

Desenvolvimento de um sistema de apoio à decisão na orçamentação de produtos por encomenda

Lucas Henrique Matte

Dissertação apresentada à

Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança para
obtenção do **Grau de Mestre em Engenharia Industrial**

no âmbito da **Dupla Diplomação** com a **Universidade Tecnológica Federal do
Paraná – Campus Pato Branco**

Este trabalho foi efetuado sob orientação de:

Professora Doutora Maria Clara Rodrigues Bento Vaz Fernandes

Coorientação de:

Professor Doutor Dalmarino Setti

Esta dissertação não inclui as críticas e sugestões feitas pelo Júri

Maio de 2022

Agradecimentos

Este trabalho é resultado da parceria, colaboração e incentivo de muitas pessoas. Agradeço primeiramente à Universidade Tecnológica Federal do Paraná (UTFPR) campus Pato Branco pela grande oportunidade proporcionada, assim como ao Instituto Politécnico Bragança, por ter me recebido e acolhido nesse período de um ano.

À minha orientadora, professora doutora Maria Clara Rodrigues Bento Vaz Fernandes por me auxiliar em todo o desenvolvimento do trabalho, por sempre estar a disposição e me apoiar quando necessário.

À empresa FactoryPlay pela recepção e pela oportunidade de realizar um estágio e uma dissertação aplicada ao ambiente de trabalho.

Aos meus companheiros Lize, Matheus, Pedro e Lucas com quem tive o prazer em dividir a casa. Por terem me transmitido uma sensação de família nesse período afastado, por terem me acolhido e me ajudado nos momentos que precisei.

Ao meu amigo Fernando, com quem tive a oportunidade não só de dividir a casa, mas também dividir quase toda a vida. Agradeço por estar sempre presente, perto ou longe, por sempre ter as palavras certas para dizer, em quaisquer momentos.

Aos meus colegas e amigos de toda a vida e aos que participaram desse período em Portugal e deixaram essa fase marcada em minha vida. Especialmente para os amigos Amanda, André, Dalila, Elias, Gabriel, Jaqueline, Maria Eduarda, Mariah e Renann que me auxiliaram desde o começo e fizeram com que Bragança fosse o meu lar, levarei vocês para toda a vida.

À minha namorada Isadora, com amor, pelo incentivo e companheirismo em todos os momentos, sendo eles tranquilos ou turbulentos. Por compartilhar momentos

inesquecíveis, por estar presente em minhas realizações e por sempre me apoiar e me ajudar a me tornar uma pessoa melhor.

A toda a minha família que me apoia em evoluir e crescer cada vez mais. Aos meus avós, pais, irmã e tios, vocês são a minha saudade diária, e estão sempre presentes no meu coração.

A minha mãe, Liane, ao meu pai, Gilmar e a minha irmã, Heloísa, toda a minha gratidão e amor, nada disso seria possível sem vocês, palavras não são suficientes para expressar o quão grato eu sou por ter vocês em minha vida. Vocês são o meu alicerce, a razão do meu viver, maior orgulho, maior força e minha maior motivação para sempre seguir em frente. Eu amo vocês infinitamente.

Por fim, agradeço a todos que contribuíram de alguma forma para toda a minha jornada acadêmica e para a realização deste trabalho.

Resumo

Este trabalho tem por objetivo identificar os custos de produção críticos, relativos à matéria-prima e mão de obra, de produtos por encomenda sem padronização, e assim desenvolver um modelo quantitativo para realizar previsões precisas destes custos na etapa inicial do projeto de uma empresa, no âmbito do processo de orçamento. Para alcançar o objetivo proposto, buscou-se, em primeiro lugar, o entendimento dos processos produtivos e matérias-primas da empresa, assim como o estudo dos principais aspectos teóricos relacionados a técnicas e métodos para estimativa de custos, criação de modelos para estimativa de custos, seleção de modelos e validação destes. Visto isso, pretende-se desenvolver um modelo de regressão linear múltipla (RLM), aplicada a um histórico de dados quantitativos da empresa, utilizando o software *RStudio*, a fim de estimar cada variável crítica relativa à quantidade de matéria-prima principal e aos tempos de mão de obra para processos críticos. Seis modelos foram analisados, dois para cada variável crítica definida, sendo estas o valor dos metros lineares (VML) de matéria-prima principal utilizado no produto, o tempo de corte da matéria-prima principal envolvido no produto e o tempo de costura ligado ao produto. Os modelos foram avaliados, selecionados e validados, definindo um melhor modelo para cada variável crítica. Os parâmetros dos modelos foram obtidos utilizando dados de treino e posteriormente os resultados dos modelos selecionados foram validados utilizando dados de teste. Os resultados obtidos, mediante a utilização da metodologia proposta, foram avaliados pela empresa e se mostraram confiáveis para utilização na fase inicial de desenvolvimento do produto. Após estes passos, foi possível utilizar estes modelos para criar uma ferramenta simples e útil para a empresa.

Palavras-chave: Orçamentação de produtos, Custos de produção, Estimativa de custos, Regressão linear múltipla.

Abstract

This work aims to identify the critical production costs, related to raw materials and labor, of ordered products without standardization in order to develop a quantitative model to predict accurately these costs in the early project stage of a company, within the budget step. In order to achieve this goal, it was necessary to understand the production processes and the raw materials of the company, as well as to study the principal theoretical aspects related to cost estimating techniques and methods, cost estimating models, model selection and validation. Therefore, it is intended to develop a multiple linear regression model, applied to an historical quantitative data of the company, using the *R* software to estimate each critical variable concerning the quantity of the main raw material and the labor times to critical processes. Six models were analyzed, in which two models are identified for each critical variable such as the linear meters value of the main raw material used in the product, the main raw-material cut time involved in the product and the sew time required by the product. The models were evaluated, selected, and validated, defining the best model for each critical variable. The model parameters were obtained using a train dataset and, afterwards, the results of the selected models were validated using a test dataset. The obtained results, through the proposed methodology, were evaluated by the company and they show to be reliable for use in the early stage of product development within the budget step. After these steps, it was possible to use these models to create a simple and useful tool to the company.

Keywords: Product budgeting, Production costs, Cost estimation, Multiple linear regression.

Índice

1.	Introdução.....	1
1.1	Objetivos.....	2
1.1.1	Objetivo Geral.....	2
1.1.2	Objetivos Específicos.....	2
2.	Caracterização da empresa.....	4
2.1	Produtos utilizados para estudo.....	9
3.	Revisão da literatura.....	12
3.1	Estimativa de Custos.....	12
3.2	Características geradoras de custo nos produtos.....	13
3.3	Método Activity Based Costing (ABC).....	16
3.4	Métodos de orçamentação.....	16
3.4.1	Método qualitativo.....	16
3.4.2	Método quantitativo.....	17
3.5	Regressão linear simples (RLS).....	18
3.6	Regressão linear múltipla (RLM).....	20
3.8	Métodos de seleção de variável nos modelos de regressão.....	21
3.8.1	Akaike Information Criteria (AIC).....	22
3.9	Validação dos modelos de regressão.....	23
3.10	Modelos de regressão para custos.....	24
4.	Metodologia.....	26
4.1	Modelo e <i>Software</i> utilizados.....	26
4.2	Desenvolvimento da base de dados.....	26
4.3	Cálculo dos custos de produção.....	29
4.3.1	Matéria-prima.....	30
4.3.2	Mão de obra.....	31
5.	Resultados e Discussão.....	37
5.1	Desenvolvimento dos modelos.....	37
5.1.2	Obtenção dos modelos de RLM.....	40
5.2	Avaliação dos pressupostos relativos aos resíduos dos modelos RLM.....	45
5.3	Análise dos resultados dos modelos de RLM.....	49

5.3.1 Modelo para estimar o VML	49
5.3.2 Modelo para estimar o tempo de corte	51
5.3.2 Modelo para estimar o tempo de costura.....	52
5.5 Modelos finais.....	53
5.6 Validação dos modelos	55
5.7 Cálculo dos custos	59
5.8 Utilização dos modelos pela empresa	59
6. Conclusão	62
7. Referências bibliográficas	64

Lista de acrónimos

ABC – *Activity based costing*

AIC – *Akaike Information criteria*

BE – *Backward elimination*

EGF – Encargos gerais de fabrico

FS – *Forward selection*

MAE – *Mean absolute error*

MAPE – *Mean absolute percentage error*

MP – Matéria-prima

MSE – *Mean squared error*

PVP – Preço de venda ao público

RLS – Regressão linear simples

RLM – Regressão linear múltipla

RMSE – *Root mean square error*

VIF – *Variance Inflation Factor*

VML – Valor dos metros lineares

Lista de tabelas

Tabela 4.1 - Cálculo da massa do produto.....	28
Tabela 4.2 - Cálculo dos custos de matéria-prima para um produto da família 1	30
Tabela 4.3 - Cálculo dos custos de matéria-prima para um produto da família 2	31
Tabela 4.4 - Cálculo dos custos de matéria-prima para um produto da família 3	31
Tabela 4.5 - Cálculo dos custos de mão de obra para um produto da família 1	32
Tabela 4.6 - Cálculo dos custos de mão de obra para um produto da família 2	32
Tabela 4.7 - Cálculo dos custos de mão de obra para um produto da família 3	33
Tabela 5.1 - Valores do VIF para os modelos (5.1) e (5.2)	42
Tabela 5.2 - Valores do VIF para o modelo "stepmp2" após a remoção da variável "Área"	42
Tabela 5.3 - Resíduos padronizados para os modelos (5.1) e (5.3)	48
Tabela 5.4 - Resultados para o modelo (5.1)	50
Tabela 5.5 - Resultados para o modelo (5.4)	51
Tabela 5.6 - Resultados para o modelo (5.6)	52
Tabela 5.7 - Resultados para o modelo (5.7)	52
Tabela 5.8 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 1	54
Tabela 5.9 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 2	55
Tabela 5.10 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 3	55
Tabela 5.11 - Comparação entre os valores estimados e os reais para o Modelo 1.....	56
Tabela 5.12 - Comparação entre os valores estimados e os reais para o Modelo 2.....	57
Tabela 5.13 - Comparação entre os valores estimados e os reais para o Modelo 3.....	57
Tabela 5.14 - Cálculo dos custos críticos, estimados e reais, para um produto da família 1	59

