresas que entendem como seus processos funcionam e conhecem sua capacidade fazem planejamentos e projeções mais confiáveis. O *feedback* imediato resultante dos indicadores de desempenho também proporciona melhor controle do sistema, uma vez que desvios na performance são detectados com antecedência, possibilitando uma intervenção para minimizar os danos ou aproveitar melhor uma oportunidade. *Feedback* é essencial para o controle de qualquer sistema, quando gestores não recebem retornos significativos e apropriados, seus processos não reconhecem oportunidades e se tornam muito mais vulneráveis às ameaças. O sistema de medição de desempenho possibilita direcionar os esforços para os maiores problemas e oportunidades a fim de alcançar melhorias tangíveis (KAYDOS, 1998).

Indicadores de desempenho são uma das melhores maneiras de comunicar as estratégias da empresa aos colaboradores. Para isso, é necessário definir o objetivo estratégico e então desdobrá-lo em objetivos menores para os níveis tático e operacional, com seus indicadores correspondentes. Quando a medição de desempenho reflete a estratégia da empresa, se garante que todos estão focados no mesmo resultado (MERCHANT; STEDE, 2007).

De acordo com o momento de observação dos indicadores, os mesmos podem ser classificados em inicial, intermediário ou final. Indicadores iniciais tem o objetivo de melhorar o planejamento e gerenciamento do projeto, envolvendo todos os recursos disponíveis, como instalações, recursos humanos, ativos tecnológicos e financeiros, serviços prestados por fornecedores, etc. indicadores intermediários mensuram a consistência entre resultados do processo e suas especificações, provendo informações sobre o estado do processo. Indicadores finais são geralmente vistos como os mais importantes, pois estimam os resultados do processo, como satisfação do cliente e custos de produção (KAPLAN; NORTON, 1993).

Na literatura, é possível encontrar várias abordagens diferentes para o desenvolvimento de um sistema de medição de desempenho, incluindo os modelos *Dupont*, *Critical Few*, *Malcom Baldrige Quality Award* e *Balanced Scorecard*. O *Balanced Scorecard*, proposto em 1992 por Kaplan e Norton, tem sido a metodologia mais difundida para medição de desempenho. Esse modelo se destaca por forçar a alta gerência a reconhecer que múltiplas atividades devem ser

executadas para que a empresa suceda. A gestão e monitoramento dessas atividades devem ser balanceados, considerando todas as características da organização e não somente o aspecto econômico (KAPLAN; NORTON, 1992).

Em última análise, o sistema de medição é responsável por manter alinhamento e coordenação na organização. Alinhamento se refere à consistência entre os objetivos estratégicos e os indicadores, Coordenação é o grau no qual os indicadores de várias áreas relacionadas são consistentes e sustentados uns pelos outros, a coordenação busca reduzir conflitos potenciais quando há objetivos contrastantes, como indicadores de produtividades conflitantes com indicadores de produtividade. Indicadores devem ser flexíveis para reconhecer e responder a mudanças e necessidade de priorizar novas estratégias.

2.2 ANÁLISE DE REGRESSÃO

Pesquisadores continuamente buscam entender como um sistema pode ser afetado por fatores internos e externos, utilizando métodos a fim de padronizar e controlar as causas e efeitos desses fatores. Frequentemente, ferramentas estatísticas são aplicadas para quantificar e manipular os componentes de um sistema e ainda que inúmeras outras sejam de significativa importância, a análise de regressão tem sido a mais influente e amplamente utilizada em diversos campos de estudo. Partindo de uma coleta de dados, onde propriedades do objeto de estudo serão processadas em escala numérica e transformadas em informações, é possível criar e testar hipóteses sobre o comportamento do mundo real. O princípio da análise de regressão é investigar a relação entre uma ou mais variáveis, com o objetivo de determinar o efeito causal que uma exerce sobre outra. Por exemplo, pode-se estudar o efeito do aumento dos preços sobre a demanda de um produto, ou como os hábitos alimentares influenciam na saúde do ser humano. A metodologia é tão universal no mundo moderno que é frequentemente subestimada e ignora-se o fato de que está sendo desenvolvida por mais de um século (FREES, 2010).

As primeiras evidências de pesquisas sobre regressão são vinculadas a Sir Francis Galton, um primo de Charles Darwin e renomado cientista do século XIX, cuja fascinação por genética levou ao estudo de regressão. As ideias que deram início ao desenvolvimento do estudo surgiram de um problema de hereditariedade, que consistia em entender quão forte as características de uma geração de seres vivos se manifestava na geração seguinte. Galton inicialmente abordou o problema examinando as características da ervilha de cheiro, escolhida pela habilidade da planta de

autofertilização, uma vez que as plantas filhas expressam as variações genéticas da planta mãe sem a influência de uma segunda planta. O cientista desenhou um gráfico bidimensional comparando o tamanho de plantas mães e plantas filhas, representação que foi utilizada para ilustrar os fundamentos básicos do que se tornou conhecido como análise de regressão. A formulação desses estudos em métodos estatísticos e até modelos mais complexos de regressão se deu ao longo dos anos seguintes (DRAPER; SMITH, 2006).

Estatisticamente, a análise de regressão consiste de técnicas de modelagem da relação entre uma variável dependente, também chamada de resposta, e uma ou mais variáveis independentes, também conhecidas como preditoras. No modelo de regressão, as variáveis dependentes são projetadas como uma função de variáveis independentes, seus coeficientes de regressão correspondentes e um parâmetro de erro que representa a variação no comportamento das variáveis que não pode ser explicada pelo modelo estatístico. Se o modelo de regressão refletir adequadamente a verdadeira relação entre a variável resposta e a variáveis preditoras, esse modelo pode ser usado para prever o comportamento de variáveis dependentes, identificar variáveis independentes importantes e estabelecer a relação causal desejada entre as variáveis (YAN, 2009).

A equação de regressão é o resultado final da análise, trata-se de um sumário estatístico que define a dependência matemática entre o conjunto de variáveis. A partir dessa função, é possível avaliar o impacto das variáveis preditoras individualmente, analisar os efeitos de variações nos valores dessas variáveis, e prever os valores de resposta com base em um determinado conjunto de variáveis preditoras. Contudo, para obter a equação de regressão, a metodologia inclui uma série de passos que incluem definição do problema; seleção de variáveis; coleta de dados; definição, desenvolvimento e validação do modelo; e execução do modelo para solucionar o problema proposto (CHATTERJEE, HADI, 2006).

Antes mesmo de formular o problema, é necessário estudar cautelosamente o sistema real para determinar se o mesmo se enquadra no escopo da análise de regressão e pode ser resolvido através dessa metodologia. A definição do problema envolve a determinação das questões a serem respondidas pela análise, trata-se da primeira e mais importante etapa do processo. Um problema mal formulado pode levar a uma seleção de variáveis irrelevantes ou a escolha de um modelo inadequado, sem alcançar o objetivo do estudo. Uma vez definido o problema, o pesquisador segue para a seleção das potenciais variáveis relevantes, aquelas que devem explicar ou prever o comportamento da variável de resposta. O próximo passo é a coleta de dados do objeto de estudo

que serão utilizados na análise. Frequentemente, os dados são coletados em condições não experimentais, onde se tem pouco controle sobre o sistema. Os dados consistem em amostras de n observações de cada uma das variáveis selecionadas na etapa anterior (YAN, 2009).

Depois de definir se o modelo utilizado na regressão é linear ou não linear, simples ou múltiplo, os parâmetros de regressão devem ser estimados por um método de estimação. Com esses parâmetros já se tem a equação de regressão, mas antes de utilizar o resultado para solucionar o problema é preciso validar o modelo, verificar novamente a acurácia dos dados e constatar que todas as suposições feitas sobre a natureza do sistema estejam corretas (LONG, 1997).

2.2.1 REGRESSÃO LINEAR SIMPLES

Um modelo de regressão é dito linear quando seus parâmetros são lineares, sistemas com essa configuração são comumente denotados na forma demonstrada na Equação 1.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

A função consiste de uma parte determinística e uma parte aleatória. A parte determinística da equação, $\beta_0 + \beta_1 x$, especifica que para qualquer valor da variável independente x , a média populacional da variável dependente y é descrita por uma função reta. De acordo com as propriedades de uma função linear, o parâmetro β_0 , também chamado de coeficiente linear ou interceptação, é o valor de y quando x se iguala à zero. Graficamente, esse coeficiente representa o ponto onde a reta de regressão irá interceptar o eixo y . O parâmetro β_1 , conhecido como coeficiente angular ou inclinação, é a variação no eixo y associada à variação de uma unidade no eixo x . Finalmente, ε é a parte aleatória do modelo e utilizada para explicar a variabilidade das respostas em relação à média real. Esse parâmetro de erro é assumido como normalmente distribuído, com média zero e desvio padrão constante (CHATTERJEE, HADI, 2006).

Os parâmetros de regressão β_0 e β_1 devem ser estimados com base nos dados coletados das variáveis estudadas para encontrar a reta que faz o melhor ajuste dos pontos no gráfico. O método mais popular para estimar os valores desses coeficientes e determinar a equação de regressão é o de mínimos quadrados, que busca fornecer a reta que minimiza a soma dos quadrados das distâncias verticais de cada ponto em relação à reta. Essas distâncias representam o erro na variável de resposta, a diferença entre a resposta real e os valores estimados. A equação obtida pelo método dos mínimos quadrados se dá conforme indicado na Equação 2.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \varepsilon \quad (2)$$

A diferença entre o valor estimado \hat{y} e a resposta real y é chamada de valor residual e denotada por e . Valores residuais de regressão têm um papel importante no diagnóstico da regressão porque são observáveis, diferente do erro ε . O erro da regressão é quanto uma observação se distancia de seu valor esperado, baseado na população total da qual as unidades foram escolhidas aleatoriamente. Em contrapartida, o valor residual é uma estimativa observável de um erro não observável que pode ser utilizado para estimar o erro do modelo (FREUND; WILSON, 2006).

Depois de encontrar uma equação linear que relacione as variáveis, a análise não se interessa apenas em descobrir se essa relação existe, mas também em mensurar a qualidade da equação encontrada. A qualidade do modelo pode ser avaliada analisando os coeficientes lineares e de correlação, quanto maior for o valor absoluto desses coeficientes, mais forte será a relação entre as variáveis. A representação gráfica da reta de regressão também pode ser usada para determinar a qualidade do modelo, quanto mais próximos os pontos da amostra estiverem da reta, melhor será a regressão (FREES, 2010).

2.2.2 REGRESSÃO LINEAR MÚLTIPLA

Quando uma variável de resposta pode se relacionar linearmente com mais de uma variável preditora, temos um modelo linear múltiplo. A função desse modelo se dá na forma da Equação 3.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n + \varepsilon \quad (3)$$

Enquanto a equação de regressão linear simples era dada por uma reta, essa equação se dá como uma superfície e é chamada de regressão linear múltipla. Nesse tipo de regressão, o coeficiente β_0 é o valor de y para $x_1 = x_2 = x_n = 0$. Os coeficientes β_n são coeficientes parciais de regressão porque representam a contribuição de x_n para a resposta y depois de ambas variáveis terem sido linearmente ajustada para outras variáveis independentes (CHATTERJEE, HADI, 2006).

Para encontrar os valores dos parâmetros de regressão, os modelos lineares múltiplos são comumente representados em forma matricial, como mostrado equação 4. A forma matricial do modelo nos permite discutir e apresentar suas propriedades de maneira mais conveniente e eficiente.

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{pmatrix}$$

(4)

Os parâmetros de modelos de regressão múltipla são diferentes da regressão simples, uma vez que as variáveis independentes estão correlacionadas. Casos de variáveis independentes não correlacionadas são possíveis, mas extremamente raros em situações reais. A fim de avaliar a qualidade da regressão múltipla, além dos coeficientes já utilizados na regressão simples, é possível calcular o coeficiente de relação múltipla que mede a força da relação entre todas as variáveis do sistema (DRAPER; SMITH, 2006).

É geralmente impossível estimar os efeitos exclusivos que as variáveis preditoras exercem individualmente na equação, os parâmetros são muito sensíveis a mudanças nos dados e inclusão ou exclusão de variáveis. Os coeficientes de regressão possuem erros amostrais que podem afetar diretamente nas afirmações e previsões baseadas no modelo. As variáveis independentes de um modelo de regressão múltipla são independentes apenas no que diz respeito a variável de resposta, mas isso não quer dizer que elas não podem depender de outras variáveis. Não obstante, variáveis preditoras podem ser dependentes umas das outras e essa relação é necessária para que os parâmetros de regressão possam ser interpretados corretamente. Contudo, algumas variáveis possuem uma relação tão forte que podem prejudicar o resultado da análise (CHATTERJEE, HADI, 2006).

A relação muito forte entre variáveis preditoras também é conhecida como multicolinearidade, um problema difícil de detectar. Não se trata de um erro de especificação ou do modelo, multicolinearidade é um problema de dados deficientes, associada a coeficientes de regressão instáveis. Na verdade, a fonte do problema pode ser muito mais sutil que a simples relação entre variáveis. Um dos sinais da presença de multicolinearidade são mudanças significativas nos coeficientes estimados quando um ponto ou variável é adicionado ou retirado. Quando coeficientes de variáveis que são consideradas muito importantes possuem erros elevados, também pode indicar um problema. Para verificar a multicolinearidade, é preciso calcular o coeficiente de relação entre as variáveis. Aquelas que se mostrarem muito relacionadas devem ser substituídas ou excluídas do modelo de regressão (YAN, 2009).

2.3 SIMULAÇÃO DE SISTEMAS

A simulação está entre as técnicas de tomada de decisão mais utilizadas atualmente. Por ser uma ferramenta tão flexível, poderosa e intuitiva, a popularidade da simulação continua a crescer rapidamente (HILLIER; LIEBERMAN, 2012). Essa técnica consiste no processo de criação do modelo de um sistema real e condução de experimentos com esse modelo com o objetivo de entender o comportamento do sistema e avaliar várias estratégias para sua operação. Sendo assim, é importante que o modelo seja projetado de maneira a imitar o comportamento do sistema real conforme eventos ocorram ao longo do tempo (SHANNON, 1998).

A simulação possui várias vantagens sobre os modelos analíticos ou matemáticos de análise de sistemas. Primeiramente, o conceito básico de simulação é simples e mais fácil de justificar para administradores ou clientes. Um modelo de simulação também pode ser mais confiável porque seu comportamento é comparado ao sistema real e requer menos suposições simplificadoras, adotando mais das verdadeiras características do sistema estudado (SHANNON, 1998).

Entretanto, a simulação possui algumas desvantagens. Algumas vezes, softwares de simulação poder ser caros. Simular requer treinamento e a qualidade do modelo depende da habilidade da pessoa responsável. Além disso, coletar dados de um sistema pode ser uma tarefa exaustiva e o modelo não proverá uma solução ideal, mas um conjunto de informações que devem ser analisados e interpretados (CHUNG, 2004).

Os modelos de simulação podem ser classificados como estáticos ou dinâmicos. Modelos estáticos representam sistemas em um momento específico ou sistemas onde o tempo simplesmente não atua. Algumas estimativas de custos podem ser simuladas por modelos estáticos. Por outro lado, um modelo dinâmico representa um sistema que depende e evolui ao longo do tempo (LAW, 2015).

Uma simulação também pode ser determinística ou estocástica. Se um modelo não utiliza probabilidade, é um modelo determinístico. Um sistema de equações diferenciais descrevendo uma reação química é um exemplo de sistema determinístico. Entretanto, a maioria dos sistemas deve ser modelada com alguns dados de entrada aleatórios. Esses modelos são chamados estocásticos. Por exemplo, um sistema de filas precisa adotar algumas probabilidades para a chegada de novos clientes. Modelos estocásticos fornecem resultados aleatórios por natureza, que devem ser tratados apenas como uma estimativa das características reais do sistema (LAW, 2015).

A classificação mais importante de modelos de simulação se divide em contínuos e discretos. Simulação contínua é aquela onde as mudanças de estado do sistema ocorrem continuamente ao longo do tempo. Por exemplo, a posição de um avião durante o voo está sendo alterada continuamente. Simulações contínuas tipicamente requerem o uso de equações diferenciais para descrever a taxa de mudança das variáveis, o que torna sua análise mais complexa. A simulação de eventos discretos é aquela onde uma mudança no estado do sistema ocorre instantaneamente em pontos aleatórios do tempo como resultado de eventos discretos. Por exemplo, o estado de um sistema de filas muda quando um novo cliente chega. A maioria das aplicações práticas de simulação possuem eventos discretos. Entretanto, é possível usar uma simulação de eventos discretos para analisar o comportamento de um sistema contínuo se algumas aproximações forem realizadas (HILLIER; LIEBERMAN, 2012).

Em simulação de sistemas, entende-se por sistema um grupo ou conjunto de componentes que interagem entre si para atingir um objetivo específico. O comportamento de um sistema ao longo do tempo é estudado pelo desenvolvimento de um modelo de simulação. Esse modelo geralmente assume a forma de um conjunto de suposições sobre a operação do sistema. Essas suposições são expressas em relações matemáticas, lógicas e simbólicas entre as entidades do sistema (CHUNG, 2004).

Simular consiste em modelar um sistema através do estudo do fluxo de suas entidades. Entidades podem ser clientes, ordens de trabalho, partes, pacotes de informações, etc. Uma entidade pode ser qualquer objeto que entre no sistema, passe por uma série de processos e deixe o sistema. As propriedades individuais de uma entidade específica são chamadas atributos. À maneira que a entidade flui pelo sistema, ela passa por atividades que requerem recursos. Atividades representam quaisquer processos que levem um período de tempo. Recurso é qualquer objeto que seja necessário para realizar essas atividades. Se um banco está sendo estudado, clientes poderiam ser as entidades, o saldo de suas contas bancárias poderia ser um atributo, e fazer depósitos se encaixaria como uma atividade e o funcionário do caixa seria um recurso (BANKS et al, 2004).

Qualquer coisa que ocorra enquanto a simulação está sendo executada que pode afetar o estado de um dos componentes do sistema é um evento. Eventos típicos em uma simulação simples incluem entrada de entidades nas filas, o começo de um serviço e fim de um serviço. Esses eventos

mudam o estado do sistema porque eles podem aumentar ou diminuir o número de entidades no sistema ou mudar o estado de recursos entre ocioso e ocupado (CHUNG, 2004).