Lista de figuras

Figura 2.1 - Fluxograma dos processos produtivos.....	4
Figura 2.2 - Desenho 3D realizado o software TopSolid Design Concept Furniture.....	5
Figura 2.3 - Planificação 2D realizada no software Rhinoceros	5
Figura 2.4 - Plano de corte realizado no software Diamino	6
Figura 2.5 - Estoque de matéria-prima	6
Figura 2.6 - Máquina de corte	6
Figura 2.7 - Setor de pintura e impressão.....	7
Figura 2.8 - Máquina de costura.....	8
Figura 2.9 - Máquina de soldadura por temperatura	8
Figura 2.10 - Artigos dobrados preparados para o estoque	8
Figura 2.11 - Área de estoque de produto acabado	8
Figura 2.12 – Bandeiras publicitárias	9
Figura 2.13 - Tendas publicitárias	9
Figura 2.14 – Totem publicitário.....	9
Figura 2.15 – Meta publicitária	9
Figura 2.16 - Artigos decorativos.....	10
Figura 2.17 - Lonas agrícolas para colheita de azeitonas.....	10
Figura 2.18 – Fluxograma da definição de componentes de custo nos produtos.....	11
Figura 3.1 - Influência da geometria e do processo produtivo na escolha da matéria-prima.....	15
Figura 3.2 - Influência da matéria-prima e da geometria na escolha do processo produtivo.....	15
Figura 3.3 - Influência da matéria-prima e do processo produtivo na escolha da geometria	15
Figura 3.4 - Exemplo de RLS.....	19
Figura 4.1 - Área de matéria-prima utilizada no produto.....	28
Figura 4.2 - Volume estimado do produto.....	28
Figura 4.3 - Perímetro total das peças do produto.....	29
Figura 4.4 - Altura média do produto.....	29
Figura 4.5 – Fragmento da base de dados para estimar o VML.....	35
Figura 4.6 – Fragmento da base de dados para estimar o tempo de corte.....	35
Figura 4.7 - Fragmento da base de dados para estimar o tempo de costura.....	35
Figura 5.1 - Diagrama da relação entre as variáveis da base de dados de treino da matéria-prima.....	38

Figura 5.2 - Diagrama da relação entre as variáveis da base de dados de treino do tempo de corte	39
Figura 5.3 – Diagrama da relação entre as variáveis da base de dados de treino do tempo de costura.....	39
Figura 5.4 - Modelos de RLM na linguagem R	40
Figura 5.5 - Método backward selection com a função stepAIC	41
Figura 5.6 - Gráfico dos resíduos em função dos valores previstos para o modelo (5.1)	45
Figura 5.7 - Gráfico dos resíduos em função dos valores previstos para o modelo (5.3)	46
Figura 5.8 - Gráfico Q-Q para o modelo (5.1)	47
Figura 5.9 - Gráfico Q-Q para o modelo (5.3)	47
Figura 5.10 - Diagrama para estimar o VML através do modelo 1.....	60
Figura 5.11 - Diagrama para estimar o tempo de corte através do modelo 2.....	60
Figura 5.12 - Diagrama para estimar o tempo de costura através do modelo 3	61

1. Introdução

O presente trabalho resultou de um estágio numa empresa que desenvolve e produz produtos por encomenda e da dificuldade sinalizada pela empresa na orçamentação de produtos novos.

Com a globalização, surge um mercado cada vez mais competitivo em que as empresas devem buscar novas formas de produzir (com menores custos e maior produtividade) e, principalmente, gerenciar seus negócios (Batista et al., 2005). Quando se fala da inclusão de um novo produto no mercado é imprescindível ter o conhecimento de todos os custos envolvidos desde a fase inicial de projeto até a entrega ao público, a fim de tomar decisões estratégicas. A primeira etapa requisitada pelo cliente, após uma ideia ou proposta apresentada, é justamente o orçamento deste artigo. É essencial que este orçamento seja realizado de forma ágil e mais precisa possível, é neste âmbito que se enquadra o problema envolvido no trabalho. Após a fabricação deste produto, o seu custo final deve ser aproximadamente igual ao valor do orçamento realizado na etapa inicial do projeto, visto que já foi apresentado ao cliente, sendo que o desvio deve ser mínimo. No caso de o custo final do produto ser superior ao anteriormente orçamentado, implica a diminuição da margem de venda, gerando possíveis prejuízos, perda de clientes por ocorrerem alterações nos valores pressupostos e menos competitividade.

Visualizando este problema, existem vários desafios para se aproximar de uma estimativa de custos precisa na etapa embrionária do projeto, precisando definir todos os seus componentes. Será necessário definir métodos para estimar estes componentes, através de uma recolha de dados intensiva, além de validá-los e transformá-los em uma ferramenta útil, simples e rápida para a empresa e o utilizador final.

No processo de fabricação da empresa do caso em estudo, são utilizadas como matéria-prima principal lonas, redes e tecidos para a produção dos produtos, alguns deles são insufláveis, produtos publicitários, produtos decorativos e lonas agrícolas. Essa matéria-prima é recebida em rolos, os quais possuem uma largura e um comprimento, que por sua vez pode ser medido em um valor de metros lineares (VML). Assim, cada produto utiliza um certo VML de matéria-prima, ou seja, um certo comprimento do rolo. Geralmente, a empresa estima o custo final do produto e o preço de venda ao público (PVP) baseando-

se apenas no VML utilizado. No caso de produtos totalmente novos e diferentes dos existentes já produzidos, há uma dificuldade para estimar o VML e os tempos de mão de obra envolvida.

Com base na análise atual de orçamentação foram desenvolvidos modelos, utilizando a regressão linear múltipla, aplicada a um histórico de dados da empresa, o que permitiu estimar a quantidade de matéria-prima principal e os tempos de mão de obra mais influentes na produção dos produtos, sendo considerados críticos no processo de orçamentação. Em seguida estão descritos os objetivos do atual trabalho.

1.1 Objetivos

1.1.1 Objetivo Geral

Neste sentido, tem-se como objetivo principal desenvolver um método para determinar a quantidade de matéria-prima utilizada e os tempos de mão de obra envolvida em produtos novos, na etapa inicial de projeto. Desse modo será possível calcular os componentes de custo mais críticos e influentes envolvidos no produto, multiplicando-se a quantidade de cada componente pelo seu preço unitário.

1.1.2 Objetivos Específicos

- Pesquisar diferentes estratégias para estimar as componentes de custo de produtos nas etapas iniciais de desenvolvimento;
- Entender e visualizar os processos produtivos da empresa;
- Entender como são feitos os cálculos de todas as componentes de custo dos produtos, em relação a matéria-prima e mão de obra;
- Criar uma base de dados com o histórico de dados da empresa, a fim de entender o comportamento dos custos;
- Desenvolver uma ferramenta para apoio à orçamentação com base no histórico de dados;
- Validar a ferramenta em questão utilizando produtos que não foram utilizados para criá-la;
- Utilização da ferramenta pela empresa para orçamentação de produtos novos.

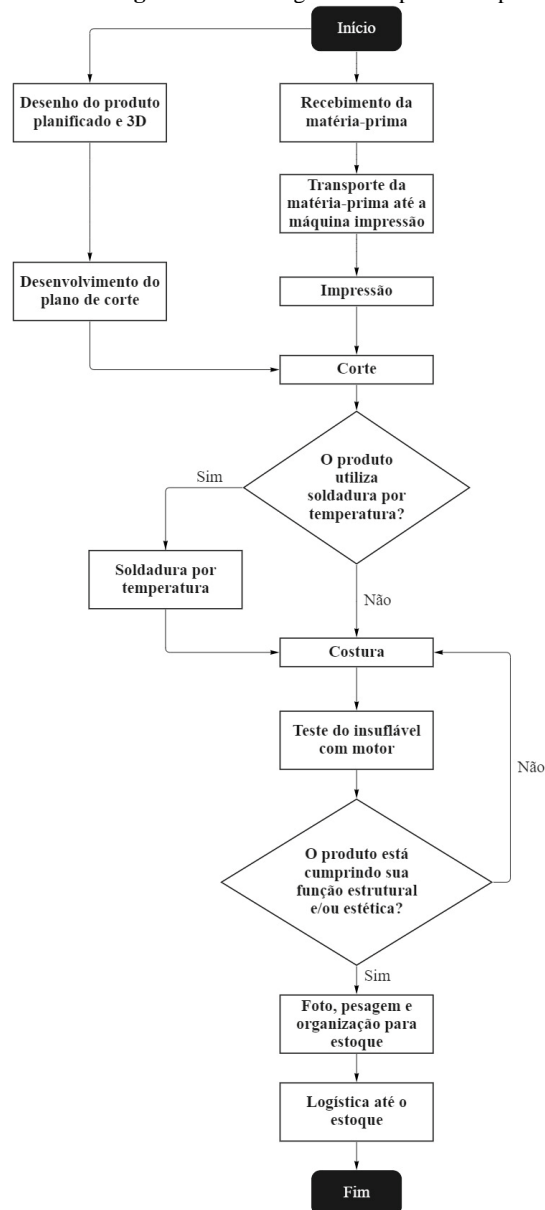
A dissertação está estruturada em seis capítulos, incluindo a introdução como primeiro capítulo. No que se refere ao segundo capítulo, será feita uma breve caracterização da empresa, a fim de visualizar o processo produtivo, os produtos produzidos e utilizados para o estudo, assim como as componentes essenciais de custo envolvidas. No terceiro capítulo será descrita a revisão da literatura que será abordada na prática, com foco em métodos para a estimativa de custos, principalmente nos métodos de regressão linear, considerando estudos de utilização desse método para custos, os métodos de seleção de variáveis para o modelo de regressão e por fim a validação desses modelos. A metodologia está presente no quarto capítulo, nela estão descritos os métodos utilizados que contribuíram para a concretização dos objetivos do trabalho. No quinto capítulo estão presentes os resultados e discussões sobre os modelos construídos, nele são descritos o desenvolvimento dos modelos, os resultados para cada modelo, a avaliação dos pressupostos da regressão linear múltipla, assim como a validação e seleção do modelo final para utilização no cálculo dos custos críticos dos produtos. O sexto capítulo compreende algumas conclusões acerca dos resultados obtidos nos modelos, evidenciando também alguns trabalhos futuros que poderão enriquecer o presente trabalho.

2. Caracterização da empresa

Neste capítulo é feita uma caracterização da empresa, a descrição de todo o processo produtivo, demonstração dos produtos produzidos e utilizados no estudo como também a exposição de todas os componentes de custo envolvidos nos produtos.

Na **Figura 2.1** é apresentado um fluxograma esquemático das etapas do processo produtivo da empresa, utilizadas em todas as famílias de produtos abordadas no estudo, que são descritas em seguida.

Figura 2.1 - Fluxograma dos processos produtivos



Fonte: o próprio autor.

O processo inicia-se no setor do *design*, no qual é feito o desenho do produto, inicialmente é realizado um esboço à mão, para melhor compreensão do produto e da sua forma, sem levar em consideração dados técnicos. Após esse passo, o desenho é realizado no software *TopSolid Design Concept Furniture* em três dimensões (3D) (**Figura 2.2**), com as dimensões aproximadas às reais para uma visualização do produto final. A seguir produto é planificado em duas dimensões (2D) através do software *Rhinoceros* (**Figura 2.3**) para identificar todos os componentes separados. O plano de corte é definido após a planificação, sendo que este arquivo 2D é transferido para o programa *Diamino* (**Figura 2.4**), onde é possível visualizar no campo preto superior todas as componentes planificadas, e no campo preto inferior a organização do corte dessas componentes no rolo de matéria-prima, definindo assim o plano de corte, já considerando uma tolerância de 15 milímetros para o futuro processo de costura. As ilustrações em 2D, desenhos personalizados dos produtos são feitas juntamente nessa etapa, no software *Illustrator*.

Figura 2.2 - Desenho 3D realizado o software TopSolid Design Concept Furniture

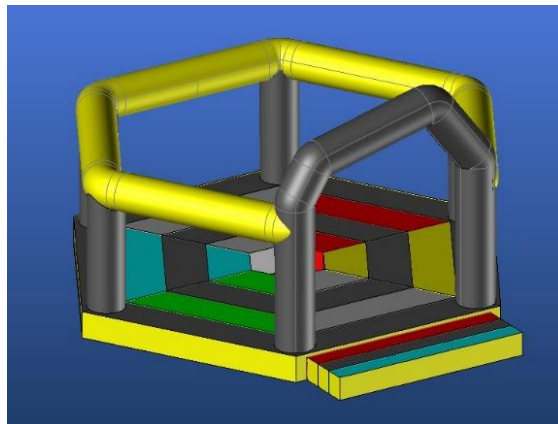
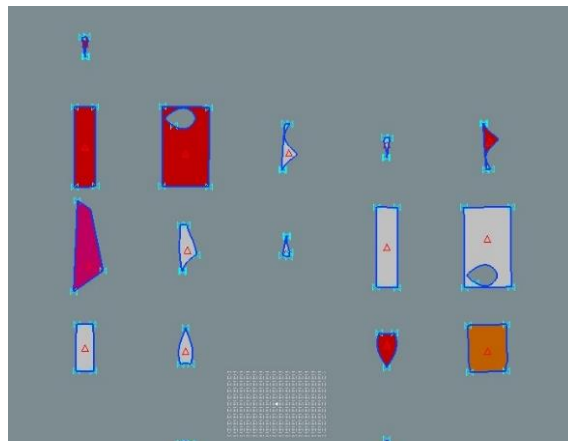
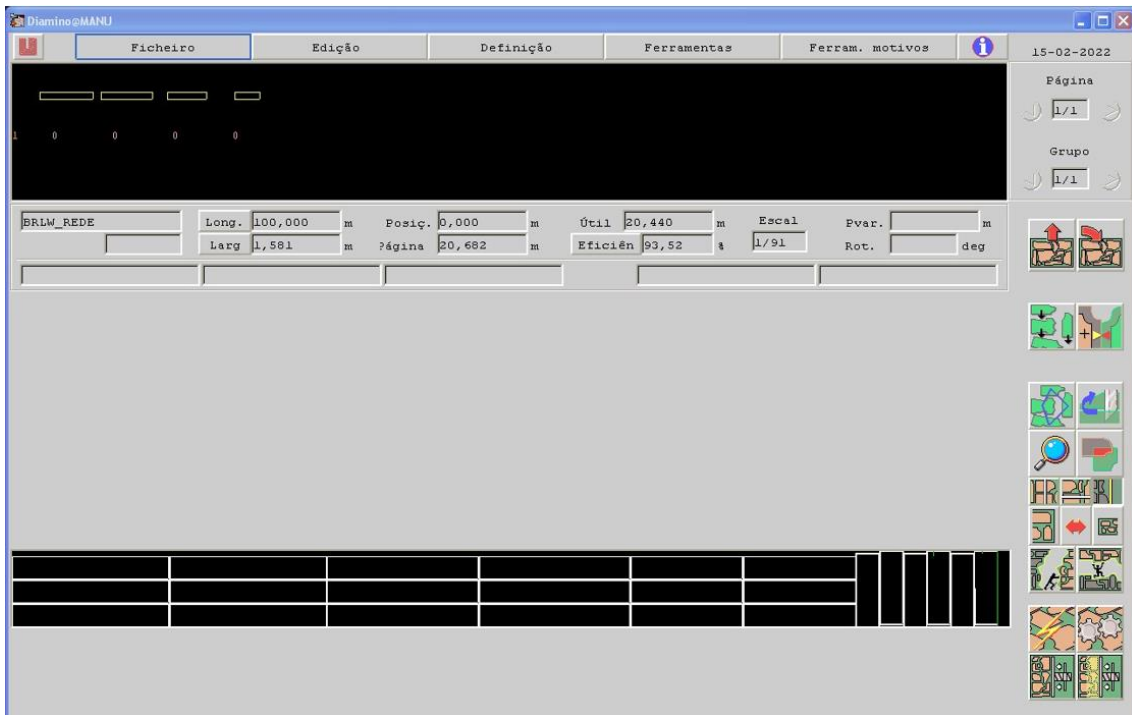


Figura 2.3 - Planificação 2D realizada no software Rhinoceros



Fonte: o próprio autor.

Figura 2.4 - Plano de corte realizado no software Diamino



Fonte: o próprio autor.

Passando para a parte de produção, primeiramente a matéria-prima é recebida diretamente no galpão do almoxarifado (**Figura 2.5**), neste local são armazenados todos os tipos da matéria prima principal, que são as lonas, tecidos e redes. Em anexo a este local está a máquina de corte (**Figura 2.6**). O primeiro processo realizado na matéria-prima é a impressão, o rolo de lona/tecido é transportado e posicionado na impressora (**Figura 2.7**), após isso, o arquivo em 2D com as ilustrações e as linhas para futuro corte manual é enviado para a máquina e executado pelo operador.

Figura 2.5 - Estoque de matéria-prima



Figura 2.6 - Máquina de corte



Fonte: o próprio autor.

Figura 2.7 - Setor de pintura e impressão

Fonte: o próprio autor.

O corte da matéria prima sempre é manual caso a peça tenha impressões e/ou pinturas personalizadas. As peças que não possuem ilustrações são encaminhadas para a máquina de corte (**Figura 2.6**) a qual utilizando o plano de corte realizado pelo setor do *design* realiza o processo de corte da matéria-prima, com as tolerâncias para costura já definidas.

Caso o produto utilize a soldadura por temperatura a matéria-prima é passada para este equipamento (**Figura 2.9**) e após para o setor da costura, caso contrário, diretamente para a costura. É no setor da costura (**Figura 2.8**) que o produto começa a tomar a sua forma final, nele é feita a junção dos componentes através das máquinas de costura, utilizando costura duplas. Finalizado o processo de costura, são feitos testes no produto, um dos testes, para os insufláveis, é a conexão do motor na entrada de ar do produto a fim de analisar sua estrutura e estética e realizar ajustes se necessário. Neste caso, o produto retorna à mesa de costura e são feitos os reparos.

Figura 2.8 - Máquina de costura



Figura 2.9 - Máquina de soldadura por temperatura



Fonte: o próprio autor.

Depois de passar por todas estas etapas, o artigo é fotografado, pesado, dobrado (**Figura 2.10**), embalado e levado até à área de estoque de produto acabado (**Figura 2.11**), finalizando assim o processo produtivo.

Figura 2.10 - Artigos dobrados preparados para o estoque



Figura 2.11 - Área de estoque de produto acabado



Fonte: o próprio autor.

2.1 Produtos utilizados para estudo

Os produtos produzidos pela empresa são divididos em diversas famílias, por apresentarem similaridades geométricas, processos produtivos similares e por terem funções/intuítos parecidos. Para o estudo, foram considerados produtos de três famílias distintas, com processos e matérias-primas parecidas, a fim de ter mais dados para análise.

Na família 1 (F1) estão presentes os produtos relacionados com a publicidade, sendo incluídos: bandeiras (**Figura 2.12**) tendas (**Figura 2.13**), totens (**Figura 2.14**), metas (**Figura 2.15**), entre outros produtos com objetivo de publicitar marcas.

Figura 2.12 – Bandeiras publicitárias



Fonte: (factoryplay.com/active-brands/pt/product/bandeiras-vrios-formatos-e-alturas-pt/)

Figura 2.13 - Tendas publicitárias



Fonte: (factoryplay.com/active-brands/pt/product/tenda-de-aluminio-3x3-reconco/)

Figura 2.14 – Totem publicitá-



Fonte: (factoryplay.com/active-brands/pt/product/totem-stihl-pt/)

Figura 2.15 – Meta publicitária



Fonte: (factoryplay.com/active-brands/pt/product/arco-de-meta-sheraton-9m-x-55m/)

A família 2 (F2) é composta pelos produtos decorativos, produtos para épocas especiais, como Natal, Páscoa, Ano Novo, e eventos em geral (**Figura 2.16**).

Figura 2.16 - Artigos decorativos



Fonte: (factoryplay.com/active-brands/pt/product/kit-decorao-natal-16-elementos-pt/)

A família 3 (F3) é composta por lonas agrícolas utilizadas por exemplo para a colheita de azeitona (**Figura 2.17**).