A essência e objetivo da simulação é ajudar na solução de problemas da tomada de decisão. Logo, um bom simulador deve unir boas técnicas de solução de problemas com boa prática em software de engenharia (SHANNON, 1978).

A primeira etapa do projeto de simulação é formular claramente o problema. Deve-se então estabelecer os objetivos e planejamento geral do projeto, esclarecendo as perguntas que a simulação deve responder. É preciso certificar que se tenham técnicos, profissionais e software suficientes e adequados para a realização do trabalho (SHANNON, 1978).

A conceituação do modelo do sistema envolve a habilidade de extrair as características essenciais do problema, selecionar e modificar suposições básicas que caracterizam o sistema e finalmente, enriquecer e elaborar o modelo até uma aproximação útil dos resultados. Assim, é melhor começar com um modelo simples e aumentar sua complexidade aos poucos. Uma boa conceituação do modelo deve aliar-se a uma boa coleta de dados. A coleta de dados toma uma grande parte do tempo necessário para a simulação. É preciso começar o mais cedo possível, geralmente junto com os estágios iniciais do projeto. Entretanto, os dados necessários para a simulação podem mudar conforme a simulação avança (BANKS et al, 2004).

Com a conceituação do modelo e os dados coletados, realiza-se a tradução do modelo. Durante a tradução do modelo, é necessário certificar que o modelo de simulação possui todos os componentes necessários e que pode ser executado pelo software. Na verdade, o simulador está interessado em não apenas executar o modelo, mas executá-lo da maneira correta. Ou seja, o modelo deve operar como foi projetado para operar. Esse processo é conhecido como verificação do modelo (CHUNG, 2004).

A etapa seguinte à verificação é a validação do modelo. Essas duas etapas são frequentemente confundidas. Validação é o processo de garantir que o modelo represente a realidade com determinado nível de confiança. Isso significa que o responsável pela simulação deve tentar criar um modelo que seja uma representação razoável do sistema real estudado. Porém, modelos cuidadosamente construídos ainda podem não representar a realidade. A inabilidade do modelo em representar o sistema real pode resultar de certas decisões tomadas pelo criador do modelo. Entre essas decisões, estão suposições, omissões, simplificações, descuidos e limitações (CHUNG, 2004).

A próxima etapa é o projeto experimental, que consiste na determinação de todas as alternativas de cenário que serão simuladas. Então, testes devem ser realizados e analisados para determinar que o modelo esteja concluído e sem defeitos. Todas as etapas devem ser documentadas através de relatórios e o modelo de simulação deve ser finalmente aplicado (BANKS et al, 2004).

Simulação é inquestionavelmente muito importante para a prática da tomada de decisões. Trata-se de uma ferramenta inestimável para os problemas onde técnicas analíticas se mostram inadequadas (HILLIER; LIEBERMAN, 2012).

2.3.1 MÉTODO DE MONTE CARLO

O Método de Monte Carlo surgiu durante uma pesquisa sobre armas nucleares de um grupo de cientistas em Los Alamos, onde foi sugerido o método para avaliar integrais matemáticas complicadas que surgem na teoria de reações nucleares em cadeia. A teoria envolvia a geração de uma amostra de números aleatórios para o cálculo das integrais e simulação dos resultados. A aleatoriedade dos modelos de simulação se baseia na geração de uma sequência de variáveis aleatórias independentes e uniformemente distribuídas em um dado intervalo. As primeiras técnicas de geração de números aleatórios utilizavam processos físicos, que eram aceitos como aleatórios, como jogos de azar (GENTLE, 2003).

Atualmente, números aleatórios são utilizados de várias maneiras associados a computadores, como em jogos e geração de dados sintéticos para testes. Contudo, isso não caracteriza uma simulação de Monte Carlo porque não produz resultados numéricos. O método de Monte Carlo envolve o uso deliberado de números aleatórios em um cálculo com a estrutura de um processo estocástico. Conforme apresentado anteriormente, processo estocástico é toda sequência de estados cuja evolução é determinada por eventos aleatórios. Em um computador, eles são gerados por um algoritmo determinístico que produz uma sequência de números pseudoaleatórios, que reproduz as propriedades de um sistema de números realmente aleatórios (KALOS; WHITLOCK, 2008).

O método de Monte Carlo fornece soluções aproximadas para uma variedade de problemas matemáticos através da execução de experimentos amostrais estatísticos em programas computacionais. Notavelmente, o método se aplica tanto a problemas sem conteúdo probabilístico quanto problemas com estrutura probabilística. Essa particularidade, porém, não representa uma

vantagem sobre outros métodos de aproximação. Contudo, entre todos os processos numéricos que se baseiam nas avaliações de n pontos em um espaço de m dimensões para apresentar uma solução, Monte Carlo possui a menor taxa de decrescimento do erro absoluto dos estimadores à medida que m aumenta. Essa propriedade coloca um grande diferencial para o método quando se trata de problemas maiores e mais complexos (FISHMAN, 1996).

Algumas distinções devem ser estabelecidas entre simulação e o método de Monte Carlo. Enquanto as observações na simulação estão serialmente correlacionadas porque são feitas em ordem cronológica, as observações no método de Monte Carlo são independentes e, conseqüentemente, tempo não tem um papel importante. No método de Monte Carlo, é possível expressar a resposta como uma simples função das variáveis estocásticas de entrada, mas a resposta da simulação é mais complexa e só pode ser explicitamente expressada pelo próprio programa computacional (RUBINSTEIN, 2006).

3 METODOLOGIA

Uma pesquisa pode ser motivada pelo simples anseio de conhecimento ou pela necessidade de se fazer algo de maneira mais eficaz e eficiente. Essa necessidade foi identificada nas operações do presente objeto de estudo e se tornou um ensejo para realização deste trabalho. Pesquisar é buscar soluções para um ou mais problemas propostos com base em um conjunto de procedimentos racionais e sistemáticos. A pesquisa se faz necessária quando não há informações disponíveis suficientes para gerar um conhecimento pleno sobre um problema, ou quando a informação disponível não se encontra organizada de forma que possa ser interpretada e solucionar o problema (GIL, 2002).

Do ponto de vista de sua natureza, o trabalho em questão se classifica como pesquisa aplicada, uma vez que busca gerar conhecimentos para a aplicação prática e solução de problemas específicos. O desenvolvimento desta pesquisa é essencialmente quantitativo, visto que utiliza métodos estatísticos e matemáticos para analisar dados das variáveis estudadas. Os resultados numéricos das análises de regressão, análises de variâncias e simulações são interpretados e transformados em informações para alcançar os objetivos da pesquisa. Diante dos objetivos deste projeto, a pesquisa pode ser vista tanto como exploratório. O estudo não somente busca descrever os métodos existentes de análise de variáveis, mas também desenvolver novos modelos facilitadores que aprimorem o entendimento do comportamento de variáveis e explique como esse fenômeno ocorre através de sequenciais análises de dados (SILVA; MENEZES, 2005).

O problema desta pesquisa surge diante do questionamento sobre a priorização de ações dentro da indústria. Internamente, acredita-se que os indicadores de qualidade e rendimento são conflitantes. A partir daí, o problema expandiu para o desconhecimento exato desse desequilíbrio e inexistência de modelos simples que facilitem a interpretação da relação entre variáveis e tomada de decisão baseada nesses resultados. Uma vez formulado o problema, foi utilizado o procedimento técnico de pesquisa bibliográfica em indicadores de desempenho e princípios estatísticos a serem utilizados na pesquisa.

Teoricamente fundamentado, foram definidas as variáveis de estudo com base em conhecimento prévio e conhecimento adquirido com a pesquisa bibliográfica e investigação in loco com pessoas mais experientes na área. Além das variáveis que representam os indicadores estudados, foram selecionadas variáveis secundárias para se relacionar com as variáveis principais.

Posteriormente, os dados históricos foram coletados para um período previamente estabelecido com o fim de garantir a normalidade e seguridade da amostra. Então, as amostras colhidas foram filtradas de maneira não tendenciosa para eliminar pontos anômalos que não representavam o comportamento de longo prazo das variáveis.

A partir dos dados coletados e selecionados, deu-se início ao desenvolvimento dos modelos estatísticos no programa de planilhas eletrônicas Microsoft Office Excel. Ainda que o programa forneça os coeficientes de uma regressão linear simples para um dado par de variáveis, foi decidido criar um modelo de equações a fim de facilitar o entendimento do cálculo e interpretação de resultados. O modelo de regressão linear simples seria utilizado para análise do relacionamento dos indicadores.

Com o objetivo de aprimorar a pesquisa, cada indicador foi analisado em relação a outras duas variáveis através de um modelo de regressão linear múltipla. Alinhado a esse modelo, foi criado um sistema de análise de variâncias para verificar a eficiência dos resultados da regressão linear. Finalmente, utilizando princípios do Método de Monte Carlo e Fatoração Cholesky alinhado modelo de análise de regressão linear, foi desenvolvido um modelo de simulação das amostras para grandes quantidades de observações. A simulação permite avaliar o comportamento das amostras ao longo prazo e prever valores de variável dependente para um dado conjunto de valores das variáveis independentes.