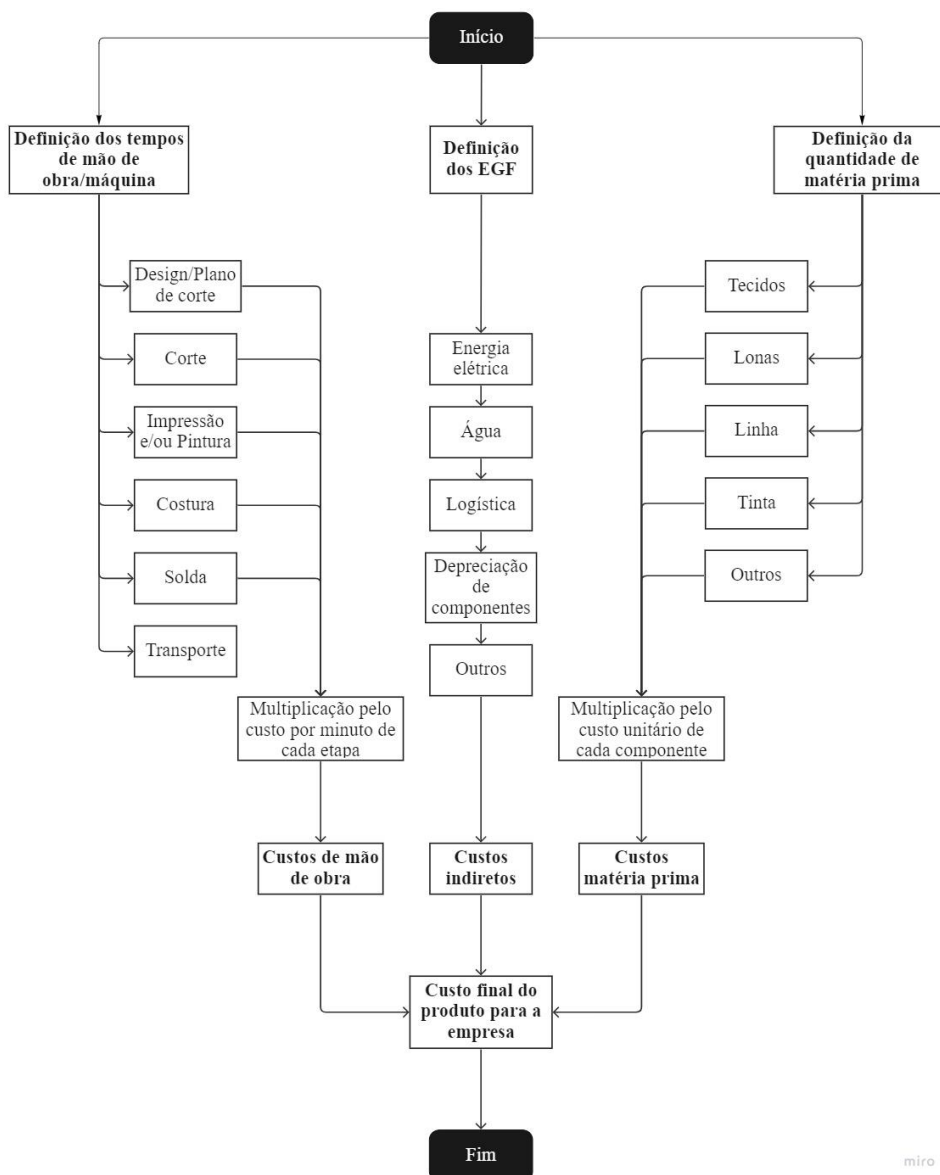
Figura 2.17 - Lonas agrícolas para colheita de azeitonas



Fonte: o próprio autor.

Conhecendo quais foram os produtos utilizados no estudo é necessário visualizar todos os elementos que geram custo nesses produtos. No fluxograma da **Figura 2.18** podemos ver as componentes de custo essenciais para a empresa relativas aos produtos manufaturados pela fábrica, os quais são custos de mão de obra/máquina, os encargos gerais de fabrico (EGF) e os custos de matéria prima, a fim de definir o custo real do produto.

Figura 2.18 – Fluxograma da definição de componentes de custo nos produtos



Fonte: o próprio autor.

3. Revisão da literatura

Este capítulo tem por objetivo apresentar alguns estudos que referem conceitos e métodos utilizados como pilares na resolução do problema descrito, ajudando a estimar os custos envolvidos na fabricação dos produtos na fase da orçamentação. Alguns tópicos como estimativa de custos, características geradoras de custo nos produtos e alguns métodos para a determinação de custos nos produtos serão apresentados nas seguintes seções.

3.1 Estimativa de Custos

Segundo a *Association for the Advancement of Cost Engineering* (CET 2010), a estimativa de custos foca-se em determinar e prever os custos de uma construção, manufatura de produtos, ou uma prestação de serviços. Na determinação do custo de produtos é geralmente envolvido o cálculo do custo de produção e do preço de venda de um tal bem, incluindo os custos de pesquisa e desenvolvimento, projeto, fabricação, *marketing*, logística e serviços ao cliente (Park et al., 2002). Complementando, a estimativa de custos está interessada em encontrar os custos relacionados com um conjunto de atividades antes que realmente sejam executadas (Weustink et al., 2000).

Além disso, o potencial consumidor final de um suposto produto frequentemente exige uma cotação de preços o mais breve possível, algumas vezes mesmo despreocupado com os fatores como a dimensão da customização, com a natureza dos dados requeridos e com a complexidade do *design*. Em algumas situações, muitas vezes o fabricante ignora fatores significativos, como disponibilidade do setor de *design*, capacidade de fabricação e o nível de precisão requerida na estimativa do tempo de produção. Esta situação pode desencadear uma subestimação, resultando em uma perda de lucro ou um dano estratégico causado pela superestimação levando à perda de credibilidade com o cliente e participação no mercado. Desse modo, a orçamentação mais precisa torna-se um ponto de grande foco para estratégias operacionais e um ponto fundamental para as decisões tomadas pela organização para produção de novos produtos nas etapas iniciais de desenvolvimento. Novas técnicas e métodos para produzir estimativas de custo precisas e consistentes não geram apenas soluções de *design* ideais, em que é possível comparar os custos de diferentes modelos de *design* a fim de obter um menor custo de produção e conseqüente maior

lucro, mas também alcançam uma alta satisfação dos clientes em termos de baixo custo, alta qualidade e entregas no tempo estipulado (Niazi et al., 2006).

3.2 Características geradoras de custo nos produtos

Segundo a análise de D.G. Ullman (2008) do processo de desenvolvimento de produto e sua tomada de decisões, pode ser concluído que os custos durante o desenvolvimento são causados pela geometria, pela matéria-prima (MP) utilizada, pelos processos produtivos e pelo planejamento da produção.

A geometria determina a quantidade de material e os processos produtivos que serão necessários. A geometria inclui a forma, dimensões e tolerâncias. A influência desse fator nos custos do produto é de certa forma clara, um maior nível de exatidão requer recursos mais precisos, como por exemplo máquinas, ferramentas e equipamentos, maiores tempos de produção pela complexidade e precisão, resultando num aumento nos custos de produção. No presente caso em estudo, a forma influencia diretamente no contorno e aspecto das costuras necessárias no produto, induzindo a um aumento direto nos gastos com a mão de obra devido ao aumento do tempo de produção e na quantidade de MP devido às dimensões determinadas.

Os custos com as MP correspondem a todos os materiais consumidos e envolvidos durante a fabricação de um tal bem, que sofrerão operações de transformação e constituirão o produto acabado (Miguel & Pires, 2017).

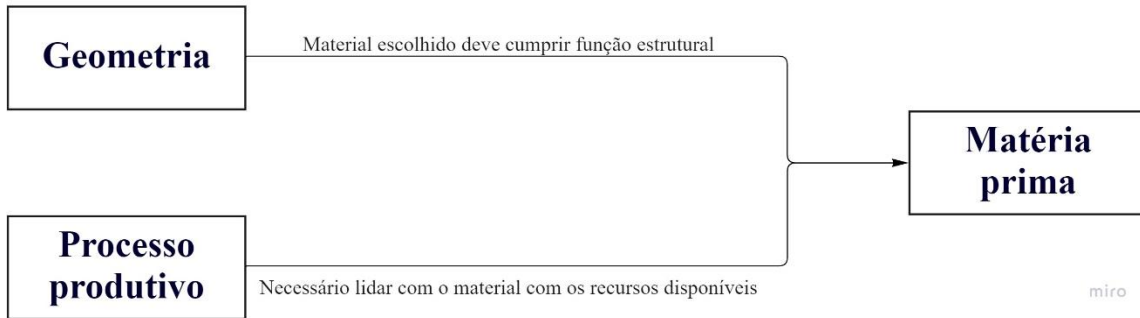
Processos produtivos são requeridos para transformar matérias primas em um componente ou incluem a montagem de componentes. Recursos (*e.g.* operadores, ferramentas, máquinas) são necessários para desenvolver essa atividade produtiva. Os recursos têm certas restrições técnicas, relativas à potência da máquina, precisão, ritmo de trabalho, dimensões, limitando a execução de certas operações que exigem uma capacidade além destas restrições (Mileski, 2011). Para cada processo produtivo, existe um tempo de mão de obra ou máquina envolvido, esse tempo gera um custo de produção, que está envolvido no custo do produto. A costura é um dos processos produtivos presentes na manufatura dos produtos para o caso estudado. Logo, o tempo de costura é uma das variáveis que geram custo no produto, sendo que cada produto possui dimensões diferentes, variando este tempo, gerando valor ao produto final.

No planejamento da produção são realizadas as decisões sobre a utilização dos recursos produtivos a fim de satisfazer a previsão da procura ao longo de um certo horizonte a um custo razoável (Katholieke, 1981). Além das restrições técnicas dos recursos, também existem restrições logísticas destes, o que significa que um recurso nem sempre estará disponível. Com o conhecimento destas situações, as operações podem ser alocadas de forma estratégica (Weustink et al., 2000) de forma a minimizar perdas relacionadas com transportes internos, resultando em uma otimização no processo, reduzindo custos desnecessários. Como outro exemplo, a fabricação de um produto com um processo produtivo planejado de forma errônea pode acarretar o aumento dos tempos de produção, relacionados diretamente com os custos de mão de obra envolvidos, tornando desse modo o produto mais caro, em comparação com um processo bem-organizado e desenvolvido.