Depois de serem submetidos a testes com amostras aleatórias não reais, os modelos foram executados com os dados coletados na pesquisa. As hipóteses sobre a relação das variáveis foram definidas para serem aceitas ou rejeitadas após execução do programa. O resultado numérico obtido nas análises contém informações que levam o pesquisador a conhecer melhor o sistema estudado. Com base nessas informações, são propostas ações para a solução dos problemas.

4 DESENVOLVIMENTO

4.1 CARACTERIZAÇÃO DA EMPRESA

A empresa utilizada para aplicação dos modelos desenvolvidos neste estudo trata-se de uma multinacional no segmento alimentício, cujo nome não será citado a fim de manter o sigilo das informações fornecidas pela mesma. Os dados foram coletados de uma unidade fabril da empresa localizada em Dourados, Mato Grosso do Sul. A unidade é um frigorífico especializado em cortes de frango pesado, produzindo uma diversa gama de produtos. Apesar de também abastecer o mercado interno, a expressiva maioria da produção é enviada ao mercado externo, para clientes no Oriente Médio, Europa, África e Ásia.

A planta com capacidade instalada de abate de 160 mil aves por dia, funcionando em dois turnos de produção cinco dias por semana, com um quadro de funcionários de aproximadamente 1700 pessoas. A estrutura organizacional da planta conta com um gerente industrial no topo da hierarquia, um especialista de produtividade e processo, um supervisor de garantia da qualidade, um supervisor de manutenção, um supervisor de recursos humanos, onze supervisores de produção e suas respectivas equipes. Paralelo à unidade industrial, há uma unidade agropecuária responsável pelo fornecimento de matéria-prima.

4.2 SELEÇÃO DE DADOS

A produção da empresa é classificada em nove famílias: peito, perna, asa, miúdos, pés, cartilagens, pertences, especiais e outros. Para análise dessa pesquisa, foi selecionada a família perna, que inclui produtos como perna inteira desossada e perna porcionada. A perna inteira desossada é exportada para mercados asiáticos e possui um dos maiores índices de valor agregado entre os produtos fabricados na unidade.

Para o desenvolvimento de análise linear simples, foram coletados resultados diários de três meses dos indicadores de rendimento e conformidade de produto. Para a análise de regressão linear múltipla, análise de variâncias e simulação, acrescentou-se ao modelo duas variáveis levantadas na fábrica como possíveis influências nos resultados de rendimento e conformidade. Essas variáveis são peso médio das aves vivas e volume total de aves abatidas por dia, das quais se colocou a mesma quantidade de observações.

Após coletados, os dados foram analisados com o objetivo de transformar as observações em amostras mais confiáveis. A primeira seleção foi a limitação da amostra a dados

correspondentes aos dias entre segunda-feira a sexta-feira, desconsiderando os resultados de sábados. Essa decisão foi tomada porque abates realizados ao sábado não são ordinários e ocorrem em condições diferentes e circunstâncias especiais, o que poderia prejudicar a confiabilidade dos resultados das análises. Posteriormente, os dados foram submetidos a um gráfico de controle para visualização dos *outliers*. Os pontos que se distanciaram da média amostral mais de três vezes o desvio padrão foram considerados pontos anômalos e eliminados da amostra. Após seleção, as amostras apresentavam 45 observações.

4.3 INDICADOR DE RENDIMENTO

O rendimento de família (R_f) avalia a taxa de aproveitamento de uma família de produto, calculando a proporção de volume real de produto acabado (V_r) sobre o volume ideal de produto acabado (V_i). Para cada parte do frango, há um rendimento médio esperado (R_p) previamente estabelecido por veterinários baseado em variações no peso, linhagem e outras características da ave. Esse rendimento médio das partes representa a proporção do peso de cada parte, por exemplo, sobre o peso total do frango vivo. Logo, se obtém o volume ideal de produto acabado multiplicando o volume aproveitado de matéria prima pelo rendimento médio da parte correspondente a família. Quando se diz volume aproveitado de matéria prima, se faz referência apenas ao volume que entrou na planta e estava de fato disponível para produção, ou seja, o volume total recebido (V_t) descontado do volume de aves mortas (V_{mort}) e volume de aves condenadas (V_{cond}). É possível sintetizar o cálculo do rendimento de família como apresentado na Equação 5.

$$R_f = \frac{V_r}{V_i} = \frac{V_r}{(V_t - (V_{mort} + V_{cond})) * R_p} \quad (5)$$

Ilusoriamente simples, o rendimento geral calculado pela empresa não é apenas a expansão dos indicadores de rendimento de família para o volume total de produção. No cálculo do rendimento geral, é atribuído um peso diferente para cada *stock keeping unit* (SKU). Esses pesos são chamados de índices unitários e são atribuídos a cada item com base na composição do produto, rendimento médio das partes, e até mesmo, valor de mercado. O volume de produção individual de cada produto (V_j) multiplicado pelo seu respectivo índice unitário (I_u) retornará a quantidade total de índices. O resultado de rendimento geral (R_g) se dá pela proporção do total de índices em relação ao volume total recebido de matéria prima (V_t). Vale ressaltar que nesse indicador, não se utiliza o

volume aproveitado de matéria prima como no rendimento de família. Assim, as taxas de mortalidade de aves durante o transporte e condenação de aves pelo serviço de inspeção afetam diretamente no valor do rendimento geral. O cálculo de rendimento geral se dá conforme estabelecido na Equação 6.

$$R_g = \frac{\sum(V_j * I_u)}{V_t} \quad (6)$$

O rendimento é calculado diariamente referente à produção do dia anterior e referente à produção acumulada no mês até o dia anterior. Os indicadores são gerados em todas as unidades da empresa, o que possibilita um benchmarking e compartilhamento de boas práticas para o alcance de melhorias.

4.4 INDICADOR DE CONFORMIDADE

O Indicador de conformidade se refere à qualidade do produto, avaliando o atendimento dos requisitos exigidos pelo cliente. O resultado de conformidade é uma nota composta de cinco notas parciais, referentes a cada um dos grupos de atributos avaliados no indicador. Para todos os produtos, os grupos são embalagem primária; ossos, cartilagens e objetos estranhos; padrão de refile; peso; e outros defeitos. Porém, os atributos que compõe cada grupo variam conforme as características de cada produto. Na composição do indicador, cada grupo de atributos tem um peso diferente, peso que também varia para o mesmo grupo de um produto diferente.

As análises são realizadas por amostragem cinco vezes ao dia. No momento da análise, um monitor retira do processo uma amostra de tamanho igual ao especificado para aquele produto. Todas as peças da amostra selecionada são avaliadas individualmente para todos os atributos previstos no indicador de conformidade. Cada atributo recebe uma nota de acordo com a quantidade de peças encontradas com seu respectivo defeito. As notas vão de 0 a 3, sendo 0 o melhor resultado e 3 o pior. A quantidade limite de peças defeituosas para o enquadramento em cada faixa é claramente estabelecida para todos os atributos e definida conforme exigência do mercado consumidor.

A partir das notas dos atributos, é determinada a nota de cada grupo e a nota final do indicador de conformidade. O indicador diário é obtido pela média simples dos resultados das cinco análises.

4.5 MODELO DE REGRESSÃO

Empregando os recursos disponíveis no software, foi criado um modelo interativo de análise de regressão linear simples e múltipla. O sistema permite ao usuário calcular os coeficientes de regressão linear para qualquer conjunto de amostras de uma variável dependente e até cinco variáveis independentes.

O primeiro passo na execução do modelo é a informação pelo usuário do número de variáveis independentes, para que o sistema calcule o tamanho das matrizes e o número total de iterações. Então, o sistema mostrará títulos correspondentes ao número informado, orientado o usuário a inserir os dados das amostras em suas respectivas colunas. O número de observações por amostra é restrito apenas às limitações para quantidade de linhas do próprio programa. Ao inserir os dados, já são exibidos os valores dos índices de correlação (R) e determinação (R^2), coeficientes de regressão linear (β_i), e a tabela de análise de variâncias contendo soma dos quadrados, graus de liberdade e resultado do teste de distribuição de probabilidade F .

4.6 MODELO DE SIMULAÇÃO

Do ponto de vista prático, o modelo de simulação foi desenvolvido para projetar o comportamento das amostras correlacionadas no longo prazo e avaliar as variações na correlação mútua entre as variáveis. Tecnicamente, o modelo mostra a aplicação do método de Monte Carlo em uma análise de regressão linear. O desafio do algoritmo é simular um conjunto de amostras de milhares de observações que possua as mesmas características do conjunto original. As características desejadas nas amostras simuladas são média, desvio padrão, coeficiente de correlação e coeficientes de regressão.

Dois modelos diferentes foram desenvolvidos para simular regressão linear simples e múltipla. Na simulação de regressão simples, foi calculado o coeficiente a com base do coeficiente de correlação R , dado pela equação 7.

$$a = \frac{R}{\sqrt{1-R^2}}$$

(7)

O coeficiente a vai determinar que a correlação das amostras simuladas aleatoriamente seja equivalente à correlação real. Duas amostras x e y de dez mil números aleatórios foram criadas utilizando a fórmula inversa da distribuição normal com média igual a zero e desvio padrão igual

a um. Uma terceira amostra z é criada a partir das duas amostras aleatórias e do coeficiente a , para servir de variável pivô da simulação. Finalmente, duas amostras são simuladas a partir das médias e desvios padrão reais associados às variáveis x e z .