As quatro características (geometria, matérias-primas, processos produtivos e planejamento da produção) supracitadas são interrelacionadas. A matéria prima é influenciada pela geometria tendo em vista que o produto tem que cumprir sua função estrutural caso necessário, e pelo tipo de processo produtivo, a fim de ser possível transformar o material, com os recursos e equipamentos disponíveis na empresa, tal como se ilustra na **Figura 3.1**. O processo produtivo é influenciado pela geometria (*e.g.* forma, dimensões), sabendo que a dimensão do produto e forma afetam diretamente nos tempos de produção e podem dificultar a manufatura, e pelo tipo de matéria-prima, pois a logística e o manejo deste é essencial, tal como se ilustra na **Figura 3.2**. Além disso a geometria também é influenciada pela MP, pois as propriedades da MP devem permitir a obtenção dessa geometria. A geometria também é influenciada pelos processos produtivos, pois não é possível fabricar um produto com uma geometria que ultrapasse os limites e capacidades técnicas do equipamento e dos operadores, tal como visualizado na **Figura 3.3**. A seleção de uma das características sempre gera restrições nas outras. A escolha do processo depende do ambiente de trabalho. Em uma produção em massa por exemplo, a ênfase é geralmente a combinação apropriada do processo e da matéria-prima, sendo que nesse meio há mais tempo disponível para fazer uma boa seleção. Em uma fabricação em pequenos lotes, o material muitas vezes é selecionado primeiramente, devido às máquinas serem de uso

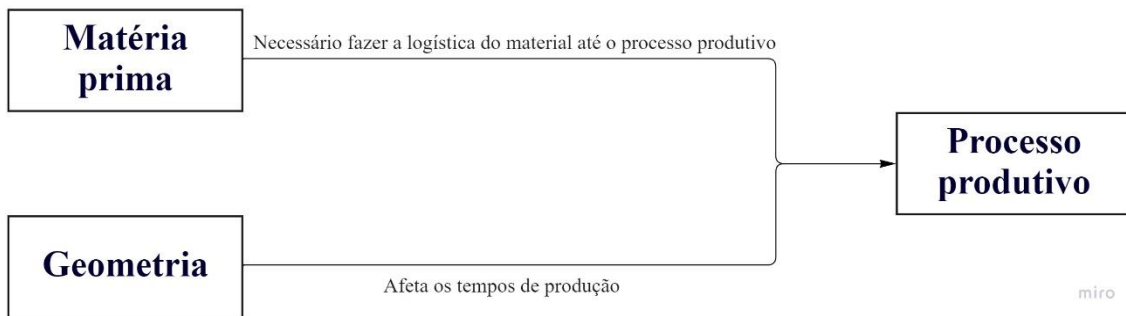
geral, que restringem o número de processos produtivos disponíveis (Weustink et al., 2000).

Figura 3.1 - Influência da geometria e do processo produtivo na escolha da matéria-prima



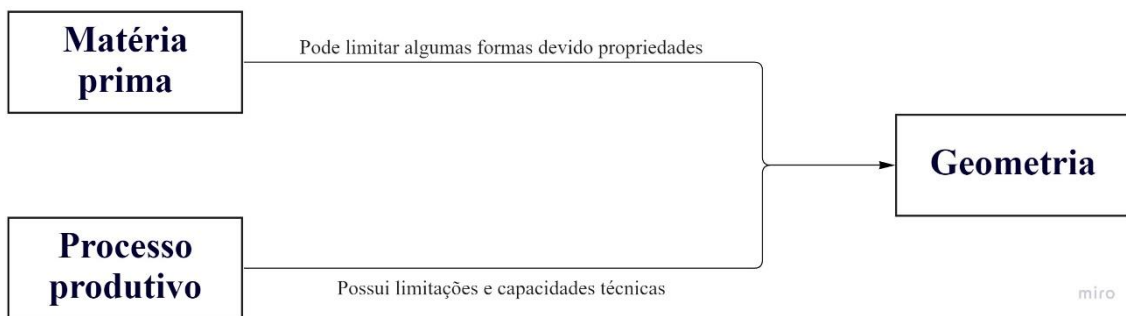
Fonte: o próprio autor.

Figura 3.2 - Influência da matéria-prima e da geometria na escolha do processo produtivo



Fonte: o próprio autor.

Figura 3.3 - Influência da matéria-prima e do processo produtivo na escolha da geometria



Fonte: o próprio autor.

3.3 Método Activity Based Costing (ABC)

Cooper & Kaplan (1991) desenvolveram o sistema ABC para obter um custo mais preciso relativo a produtos que resultam de processos produtivos. Neste sistema ABC, a atribuição dos custos é baseada nas atividades requeridas para obter o produto, focando-se na medição do custo e da performance das atividades, sendo baseado em três premissas básicas: produtos requerem atividades, atividades consomem recursos e recursos custam dinheiro (Cooper & Kaplan, 1988). Segundo Ching (1997), o ABC é um método para rastreamento das despesas de uma empresa ou departamento nas atividades realizadas e para examinar como essas atividades estão relacionadas à geração de receita e consumo de recursos. O ABC avalia o valor que cada atividade agrega para o desempenho da empresa. O sistema ABC possibilita uma visão profunda do próprio negócio ao dividir em partes os processos. Desse modo identifica o envolvimento entre as diferentes estruturas da organização e a composição das atividades executadas e conseqüentemente o consumo dos recursos para produzir produtos ou serviços, a fim de satisfazer a necessidade dos clientes (Batista et al., 2005). Sistemas ABC provaram ser mais precisos que os sistemas tradicionais para determinação de custos, ajudando os gestores a entender precisamente onde tomar ações para gerar maiores lucros (Morgan, M. J., & Bork, 1993). De seguida, descrevem-se alguns métodos de orçamentação encontrados na literatura.

3.4 Métodos de orçamentação

Niazi et al. (2006) categoriza as técnicas de *Product Cost Estimation* (PCE) em qualitativa e quantitativa. A classificação é baseada em um agrupamento de técnicas com recursos similares e com tipos de dados similares.

3.4.1 Método qualitativo

As técnicas de orçamentação qualitativas são baseadas em uma análise comparativa de um novo produto com produtos que já foram fabricados com finalidade de identificar similaridades. Estas ajudam a incorporar os dados antigos no novo design, reduzindo consideravelmente a necessidade de obter o custo totalmente do zero. A utilização de sistemas baseado em regras, assim como árvores de decisão podem ajudar a gerar estimativas confiáveis, aliando estes a uma base de dados de *design*, juntamente com dados de manufatura.

As técnicas qualitativas podem ser divididas entre intuitivas e analógicas (Ben-Arieh & Qian, 2003). As intuitivas são baseadas no uso de experiências passadas. O conhecimento prático já adquirido (*know-how*) pode ser definido em forma de regras, diagramas de decisão, como por exemplo uma base de dados, para auxiliar o utilizador final a melhorar/otimizar o processo de tomada de decisão e preparar a orçamentação para novos artigos baseado em informações prévias. As técnicas analógicas usam critérios de similaridade baseados em dados históricos de custo para produtos com custo já definidos, porém com informações limitadas, como modelos de análise de regressão com variáveis qualitativas, binárias (*dummy*), e redes neurais (*back propagation methods*) (Niazi et al., 2006).

O método qualitativo é geralmente favorável para realizar estimativas nos estágios iniciais de projeto, utilizando dados históricos sem requerer informações muito detalhadas, como dados geométricos ou resultados de planejamento dos processos. À vista disso, a precisão do método qualitativo é por vezes questionável, porém essa aproximação utilizada no início permite obter um bom suporte para a tomada de decisão, sendo utilizada durante todos os estágios de projeto e desenvolvimento. Desse modo, o custo do produto pode ser controlado desde o período de *design* conceitual, início de projeto, quando as alternativas desse *design* têm uma influência direta no custeio para as etapas de produção, visto que os processos serão realizados e adaptados em cima das escolhas prévias (Niazi et al., 2006).

3.4.2 Método quantitativo

Por outro lado, as técnicas quantitativas são fundamentadas em uma análise detalhada de um produto específico, seus componentes e singularidades, e os processos produtivos ao invés de apenas confiar em um histórico de dados/conhecimentos. Nesta metodologia os custos são calculados utilizando funções analíticas de certas variáveis, que representam os parâmetros do produto ou unidades elementares que representam o consumo de diferentes recursos consumidos durante o ciclo de manufatura do produto. Embora esses métodos gerem resultados mais precisos, são normalmente restritamente utilizados nas fases finais de desenvolvimento devido a uma exigência de informações mais detalhadas do produto (Niazi et al., 2006).

As técnicas quantitativas podem ser categorizadas em paramétricas e analíticas (Ben-Arieh & Qian, 2003). Modelos paramétricos são derivados aplicando metodologias estatísticas e expressando o custo como uma função das suas variáveis constituintes, consideradas independentes. Esse modelo pode ser efetivo nas situações em que os parâmetros, algumas vezes conhecidos como fatores de custo, possam ser facilmente identificados. Estes modelos são geralmente utilizados para quantificar o custo unitário de um tal produto. A aproximação analítica requer a decomposição do produto em várias unidades elementares (produto dividido em componentes), separando as operações e atividades para cada uma dessas unidades, que representam diferentes recursos consumidos durante o ciclo de produção, expressando o custo como um somatório de todas estas unidades (Niazi et al., 2006). Uma forma de estimar os custos através de métodos quantitativos também é a utilização do método de regressão. É possível usar este método criando-se uma base de dados, com a informação dos produtos já existentes relativa ao cálculo das componentes para cada operação, separando entre os diversos processos e definindo os custos para cada um separadamente, com relação a mão de obra. Esta abordagem reflete a atribuição dos custos baseada nas atividades requeridas para obter o produto, de acordo com o método ABC. Da mesma forma pode ser utilizada a regressão para estimar a quantidade de matéria-prima, criando-se uma base de dados com o cálculo dos custos de matéria prima para produtos já existentes. Em seguida, introduzem-se os conceitos da regressão linear simples e da regressão linear múltipla.