Os resultados dos coeficientes reais e simulados e o erro residual da simulação são apresentados em uma tabela de sumário. Pelo modelo, é possível obter os valores máximos e mínimos que cada variável pode exercer no longo prazo nas circunstâncias atuais, assim como o ponto de equilíbrio do sistema, onde as variáveis dependente e independente se igualam. Outro resultado importante da análise é o valor chamado no modelo de meta, que representa par de pontos com o melhor resultado mútuo para ambas variáveis. Essa meta é encontrada pelo produto máximo de cada par de observações das amostras simuladas. Ainda a partir da equação de regressão linear, o usuário pode inserir o valor de uma variável e receber a previsão do valor esperada para a outra variável.

Além disso, o modelo fornece uma tabela de probabilidades, onde o usuário fornece as metas ou valores desejados para cada variável e recebe do sistema a probabilidade das amostras alcançarem esses valores individualmente e em conjunto. Primeiramente, a tabela apresenta a probabilidade calculada a partir da média e desvio padrão das amostras simuladas, considerando uma distribuição normal. Em seguida, consta na tabela a proporção real dos pontos dentro das amostras simuladas que atingiram a meta determinada pelo usuário. Finalmente, é calculado o erro diferencial entre as duas porcentagens.

Para simular um grupo de mais de três amostras, a fatoração de Cholesky fez o papel do coeficiente a utilizado anteriormente. Comumente utilizada na solução matemática de problemas lineares, a fatoração de Cholesky é a decomposição de uma matriz simétrica e definida positiva no produto da matriz triangular inferior e sua transposta hermitiana. A decomposição de Cholesky se dá na forma da Equação 8, onde L é uma matriz triangular inferior com diagonal real e positiva e L^T é a transposta hermitiana de L . (Krishnamoorthy)

$$A = LL^T \tag{8}$$

A matriz de Cholesky é usualmente aplicada no método de Monte Carlo para simular sistemas com múltiplas variáveis correlacionadas. A matriz L , obtida pela decomposição da matriz de correlação do sistema, é aplicada a um vetor u de amostras aleatórias não correlacionadas. O vetor produzido Lu possui as propriedades de correlação e covariação do sistema sendo simulado.

A matriz de correlação e sua decomposição são calculadas automaticamente pelo modelo quando as amostras são inseridas pelo usuário. Executada a simulação, é possível observar os valores máximos e mínimos alcançados pelos indicadores com base nas variáveis preditoras, assim como prever o valor da variável de resposta.

5 RESULTADOS E DISCUSSÕES

5.1 ANÁLISE SIMPLES DE INDICADORES

Os indicadores de Rendimento e Conformidade do produto perna desossada foram analisados a fim de testar uma tendência inversamente proporcional no seu relacionamento. Para os objetivos desse estudo, a definição das variáveis dependentes e independentes não interfere no resultado. Assim, os dados foram inseridos em ordem cronológica, onde y representa o indicador de rendimento de família e x1 o indicador de conformidade do produto. As 45 observações das amostras estão no Quadro 1.

Y	X1	Y	X1	Y	X1	Y	X1	Y	X1
97,63	74,58	96,67	75,17	96,60	77,68	91,04	81,65	93,87	75,41
97,07	75,53	95,82	71,27	98,03	79,19	91,35	81,34	95,25	81,39
97,76	72,67	94,92	76,00	94,63	72,80	98,62	74,73	98,57	77,71
98,23	75,39	96,38	74,73	96,03	77,11	92,04	81,15	99,78	78,71
94,06	74,00	98,84	72,65	99,78	73,00	95,17	83,82	97,61	73,58
97,64	74,40	97,69	77,79	98,34	74,98	92,33	82,58	94,52	73,52
98,96	73,97	94,73	78,17	90,35	82,75	92,37	87,22	93,11	80,50
94,38	74,65	98,25	75,64	94,52	73,52	90,05	84,89	90,02	83,92
94,18	75,31	99,98	68,00	90,01	83,03	97,89	75,38	99,82	75,47

Quadro 1 - Dados coletados dos indicadores de rendimento e conformidade

Após a inserção das amostras, o modelo apresenta os resultados em tabelas, como pode ser visualizado na Quadro 2, que mostra a interface do modelo com os resultados da análise. Antes de estudar os resultados da análise, as mesmas amostras foram analisadas por um software especializado para testar a eficácia do modelo. O software utilizado para teste foi o Minitab 17, que apresentou os mesmos resultados do modelo desenvolvido em Excel. A diferença nula na comparação dos coeficientes β_0 , β_1 , R^2 e R gerados pelos dois sistemas confirma a eficácia e acurácia no projeto e concepção do modelo. A Figura 3 mostra um sumário estatístico gerado no Minitab.

Insira um número de variáveis dependentes entre 1 e 5	1	
Insira os valores das variáveis na tabela ao lado		

R	R ²	β ₀	β ₁
0,7872	0,6196	79,8040	0,2051

Fonte	Soma dos quadrados	Graus de liberdade	Soma dos q. média	F calculado
Devido à regressão	1076,89	1	1076,89	79,82
Devido ao erro	661,12	49	13,49	
Total	1738,01	50	F tabelado	4,04

Y	X1
97,63	74,58
97,07	75,53
97,76	72,67
98,23	75,39
94,06	74,00
97,64	74,40
98,96	73,97
94,38	74,65
94,18	75,31
96,67	75,17
95,82	71,27
94,92	76,00
96,38	74,73

Quadro 2 - Resultados do Modelo de Análise de Regressão Simples

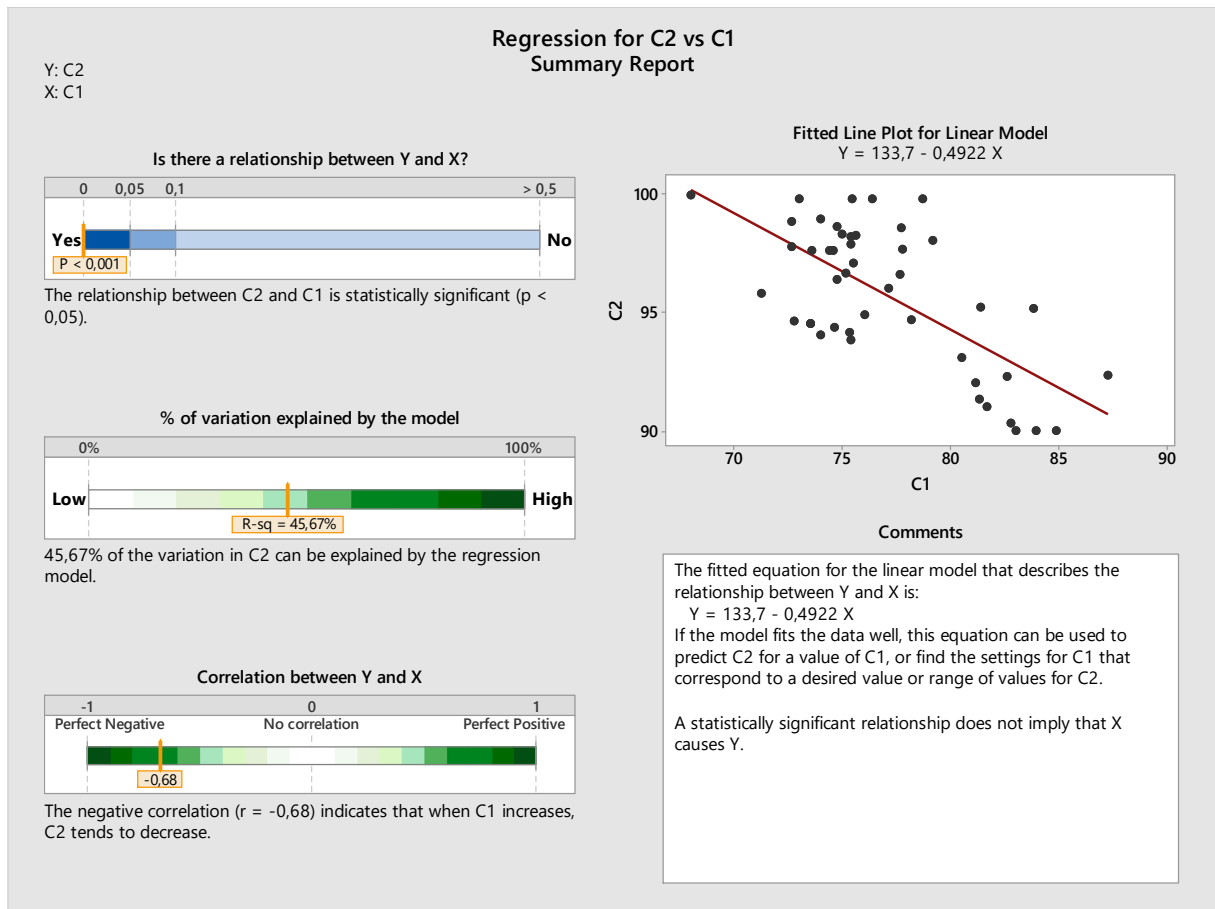


Figura 1 - Relatório de Análise de Regressão do Minitab

O valor absoluto do coeficiente de correlação R de 0,6758 é bastante expressivo e indica uma forte dependência linear entre as variáveis. Por ser negativo, o resultado indica que quando um indicador aumenta, o outro tende a diminuir. O coeficiente de correlação não avalia variáveis dependente e independente, mas sim o comportamento que as duas apresentam em um mesmo ponto no tempo. Logo, é importante ressaltar que correlação não indica causalidade, apenas que as variações em uma amostra se assemelham às variações na outra. O resultado confirma a hipótese levantada durante pesquisa de campo que sugeria um desequilíbrio entre os indicadores de rendimento e conformidade.

$$\text{Rendimento} = 133,7 - 0,4922 * \text{Conformidade}$$

(9)

Na equação da reta de regressão linear, dada pela Equação 9, o coeficiente β_1 é negativo, indicando mais uma vez a relação inversa entre as variáveis, e baixo, indicando que ambas variáveis variam em escalas aproximadas. A eficiência do sistema em representar o sistema real é dada por R^2 ajustado de 0,4567, ou seja, 45,67% das variações nas variáveis podem ser explicadas pela equação de regressão. O ponto de equilíbrio entre as duas amostras calculado a partir da equação de regressão é de 89,46, considerado na empresa como um resultado razoável para o indicador de conformidade e um resultado ruim para o indicador de rendimento. A análise de regressão não prova que é impossível se alcançar resultados satisfatórios para ambos indicadores de desempenho, apenas mostra isso não acontece na situação real e atual. Os resultados da tabela ANOVA não serão utilizados nesse caso, uma vez que são mais direcionados às análises de regressão múltipla.

A análise de regressão foi concluída e as amostras foram submetidas ao modelo de simulação. O usuário é orientado a inserir os dados das amostras nas colunas correspondentes. A partir desses dados, o modelo irá executar mil replicações do conjunto de variáveis. As variáveis reais e simuladas são apresentadas em tabelas, como é evidenciado no Quadro 3 com parte das tabelas geradas na análise. As primeiras colunas da tabela, na cor amarela, são as colunas onde o usuário deve inserir os dados coletados durante a pesquisa, determinando a variável dependente e a variável independente. As amostras aleatórias geradas pela simulação das propriedades de regressão das variáveis originais em milhares de replicações são apresentadas na segunda parte da tabela. A variável de resposta é chamada de y' e a variável preditora é chamada de x' .

Insira os dados das amostras nas colunas correspondentes.		Aqui estão as amostras simuladas com mil observações.	
X	Y	X'	Y'
74,58	97,63	82,34	91,86
75,53	97,07	82,00	93,38
72,67	97,76	78,90	96,74
75,39	98,23	68,76	97,63
74,00	94,06	75,72	96,88
74,40	97,64	75,54	95,13
73,97	98,96	71,49	100,03
74,65	94,38	74,40	95,31
75,31	94,18	82,94	95,33
75,17	96,67	77,73	98,94

Quadro 3 - Valores amostrais e simulados dos indicadores de rendimento e conformidade

O primeiro resultado da simulação é uma comparação dos coeficientes das amostras resultantes da simulação em relação às amostras de dados coletados no sistema real. Os erros médios, que são apresentados no Quadro 4, são muito satisfatórios. O índice de correlação obteve um erro de 0,12%, enquanto os coeficientes linear e angular apresentaram desvios de 0,04% e 0,12%, respectivamente. Esse desempenho indica que a amostra gerada pelo modelo se aproxima muito das amostras reais, aumentando o nível de segurança para tomada de decisão baseada em informações extraídas da análise de simulação.

Coefficientes	R	Linear (a)	Angular (b)
Real	-0,68	133,66	-0,49
Simulado	-0,68	133,71	-0,49
Erro	0,12%	0,04%	0,12%

Quadro 4 - Comparação dos coeficientes de regressão e correlação originais e simulados

O modelo de simulação oferece um resumo das características das variáveis, por exemplo, os valores máximos e mínimos que podem ocorrer diante das condições atualmente vigentes. Como pode ser visto na Quadro 5, os pontos máximos dos indicadores de conformidade e rendimento foram 92,92 e 104,90, indicando que mesmo sem nenhuma alteração ou melhoria no sistema, é

possível obter resultados significativamente melhores que os atuais. Por outro lado, os valores mínimos obtidos na simulação foram 63,35 para conformidade e 85,94 para rendimento, mostrando que o sistema ainda é suscetível a resultados muito inferiores. Esses resultados expõem a falta de controle no desempenho de cada indicador individualmente. A distância entre os pontos máximos e mínimos deve ser a menor possível, aqui os dois indicadores apresentaram uma extensão provável de variação muito grande. O sistema se encontra em uma situação muito instável, mostrando a necessidade de ações imediatas para eliminação de anomalias e melhoria no controle das variáveis.

Simulação	Conformidade	Rendimento	Paridade Média
Máximos	92,92	104,90	
Mínimos	63,35	85,94	0,54
Meta	90,87	91,04	0,82

Quadro 5 - Resultados da Simulação e Meta

No Quadro 5, ainda encontramos outro resultado importante da simulação para análise de regressão simples, o índice de paridade. Esse índice foi desenvolvido com o objetivo de encontrar o melhor resultado mútuo possível, alinhando equilíbrio e grandeza dos indicadores. O índice de paridade pode ser utilizado apenas para variáveis dentro da mesma escala numérica, o que normalmente é o caso de indicadores de desempenho. Isso porque esse valor não está ligado à correlação, mas sim à distância entre as médias de cada variável. Se o índice de paridade for baixo, podemos afirmar que os resultados dos indicadores estão ruins e desequilibradas. Quanto mais próximo de um estiver o resultado, melhores e mais próximos estarão os indicadores. O índice de paridade é dado por uma função dos indicadores e da diferença entre eles, pois de nada adianta ter resultados equilibrados se os mesmo forem muito ruins.

A paridade média de 0,54 mostrada na tabela mostra como o sistema opera atualmente, sem nenhum equilíbrio entre os indicadores. O índice que aparece na linha de meta representa a melhor paridade encontrada entre todos os pontos das amostras simuladas, mostrando o valor máximo que se espera alcançar para esse índice nas condições atuais do sistema. Os valores referentes aos indicadores de rendimento e conformidade na linha meta são os que apresentam as melhores condições para atingir a paridade máxima.

Esse índice de paridade pode ser utilizado na definição de metas para o processo, uma vez que se busca não somente o equilíbrio, mas também elevar os resultados de ambos indicadores. No conjunto de valores apresentados como resultados ideais dos indicadores para atingimento da meta

de paridade, o rendimento é menor que a média já realizada atualmente. Isso se dá porque os resultados de conformidade estão muito inferiores aos de rendimento, portanto pode-se definir a priorização de ações de qualidade. Uma vez alcançado o ponto de equilíbrio, ambos indicadores podem evoluir juntos, elevando ainda mais o índice de paridade do sistema.

Além de prover o usuário com resultados ideais a serem estabelecidos como metas e apresentar cenários favoráveis ao alcance dos mesmos, o modelo de simulação informa a probabilidade de atingimento dessas metas ou de qualquer outro valor que se deseje conhecer. O recurso pode ser visualizado na Quadro 6.

Insira o valor desejado para as variáveis	Conformidade	Rendimento
	90,00	94,00

Probabilidade de obter o resultado desejado			
Distribuição	Conformidade	Rendimento	Ambos
Normal	0,08%	72,17%	0,06%
Simulação	0,00%	72,20%	0,00%
Erro	0,08%	0,03%	0,06%

Quadro 6 - Ferramenta de Probabilidades de Atingimento de Meta

Nesta tabela de resultados, é dada a probabilidade de cada variável atingir os valores desejados individualmente, assim como a probabilidade de as duas variáveis alcançarem os números exigidos simultaneamente. Primeiramente, a probabilidade é calculada pela distribuição normal a partir da média e desvio padrão das amostras geradas na simulação. Então, o modelo apresenta a proporção real de pontos das amostras que alcançaram os valores inseridos pelo usuário. Nesse caso, foram inseridas as metas previamente estabelecidas pela empresa analisada, 90 para o indicador de conformidade e 94 para o indicador de rendimento. As diferenças entre as probabilidades calculadas e reais foram significativamente baixas. A probabilidade de o rendimento atingir a meta estabelecida na empresa é de 72,20%, enquanto a conformidade tem apenas 0,08% de chances de atender a meta dentro das circunstâncias correntes do sistema. A probabilidade múltipla calculada pela distribuição normal para os dois indicadores atingirem a meta é 0,06%. Contudo, em nenhum dos milhares de pares de pontos das amostras, as duas metas foram atingidas ao mesmo tempo.

5.2 ANÁLISE MÚLTIPLA DE RENDIMENTO

Uma relação inversamente proporcional entre os indicadores de rendimento e conformidade foi encontrada a partir dos modelos de análise simples. Nos modelos de análise múltipla, é possível adicionar variáveis ao sistema e aprimorar os resultados da pesquisa como um todo. As variáveis escolhidas para análise múltipla foram o peso médio das aves e a quantidade de aves abatidas no dia, a serem confrontadas com os indicadores de rendimento e conformidade em análises separadas.