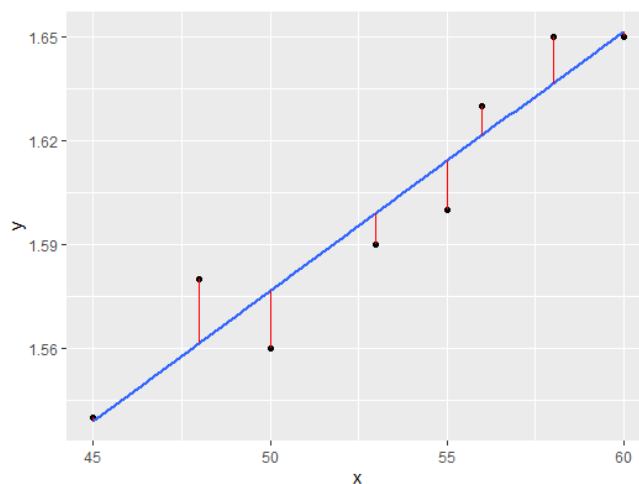
3.5 Regressão linear simples (RLS)

O modelo de RLS estima a relação entre uma variável dependente (y), que representa a resposta do sistema e uma variável independente (x). Este tipo de modelo é conhecido como uma regressão linear de um fator (Kuonen, 2004). Modelos de regressão são utilizados para prever um comportamento de um sistema extrapolado através de valores de variáveis dependentes mensurados anteriormente quando o sistema é testado com variáveis particulares conhecidas.

Observando a **Figura 3.4**, observa-se uma tendência linear entre os valores reais y e x (pontos representados a preto) observados na amostra de dados. Assim, é possível obter a função linear $\hat{y} = f(x)$, representada pela linha a azul, minimizando a soma dos quadrados dos desvios entre cada observação real e o valor estimado, $y - \hat{y}$ (método dos

mínimos quadrados) para todos os pontos observados. Considerando este exemplo a melhor maneira de generalizar um modelo de regressão linear é assumir que o valor esperado de y é uma função linear de x , sendo esse valor residual representado pelo segmento a vermelho no gráfico da **Figura 3.4** (Douglas C. Montgomery, 2011).

Figura 3.4 - Exemplo de RLS



Fonte: o próprio autor.

Os valores residuais, $y - \hat{y}$, são as diferenças entre os valores reais observados e os valores correspondentes estimados pela reta de regressão. Caso o valor estimado esteja acima da reta, terá um valor residual positivo, caso contrário, o valor estimado está abaixo da reta, sendo o erro negativo. Caso a reta seja um bom preditor do modelo com os dados utilizados, será expectável que os valores residuais sejam independentes e sigam uma distribuição normal (ou distribuição Gaussiana) com uma média próxima de zero (Douglas C. Montgomery, 2011). Assim, os pressupostos do modelo de RLS são:

- Relação linear entre a variável dependente e a independente;
- Independência dos resíduos;
- Normalidade dos resíduos;
- Homocedasticidade, ou seja, uma variância constante para os resíduos.

O primeiro passo para a análise do modelo de regressão é averiguar se os pressupostos são satisfeitos. Após essa verificação é obtida a equação que melhor descreve os dados seguida da validação da equação analisando se a capacidade preditiva do modelo é boa

ou não. Essa capacidade preditiva pode ser melhor analisada através dos coeficientes de determinação R^2 e R^2 ajustado.

3.6 Regressão linear múltipla (RLM)

A RLM é uma generalização da RLS, utilizada para modelar a relação entre uma variável dependente de saída (y) e múltiplas variáveis independentes (x_i) de entrada. Embora a técnica seja limitada para expressar uma função linear, ela se beneficia de uma estrutura matemática bem desenvolvida que produz estimativas pontuais e intervalos de confiança para os coeficientes de regressão (Marill, 2004).

Assim como a RLS possui os mesmos pressupostos, com a adição do pressuposto da ausência da multicolinearidade, o que significa que as variáveis x_i são independentes entre si. A existência de multicolinearidade refere-se à relação linear entre duas ou mais variáveis, o que também significa uma falta de ortogonalidade entre elas (Daoud, 2018). Um dos efeitos mais sérios da multicolinearidade é que as variáveis independentes terão desvio padrão elevado, gerando uma alta variabilidade nos valores, não distribuídos ao redor de uma média. Esta situação faz com que os coeficientes não sejam confiáveis, diminuindo assim suas precisões (Alin, 2010). A multicolinearidade também pode implicar que algumas variáveis não sejam estatisticamente significativas quando elas deveriam ser (Farrar & Glauber, 1967). Uma das formas de medir o grau de multicolinearidade entre as variáveis independentes é a utilização do fator de inflação da variância (VIF). Quando o VIF ultrapassa um limite de 10 é um sinal de um sério problema de multicolinearidade (Paul, 2006). Quando este limite é ultrapassado, o problema de multicolinearidade geralmente é resolvido eliminando uma ou mais variáveis da análise (O'Brien, 2007). Portanto, para um modelo de RLM, não deve existir multicolinearidade. No próximo tópico descreve-se como é realizada a seleção das variáveis independentes para o modelo.

Quando utiliza-se a RLM, será preferível utilizar o valor de R^2 ajustado, já que o valor de R^2 sempre aumenta quando novas variáveis são adicionadas ao modelo e o R^2 ajustado aumenta somente se os novos preditores tendem a melhorar o desempenho do modelo (Harel, 2009). Essa medida R^2 ajustado significa a fração da variação em Y que pode ser explicada pela variação em X através da sua relação linear. Também é necessário

avaliar se os dados se ajustam ao modelo linear antes que a equação seja aplicada para previsão e estimativa (Pak & Oh, 2010).

3.8 Métodos de seleção de variável nos modelos de regressão

O objetivo geral da seleção de variáveis no modelo de regressão seria a inclusão de todas as variáveis particulares que adicionam informações independentes e que evitam multicolinearidade e sobreposição excessiva. Quando a multicolinearidade existe, os coeficientes das variáveis independentes podem ser influenciados, sendo que os resíduos podem aumentar pela inclusão de poucas ou muitas variáveis no modelo de regressão (Marill, 2004).

Mesmo utilizando métodos manuais para selecionar as variáveis particulares, existe um amplo debate sobre se os investigadores deveriam geralmente utilizar o menor número de variáveis particulares que geram um bom ajustamento aos dados, ou se todas as variáveis independentes deveriam ser inclusas para melhorar a capacidade preditiva do modelo (Wears & Lewis, 1999).

Uma das abordagens é desenvolver uma experiência para coletar dados sobre as variáveis preditoras fundamentais conhecidas e suspeitas de serem mais importantes com base no conhecimento científico atual. Desse modo o investigador pode analisar os resultados para confirmar que os resultados satisfazem os pressupostos do modelo linear, e que todas as variáveis particulares incluídas contribuem com informações exclusivas e não demonstram um excesso de multicolinearidade (Marill, 2004).

Por causa das tarefas computacionais para avaliar todas as possíveis regressões, vários métodos foram propostos para avaliar apenas um número pequeno de subconjuntos de variáveis por adição ou subtração de variáveis, uma de cada vez, de acordo com um critério específico. Estes procedimentos, consistem em duas ideias básicas chamadas de *forward selection* (FS) e *backward elimination* (BE) (Efroymson, 1966).

A técnica FS inicia sem variáveis na equação e adiciona uma variável de cada vez até que todas as variáveis estão inclusas ou até que um critério de paragem é satisfeito (Hocking, 1976). As variáveis independentes adicionadas ao modelo são as que contribuem para explicar uma proporção significativa da variável dependente, o que pode ser

verificado através do valor de prova (*p-value*) (Lilja, 2016). A BE inicia com a equação em que todas as variáveis estão inclusas, eliminando as variáveis uma de cada vez (Hocking, 1976). As variáveis são retiradas também se baseando no *p-value* de cada uma, sendo que para esta técnica as variáveis não significativas estatisticamente são retiradas do modelo.

3.8.1 Akaike Information Criteria (AIC)

O *Akaike Information Criteria* (AIC) foi desenvolvido pelo estatístico japonês Hirotugu Akaike em 1973 (Profillidis & Botzoris, 2019). O AIC é conhecido como um avaliador de erros de predição e, conseqüentemente, da qualidade dos modelos estatísticos relativos a uma base de dados (McElreath, 2020). É calculado através da equação (3.1).

$$AIC = 2k - 2\ln(\hat{L}) \quad (3.1)$$

Sendo,

- k : número de parâmetros estimados no modelo;
- \hat{L} : valor máximo da função de verossimilhança;

Visualizando um conjunto de dois ou mais modelos de regressão que se adequam aos dados, o AIC estima a qualidade de cada modelo, comparando-os um ao outro (Taddy, 2014) e pode ser utilizado para a comparar modelos de regressão com diferentes números de parâmetros (Scott et al., 1991). Como o AIC estima a qualidade e define o melhor modelo, seleciona ao mesmo tempo as variáveis mais significativas e importantes para o modelo. O AIC faz a previsão da quantidade relativa de informações perdidas para um determinado modelo, sendo que quanto menos informações um modelo tiver a perder, maior será sua qualidade (McElreath, 2020). Ao estimar a quantidade de informações perdidas por um modelo, o AIC lida com um equilíbrio entre a qualidade de ajuste do modelo e simplicidade do modelo. Em outras palavras, o AIC lida com risco de um sobreajuste (“*overfitting*”) e de um sob ajuste (“*underfitting*”) (McElreath, 2020). Numa situação de seleção de modelos, o modelo que melhor se ajusta aos dados é o que possui um valor de AIC inferior (Lucambio, 2020).