Os dados do indicador de rendimento foram inseridos no modelo misto de regressão como variável de resposta, enquanto os valores de peso médio e quantidade de aves abatidas foram inseridos como variáveis preditoras. Os dados de cada variável foram inseridos para suas datas correspondentes, ou seja, cada conjunto de valores em uma mesma linha do modelo correspondia a valores coletados do mesmo dia.

O sistema apresentou um coeficiente de determinação R^2 muito baixo de 0,0542, resultado que indica que a relação entre as três variáveis é pouco significativa. Após avaliação desse resultado, é necessário analisar a tabela de análise de variâncias. Para um nível de confiança estabelecido para o modelo de 95%, o valor de F tabelado de um sistema com os mesmos graus de liberdade é 3,19. Para inferir que a análise de regressão confirma uma relação interdependente no sistema, espera-se que o valor de F calculado seja maior que F tabelado. Contudo, o resultado obtido de 1,35 é menor que 3,19. Logo, podemos afirmar que a hipótese de pelo uma das variáveis apresentar uma variância igual é falsa.

Esse resultado significa que não há nenhuma relação entre as amostras selecionadas, derrubando uma teoria existente dentro da empresa de que quanto maior o peso médio, melhor será o indicador de rendimento. Apesar de indicar significativamente que o peso médio e quantidade de aves abatidas não são fatores causadores de variações no rendimento, não se pode anular completamente a hipótese deduzida por trabalhadores experientes na área. O processo pode estar operando muito fora dos limites de controle, o que provoca muitas anomalias que são impossíveis de detectar através desse modelo. Não é possível observar tendências em um cenário fora de controle porque nenhuma estabilidade se sustenta por muito tempo.

5.3 ANÁLISE MÚLTIPLA DE CONFORMIDADE

Após executar a regressão múltipla para o indicador de rendimento, a mesma foi realizada para o indicador de conformidade. A variável correspondente à conformidade foi definida como variável dependente e as outras duas como variáveis independentes, sendo x_1 o peso médio e x_2 a quantidade de aves abatidas. Os resultados são apresentados no Quadro 7.

Insira um número de variáveis dependentes entre 1 e 5		2		
Insira os valores das variáveis na tabela ao lado				
R	R ²	β_0	β_1	β_2
0,4852	0,2354	22,7069	6,7462	0,0002
Fonte	Soma dos quadrados	Graus de liberdade	Soma dos q. média	F calculado
Devido à regressão	315,18	2	157,59	7,04
Devido ao erro	1051,70	47	22,38	
Total	1366,88	49	F tabelado	3,20

Y	X1	X2
75,53	2,611	159956
72,67	2,643	152458
80,79	2,722	169484
75,39	2,670	163395
74,00	2,548	162704
74,40	2,661	153078
73,97	2,589	153444
74,65	2,540	156574
75,31	2,586	170350
75,17	2,536	166559
71,27	2,407	165060
82,62	2,549	155566
76,00	2,557	159514

Quadro 7 - Resultados do Modelo de Análise de Regressão Múltipla

Os coeficientes calculados pelo modelo de regressão apresentam erro de menos de 1% em relação aos resultados obtidos em softwares estatísticos mais complexos, como o Minitab. O índice de correlação de 0,4852 indica uma relação linear razoável entre as três variáveis, enquanto o coeficiente de determinação mostra que 23,54% das variações na variável de resposta podem ser explicadas pela equação de regressão linear. Com base nos coeficientes resultantes da análise de regressão, a equação da reta é dada da Equação 10.

$$\text{Conformidade} = 22,7069 + 6,7462 * \text{Peso Médio} + 0,0002 * \text{Quantidade Abatida}$$

(10)

Contudo, esses números ainda não são suficientes para confirmar uma relação linear entre as variáveis, para isso é realizada a análise de variâncias. Na tabela de análise de variâncias, é possível identificar que o valor de F calculado é maior que F tabelado. Assim, pode-se concluir com risco de 0,05% que existe regressão linear múltipla e o modelo pode explicar e prever a variável dependente. As variáveis aplicadas no modelo estão em escalas muito diferentes, o que

torna difícil avaliar a força de cada variável apenas pela interpretação da equação da reta de regressão.

Estabelecida a relação linear entre as variáveis, as mesmas podem ser submetidas ao modelo de simulação para aprimorar a pesquisa e obter informações importantes. As variáveis foram inseridas no modelo de regressão e parte das tabelas é apresentada na Quadro 8.

Insira os dados das amostras nas colunas correspondentes.			Aqui estão as amostras simuladas com mil observações.		
X1	X2	Y	X1'	X2'	Y'
2,611	159.956	75,53	2,766	167.041	77,40
2,643	152.458	72,67	2,795	157.838	80,64
2,722	169.484	80,79	2,477	152.478	75,91
2,670	163.395	75,39	2,678	148.648	74,90
2,548	162.704	74,00	2,797	149.886	69,81
2,661	153.078	74,40	2,558	141.382	74,61
2,589	153.444	73,97	2,917	168.630	79,35
2,540	156.574	74,65	2,592	160.284	71,46
2,586	170.350	75,31	2,487	150.067	74,14
2,536	166.559	75,17	2,697	151.419	69,36

Quadro 8 - Valores amostrais e simulados de peso médio, volume de abate e indicador de conformidade

Conforme explicado anteriormente, o modelo de simulação para análise de regressão linear múltipla gera um conjunto de variáveis com os mesmos índices de correlação das variáveis inseridas pelo usuário. O quadro 9 mostra a matriz de correlação das amostras de dados coletados no sistema real e a matriz de correlação das amostras geradas pelo modelo de simulação. O erro de cada um dos índices simulados em relação ao valor real é muito baixo, o que evidencia a eficiência da utilização da decomposição matricial de Cholesky na reprodução das amostras do sistema real.

Matriz de Correlação Real			
	X1	X2	Y
X1	1,0000	0,0352	0,1844
X2	0,0352	1,0000	0,4496
Y	0,1844	0,4496	1,0000

Matriz de Correlação Simulada			
	X1	X2	Y
X1	1,0000	0,0184	0,1981
X2	0,0184	1,0000	0,4352
Y	0,1981	0,4352	1,0000

Quadro 9 - Comparação entre Matrizes de Correlação Real e Simulada

As médias e desvios padrão são utilizados como base das operações matriciais que geram as amostras de números aleatórios, logo esses números não variam no processo de simulação. Os coeficientes de regressão constam na Quadro 10, com um erro de até 6,0% em relação aos coeficientes originais. Analisando os erros nos índices de correlação e coeficientes de regressão, aceita-se a reprodução das amostras do modelo de simulação como uma representação confiável do sistema real.

Coefficientes de Regressão			
	β_0	β_1	β_2
Real	22,71	6,75	0,00
Simulado	21,35	7,13	0,00
Erro	6,0%	5,7%	0,8%

Quadro 10 - Comparação entre Coeficientes de Regressão Real e Simulado

Assim como no modelo para regressão simples, esse modelo múltiplo também oferece os valores máximos e mínimos para cada variável, como demonstrado no Quadro 11. A simulação da análise simples já mostrava uma extensão de variação muito grande no índice de conformidade, o mesmo se repete para as variáveis de peso médio e quantidade abatida. Para um indicador de conformidade dependente de ambas variáveis, esse descontrole pode levar a condições preocupantes.

Resultados da Simulação			
	X1	X2	Y
Máximos	3,121	186359	93,30
Mínimos	2,314	121839	60,79
Medianas	2,704	158561	77,32

Quadro 11 - Resultados da Simulação

Na simulação de múltiplas variáveis, o usuário é orientado a inserir o valor desejado para a variável de resposta, a partir da qual o sistema calculará o conjunto de valores ideais para as variáveis independentes. Os valores ideais de cada variável independente são calculados pelas médias dos pontos nas amostras de simulação nos quais a resposta correspondente atingiu o valor desejado pelo usuário. Os resultados estão no Quadro 12 e mais uma vez, foi utilizada como exemplo a meta de indicador de conformidade buscada pela empresa, que é 90,0. Para essa meta, o cenário mais favorável é aquele com um peso médio de 2,716kg e um volume diário de abate de 165656 aves.

Insira o valor desejado da variável de resposta.	Y
	90,00

Aqui estão os valores ideais para obter o resultado desejado.	X1	X2
	2,716	165656

Quadro 12 - Ferramenta de Determinação de Valores para Atingimento de Meta

A partir dos valores ideais encontrados para cada variável independente, o modelo traz as probabilidades da simulação. A probabilidade apresentada no Quadro 13 é calculada individualmente para cada variável pela função cumulativa de distribuição normal e pela proporção real encontrada nas amostras geradas pela simulação.