3.9 Validação dos modelos de regressão

A seleção do modelo e sua validação são críticos na previsão de uma variável dependente dada a variável independente ou conjunto de variáveis independentes. A validação é um passo importante no processo de desenvolvimento dos modelos de regressão e auxilia na avaliação da confiabilidade dos modelos antes que possam ser utilizados na tomada de decisão. A validação envolve a comparação das previsões do modelo com observações análogas dos experimentos não utilizados para desenvolver o modelo ou valores da literatura. Os resultados devem ser eventualmente validados utilizando um conjunto de dados semelhante, mas separado (Diaconis & Efron, 2018). Sendo assim, modelos não podem ser usados com confiança até que tal comparação seja feita e, portanto, a validação é essencial, permitindo os pesquisadores entenderem a faixa aplicável dos modelos e os limites do seu desempenho (Jagannath & Tsuchido, 2003). Os modelos podem ser examinados através do R^2 ajustado e da análise dos resíduos, a fim de averiguar a qualidade do modelo. A validação pode ser feita dividindo a base de dados em dados de treino e dados de teste, e verificar como o modelo se comporta na previsão dos valores dos dados de teste (Lilja, 2016). Para analisar e validar os modelos, são calculados além dos erros das previsões, o erro absoluto médio – *Mean Absolute Error* (MAE) através da equação (3.2), que representa a média aritmética dos erros absolutos, o erro percentual absoluto médio – *Mean Absolute Percentage Error* (MAPE) representado na equação (3.3), que significa a média de todas as percentagens dos erros absolutos e o erro quadrático médio – *Mean Squared Error* (MSE), definido como a média do quadrado dos erros, representado na equação (3.4), sendo que neste caso a unidade de medida é alterada. De modo a obter o MSE na unidade de medida original, calcula-se a raiz do erro quadrático médio, obtendo-se o valor de *Root Mean Squared Error* (RMSE), representado na equação (3.5).

$$MAE = \frac{1}{n} * \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.2)$$

$$MAPE = \frac{1}{n} * \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100\% \quad (3.3)$$

$$MSE = \frac{1}{n} * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.4)$$

$$RMSE = \sqrt{MSE} \quad (3.5)$$

Sendo,

- n : número total de observações;
- y_i : valor real;
- \hat{y}_i : valor estimado.

Na próxima secção será possível visualizar alguns exemplos de modelos de regressão utilizados justamente para a estimativa de custos.

3.10 Modelos de regressão para custos

A aplicação de diferentes ferramentas para estimativa de custos mostrou que os profissionais desta área geralmente utilizam a regressão para construir os seus modelos de análise de custo (Smith & Mason, 1997).

Esses modelos utilizam dados históricos de custo para estabelecer uma relação linear entre os custos de produtos existentes e os valores de certas variáveis selecionadas para que essa relação possa ser utilizada para prever o custo de um novo produto (Huang et al., 2012). Assim, o problema de estimativa de custos pode ser aproximado no âmbito da regressão. A ideia é utilizar esta técnica estatística para estimar relações entre variáveis independentes de entrada e a variável dependente de saída. O modelo estimado é essencialmente um preditor, que uma vez alimentado com um valor particular para cada uma das variáveis independentes, retorna uma previsão do valor da variável dependente. Em outros termos, este método tem por objetivo determinar e avaliar, com base nas observações do histórico, potenciais relações entre a variável dependente e um conjunto de

variáveis independentes e utilizá-las para extrapolar para novos cenários (Adalier et al., 2007). Desse modo, caso um novo produto tenha relações não-lineares ou irregulares, este modelo não é o mais adequado (Huang et al., 2012)

Marchionni et al (2016) utilizou a RLM para obter funções de custo para diferentes tipos de recursos (bombas, tubos, tanques) ligados a sistemas de abastecimento de água, baseando-se nas suas características hidráulicas e físicas. Neste trabalho foi utilizada uma metodologia de cinco passos, construção de uma base de dados e caracterização dos recursos, cálculo atual dos valores de custo, definição de parâmetros chave e estabelecimento da função de custo, especificação do modelo e por fim o teste e a validação do modelo proposto utilizando-se 20% de dados selecionados aleatoriamente.

No artigo de Sayadi et al. (2012) é apresentado um modelo de estimativa de custos para máquinas utilizadas na mineração subterrânea, carregamento, transporte e descarregamento, utilizando o modelo de RLM, na qual a variável independente era a capacidade de caçamba (dada em metros cúbicos). No modelo de RLM, as variáveis independentes incluíam a capacidade de caçamba, largura total da máquina, altura total, e potência. A RLM foi conduzida em três passos, primeiramente com o auxílio de um *software* é estabelecida uma correlação entre as variáveis independentes, em seguida apenas as variáveis com correlações significativas são selecionadas e utilizadas como variáveis independentes nos modelos de RLM. No final, as relações de custo são estabelecidas como funções das variáveis iniciais, sendo elas a capacidade da caçamba, a largura total da máquina, a altura total e a potência. Após a análise das diferentes abordagens, os autores concluíram que o modelo de RLM poderia ser uma ferramenta útil para estimativas de custo na fase de orçamento e inicial do projeto e para estudos detalhados de viabilidade de projetos.

Estas funções de custo obtidas são um passo à frente no ramo de análise de custo, pois gera um conhecimento prévio dos custos do produto para a empresa, podendo ser utilizadas na escolha do *design* de produtos novos e na gestão da produção de produtos existentes. Sendo assim, contribui para uma administração mais sustentável, a curto, médio e longo prazo, permitindo a redução dos gastos e maximização dos lucros aumentando as margens de venda e mesmo tempo manter a competitividade no mercado (Ben-Arieh & Qian, 2003).

4. Metodologia

Este capítulo tem como finalidade apresentar a metodologia utilizada na dissertação, tanto no desenvolvimento da parte teórica, nos métodos de regressão múltipla como no desenvolvimento das bases de dados relativas aos produtos, na determinação das variáveis mais críticas no custo total do produto.

4.1 Modelo e *Software* utilizados

O modelo utilizado para o desenvolvimento do trabalho foi a RLM o que permitiu identificar as variáveis independentes que realmente impactavam diretamente nas variáveis dependentes. A análise foi desenvolvida no *software RStudio*, por ser um programa de fácil acesso, gratuito e ser essencialmente utilizado para a análise estatística, facilitando a visualização dos resultados e validação dos pressupostos, imprescindíveis no método da RLM.

Em geral uma RLM com k variáveis independentes é definida pela equação (4.1):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (4.1)$$

Sendo,

- Y : variável dependente prevista pelo modelo;
- β_0 : valor em que a reta intercepta o eixo y ;
- $\beta_1, \beta_2, \dots, \beta_k$: coeficientes de regressão;
- x_1, x_2, \dots, x_k : variáveis independentes do sistema;
- ε : erros residuais.

Os coeficientes de regressão representam a mudança média para cada uma das variáveis independentes para uma unidade de mudança na variável dependente. Ou seja, representam o quanto cada variável independente impacta no valor futuramente previsto. As variáveis independentes nesse caso serão as características quantitativas dos produtos, definidas e organizadas numa base de dados que é desenvolvida na próxima secção.

4.2 Desenvolvimento da base de dados

Os dados utilizados para realizar a análise foram recolhidos na empresa, a qual possui uma base de dados em nuvem para controle interno. Com esses dados foi possível

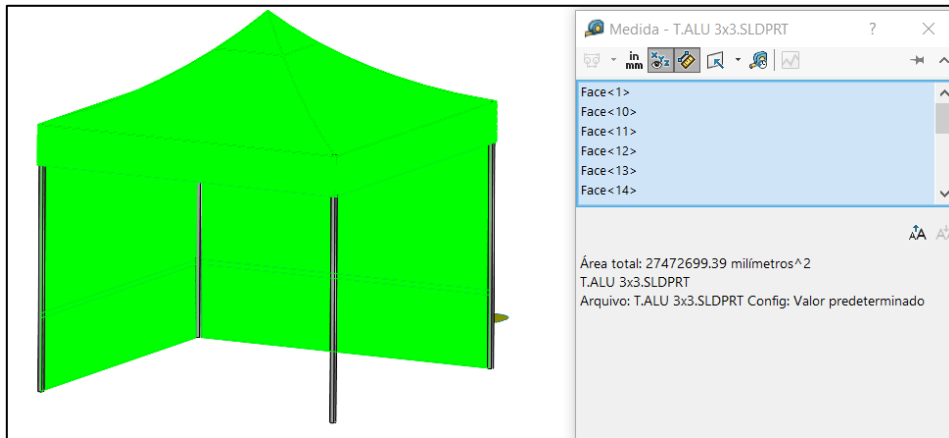
desenvolver uma planilha mestre com todas as informações necessárias para o estudo tais como os custos de matéria-prima e custos de mão de obra. Estes dados foram obtidos diretamente com o setor responsável pela orçamentação dos produtos, para além das informações técnicas, como os desenhos técnicos (3D), as planificações dos produtos (2D), dimensões, densidade das matérias-primas, para futuros cálculos de massa do produto, que foram obtidas com o setor do *design*. Apenas foram utilizados dados a partir do ano de 2020, devido à mudança nos processos produtivos, que impactou num melhor aproveitamento da matéria-prima e, conseqüentemente, na redução dos desperdícios. Para visualização e aquisição das informações técnicas dos produtos, o *software SolidWorks* foi utilizado para abrir arquivos do tipo “.igs”, que constitui o formato dos desenhos disponibilizados pela empresa.

As variáveis independentes do sistema consideradas tanto na base de dados das matérias-primas quanto para a base de dados da mão de obra, utilizadas para o desenvolvimento dos modelos de RLM, foram as seguintes:

- A área de matéria-prima (a) utilizada no projeto do produto (obtida com auxílio do SolidWorks), visualizado na **Figura 4.1**;
- A massa do produto (m), calculada multiplicando a densidade (quilograma por metro quadrado) para cada tipo de matéria-prima pela área de matéria-prima (a), conforme a **Tabela 4.1**;
- O volume aproximado (V), sendo este obtido através do volume do sólido externo ao produto, calculado multiplicando a área da maior seção transversal do produto pela sua altura máxima, demonstrado na **Figura 4.2**. Foi escolhida essa forma de determinar o volume para facilitar a etapa realizada pela empresa, já que muitos desenhos 3D possuem contornos abertos e o software não permite o cálculo do volume automaticamente;
- O perímetro total (P) de todas as peças que compõem o produto, obtido (com auxílio do SolidWorks) da forma observada na **Figura 4.3**;

- A altura média do produto (h), calculada por uma média aritmética entre todas alturas do produto (obtida com auxílio do SolidWorks), visualizado na **Figura 4.4**.