Probabilidade de obter o resultado desejado				
Distribuição	X1	X2	X1 e X2	Y
Normal	46,1%	23,6%	10,9%	0,9%
Simulação	45,7%	23,4%	11,0%	0,7%
Erro	0,4%	0,2%	0,1%	0,2%

Quadro 13 - Ferramenta de Determinação das Probabilidades de Atingimento da Meta

O processo apresenta chances bem difíceis na estabilidade de resultados para todas as variáveis. Com oportunidades ainda ruins, o processo apresenta a melhor eficiência quando se trata de peso médio, onde se identificou uma probabilidade de aproximadamente 46% de atingimento da meta. A probabilidade da variável de quantidade abatida alcançar o valor ideal é ainda mais baixa, em torno de 23,5%. Apenas 11% dos milhares de conjuntos de variáveis gerados pela simulação obtiveram os valores ideais tanto para peso médio quanto para quantidade abatida,

configurando assim o cenário definido como ideal para alcançar a meta estabelecida para o indicador de conformidade. Contudo, a quantidade de pontos da variável de resposta que estiveram acima da meta foi inferior a 1%. Essa diferença se dá porque as variáveis analisadas nesse caso não são as únicas atuando no sistema. Portanto, operar com os valores ideais de peso médio e volume de abate aumenta as chances de bons resultados, mas não garante o atingimento da meta do indicador de conformidade. Em uma simulação ideal, incluindo todas as variáveis atuantes no sistema, a probabilidade de obter os valores ideais em todas as variáveis independentes deve ser a mesma para o atingimento da meta da variável dependente.

No estabelecimento de metas, o conjunto ideal de valores deve ser definido como o objetivo para obter estabilidade no sistema e oferecer um cenário propício à melhoria nos resultados de conformidade. A estabilidade na variável de resposta será obtida através da estabilidade de seus fatores influenciadores.

6 CONSIDERAÇÕES FINAIS

A análise de regressão tem o objetivo de estabelecer a relação entre duas ou mais variáveis, o que foi realizado com sucesso pelo modelo desenvolvido nesta pesquisa. O modelo baseado no método dos mínimos quadrados foi bem sucedido na entrega da equação da reta de regressão para as variáveis analisadas, apresentando resultados muito aproximados àqueles gerados por programas que são referência neste tipo de análise. Com uma estrutura interativa e agradável ao usuário, o modelo possibilitou a execução de análises de regressão linear simples e múltiplas, entregando informações essenciais como os coeficientes de regressão, coeficiente de correlação e coeficiente de determinação. Tornando o modelo ainda mais completo, a análise de variâncias ofereceu um nível de confiança maior aos resultados.

A simulação da análise de regressão abriu novos horizontes na interpretação dos indicadores de desempenho. Informações essenciais ao gerenciamento dos sistemas de medição de desempenho foram extraídas, entre elas, o índice de paridade. O indicador desenvolvido pela pesquisa oferece uma maneira eficaz de controle das medidas de desempenho. Ao acompanhar a evolução no índice paridade, é possível avaliar as melhorias individuais nos indicadores de conformidade e rendimento enquanto se controla o equilíbrio de resultados. O controle permite identificar quando um indicador é tratado com mais prioridade, ou ações que melhoram o resultado em um indicador enquanto prejudicam o outro. Contudo, a pesquisa leva a crer que a existente disparidade entre rendimento e conformidade trata-se apenas de uma configuração atual do processo e não, necessariamente, uma propriedade do sistema. O entendimento mais claro dessa situação por meio das análises de regressão e simulação vai facilitar a gestão e levar à proposição de ações eficazes que atuaram direto no problema.

No desdobramento dos índices em análises de regressão múltipla, verificou-se pouca ou nenhuma relação de dependência entre o indicador de rendimento e as variáveis de peso médio e quantidade abatida. Essas variáveis independentes foram selecionadas depois de pesquisa na empresa estudada, pois a grande maioria dos funcionários afirmava seguramente que o peso da matéria prima e quantidade de aves tinham impacto direto e explícito no índice de rendimento. As análises provaram o contrário, mesmo que as pessoas com experiência na área e diretamente envolvidas nas operações tem um embasamento para afirmar tal relação. O problema pode estar no próprio indicador, no que se refere ao cálculo e composição e interferências externas. Sem análises mais detalhadas, conclui-se com embasamento teórico e estatístico que essa relação é inexistente.

Enquanto na análise múltipla do indicador de rendimento, o resultado contradisse a experiência dos especialistas na empresa, na análise múltipla de conformidade com as mesmas variáveis independentes, a correlação apareceu. Ainda que relativamente baixa, a inter-relação entre as três variáveis foi encontrada na análise de regressão e comprovada na análise de variâncias. A simulação para análise múltipla também foi bem sucedida, apresentando resultados satisfatórios. Através de cálculos implícitos do modelo, foi possível identificar os valores ideais de peso médio e quantidade abatida que comporiam o melhor cenário para obtenção de determinado valor de conformidade. Estabelecendo esses valores como meta dentro da organização, é possível alcançar melhorias graduais. Ao alcançar o resultado indicado pelo modelo, a análise pode ser executada novamente para atualização da meta e avaliação da evolução no desempenho dos indicadores.

Os modelos de análise de regressão e simulação oferecem ao usuário um diagnóstico do sistema de medição de desempenho. Para o caso exemplificado nesta pesquisa, foi evidenciada uma situação muito desfavorável, principalmente em relação ao índice de conformidade que raramente consegue bons números. O descontrole e instabilidade são sintomas de um sistema falho, que em nenhuma circunstância irá operar com a eficácia e eficiência que deveria. A interpretação dos resultados não somente identifica a saúde dos indicadores de desempenho, mas viabiliza oportunidades e define objetivos para a melhoria do processo. Cada resultado pode ser interpretado de maneiras diferentes, incitando o surgimento de novas ideias.

Essa flexibilidade dos métodos utilizados neste estudo é que o tornam diferente. Na execução das análises, foi possível verificar um grande potencial do modelo para aprimoramento. Outras avaliações podem ser incorporadas ao modelo, por exemplo, variáveis categóricas. Ainda no setor empresarial aqui pesquisado, poderia se incluir uma variável referente à linhagem das aves abatidas. A ideia que fundamenta esta pesquisa pode ser moldada na melhoria desse modelo, na criação de outros modelos e até mesmo, no desenvolvimento de novas práticas a serem adotadas na área de solução de problemas.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- BANKS, J.; CARSON J. S.; NELSON, B.; NICOL D. M. **Discret-event system simulation**. 4th ed. 2004.
- CHATTERJEE, S; HADI, A. S. **Regression Analysis by Example**. 4th ed. New Jersey, 2006 (Wiley Series in Probability and Statistics)
- CHUNG, C. **Simulation Modeling Handbook: A practical approach**. Florida, 2004.
- DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. 3rd ed. New Jersey, 2006 (Wiley Series in Probability and Statistics).
- FISHMAN, G. S. **Monte Carlo: Concepts, Algorithms and Applications**. Hong Kong, 1996
- FRANCESCHINI, F.; GALETTO, M.; MAISANO, D. **Management by Measurement: Designing Key Indicators and Performance Measurement Systems**. New York, 2007.
- FREES, E. W. **Regression modeling with actuarial and financial applications**. New York, 2010. (International Series on Actuarial Science) ISBN 978-0-511-67528-7.
- FREUND, R. J.; WILSON, P. S. **Regression analysis: statistical modeling of a response variable**. 2nd ed. New York, 2006. ISBN 0-12-088597-2.
- GENTLE, J. E. **Random number generation and Monte Carlo methods**. New York, 2003. ISBN 0-387-0017-6
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4 ed. São Paulo: Atlas, 2002. ISBN 85-224-3169-8.
- HARBOUR, J. L. **The Basics of Performance Measurement**. 2nd ed. New York, 2009.
- HILLIER, F. S.; LIEBERMAN, G. J. **Introduction to Operations Research**. 9th ed. New York, 2012. McGraw Hill Higher Education, ISBN 007123828X
- KALOS, M. H.; WHITLOCK, P. A. **Monte Carlo Methods**. 2nd ed. Weinheim, 2008.
- KAPLAN, R. S.; NORTON, D. **Putting the Balanced Scorecard to Work**. Harvard Business Review 71, no. 5 (Sep-Oct 1993): 134–147.
- KAPLAN, R. S.; NORTON, D. **The Balanced Scorecard: Measures that Drive Performance**. Harvard Business Review 70, no. 1 (Jan-Feb 1992): 71–79.
- KAYDOS, W. J. **Operational Performance Measurement: Increasing total productivity**. Florida, 1998.
- LAW, A. M. **Simulation Modeling and Analysis**. 5th ed. McGraw-Hill, 2015.
- LONG, J. S. **Regression models for categorical and limited dependent variables**. India, 1997.
- MARK, S.; MORDECHAI, S. **Applications of Monte Carlo Method in Science and Engineering**. India, 2011.
- MERCHANT, K.; STEDE, W. A. **Management control systems: performance measurement, evaluation and incentives**. Harlow, England, 2007.

NEELY, A.; ADAMS, C.; KENNERLEY, M. **The Performance Prism: The Scorecard for Measuring and Managing Business Success.** Great Britain, 2002. ISBN 0-273-65334-2.

RUBINSTEIN, R. Y. **Simulation and the Monte Carlos Method.** New Jersey, 2006 (Wiley Series in Probability and Statistics)

SHANNON, R. E. **Introduction to the Art and Science of Simulation.** In Proceedings of the 1998 Winter Simulation Conference, Association for Computing Machinery, New York, 1998.

SILVA, E. L.; MENEZES, E. M. **Metodologia da Pesquisa e Elaboração da Dissertação.** 4 ed. Florianópolis: UFSC, 2005.

YAN, X. **Linear regression analysis: theory and computing.** Singapore, 2009. ISBN 981-283-410-9.