Figura 4.1 - Área de matéria-prima utilizada no produto



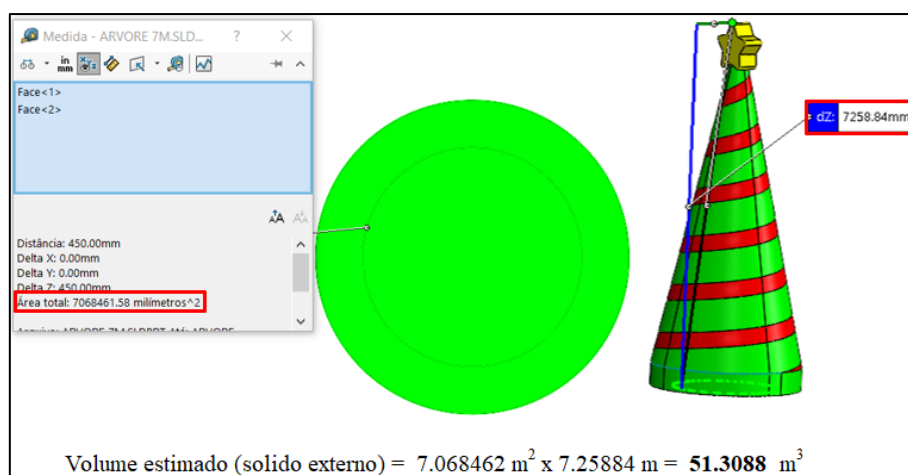
Fonte: o próprio autor.

Tabela 4.1 - Cálculo da massa do produto

Produto Totem Specialized 6m		Família F1	
Tipo de matéria prima	Área [m ²]	Densidade [kg/m ²]	Massa [kg]
Lona 140	28.88587886	0.45	12.99864549
Rede Septos	12.72323819	0.27	3.435274311
Total			16.4339198

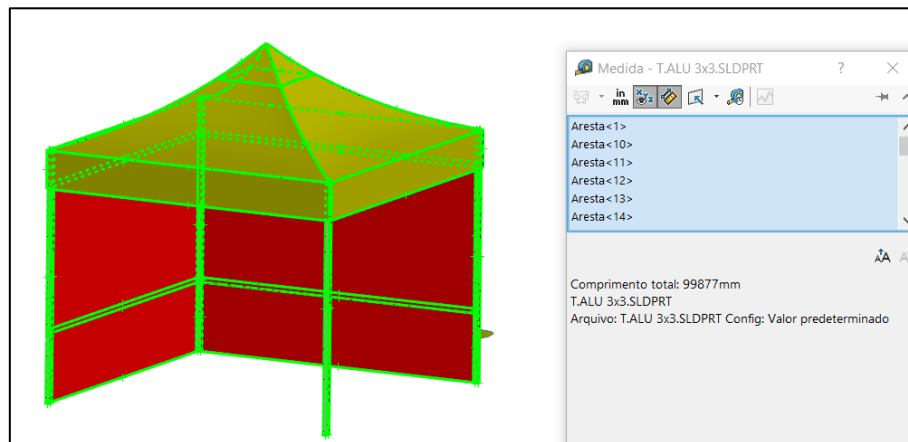
Fonte: o próprio autor.

Figura 4.2 - Volume estimado do produto



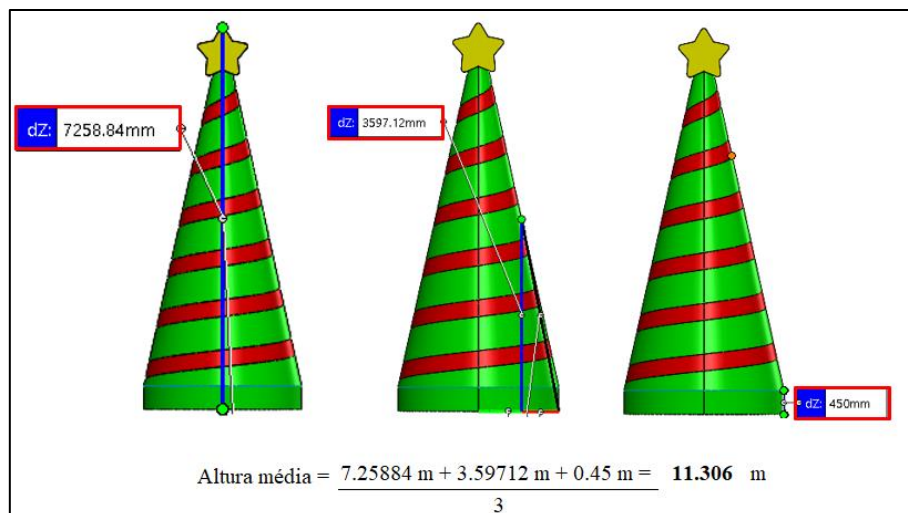
Fonte: o próprio autor.

Figura 4.3 - Perímetro total das peças do produto



Fonte: o próprio autor.

Figura 4.4 - Altura média do produto



Fonte: o próprio autor.

Seguidamente estão descritos como serão feitos os cálculos dos custos de produção neste trabalho, tanto para matéria-prima quanto para a mão de obra.

4.3 Cálculo dos custos de produção

O cálculo dos custos de produção para todas as famílias de produtos é similar, tendo algumas diferenças apenas nas matérias primas utilizadas. Os processos utilizados na confecção dos produtos são os mesmos para todas as famílias, nos quais são utilizadas máquinas de corte, impressão, costura e solda. Desse modo os custos de mão de obra são

calculados similarmente para as famílias 1, 2 e 3. Nas próximas secções estão descritas como foram feitos os cálculos de custos relacionados com a matéria-prima e a mão de obra.

4.3.1 Matéria-prima

Inicialmente, através da base de dados da empresa, foram obtidas as quantidades de cada matéria-prima utilizadas para o desenvolvimento e para a fabricação de cada produto. Juntamente com estas quantidades, foi também fornecido nessa base de dados o custo unitário de cada componente. Como por exemplo para as lonas/tecidos/redes, tem-se o custo por cada metro linear utilizado, para tintas, tem-se o custo por cada litro (l) utilizado, e para componentes tais como motores, velcros, ilhós, tem-se o custo por cada unidade. Na **Tabela 4.2** pode-se visualizar como foi feito o cálculo do custo total de matéria-prima para um produto da família F1, assim como na **Tabela 4.3** se apresenta o cálculo do custo para um produto da família F2 e na **Tabela 4.4** se pode visualizar o cálculo para um produto da família F3, de forma a identificar todos os custos de matéria-prima envolvidos no produto final. O custo por MP, apresentado na última coluna “Custo”, é determinado multiplicando os dados da coluna “Quantidade” pelos dados da coluna “Preço unitário”. Somando todas as parcelas da coluna “Custo”, obtemos o custo total de matéria-prima utilizada neste produto. Esses cálculos foram feitos para todos os produtos com os dados disponibilizados pela empresa em uma planilha mestre, a fim de poder integrar os dados numa base de dados para depois analisá-los.

Tabela 4.2 - Cálculo dos custos de matéria-prima para um produto da família 1

Produto: Meta 5x3 FactoryPlay			Família: F1		
Matéria-prima	Quantidade	Unidade de medida	Preço unitário	Custo	
Base Lona	0.436	metros lineares	€ 3.67	€ 1.60	
Lona 140 Turim	11	metros lineares	€ 2.96	€ 32.56	
Motor Totem	1	unidade	€ 65.00	€ 65.00	
Linhas Cusbor / 3000m	1	unidade	€ 8.57	€ 8.57	
Tinta MIMAKI	0.121568	litros	€ 112.00	€ 13.62	
Custo total de matéria prima				€ 121.35	

Fonte: o próprio autor.

Tabela 4.3 - Cálculo dos custos de matéria-prima para um produto da família 2

Produto: Boneco de Neve			Família: F2		
Matéria-prima	Quantidade	Unidade de medida	Preço unitário	Custo	
Lona 150	11	metros lineares	€ 3.67	€ 40.37	
Rede Septos	18	metros lineares	€ 2.29	€ 41.22	
Tecido 140	16.5	metros lineares	€ 3.30	€ 54.45	
Tecido 300	23	metros lineares	€ 4.71	€ 108.33	
Linhas Cusbor / 3000m	1	unidade	€ 8.57	€ 8.57	
Tintas	0.00079	litros	€ 15.00	€ 0.01	
Custo total de matéria prima				€ 252.95	

Fonte: o próprio autor.

Tabela 4.4 - Cálculo dos custos de matéria-prima para um produto da família 3

Produto: AP100CA750			Família: F3		
Matéria-prima	Quantidade	Unidade de medida	Preço unitário	Custo	
Fita	50	metros lineares	€ 0.20	€ 10.00	
Ilhós	52	metros lineares	€ 0.10	€ 5.20	
Lona 250	20	metros lineares	€ 5.38	€ 107.60	
Tecido 140	4	metros lineares	€ 3.30	€ 13.20	
Linhas Cusbor / 3000m	1	unidade	€ 8.57	€ 8.57	
Tintas	0.00079	litros	€ 15.00	€ 0.01	
Custo total de matéria prima				€ 144.58	

Fonte: o próprio autor.

Analisando as tabelas dos cálculos dos custos de MP, salienta-se que as lonas, tecidos e redes são as MP que mais contribuem para o custo de cada produto das famílias F2 e F3, sendo 97% do custo do produto para a F2 e perto de 84% para a F3. Para a família F1 a contribuição é menor, mas também bastante considerável, cerca de 28% do custo do produto. Portanto, o modelo a ser criado buscará estimar o VML do conjunto dessas MP críticas, lonas, tecidos e redes na fase de orçamentação do produto, considerando as características captadas por variáveis quantitativas do produto.

4.3.2 Mão de obra

De forma similar à informação recolhida para a matéria-prima, através da base de dados da empresa, foram obtidos os tempos de mão de obra para a produção dos produtos

em estudo, definidos em horas (h) de trabalho, assim como os custos unitários de cada atividade do processo, desenho/plano de corte, corte, pintura/impressão, costura, soldadura e logística, refletindo a aplicação do método ABC. Considerando o método ABC, na **Tabela 4.5** é demonstrado o cálculo do custo total de mão de obra para um produto da F1, e na **Tabela 4.6** se apresenta o mesmo cálculo para um produto da F2. Pode-se ver na **Tabela 4.7** que para a F3 temos processos diferentes, devido a esta família incluir produtos com funções totalmente diferentes e possuir peças/componentes distintas. O custo por atividade é representado na última coluna “Custo” de cada tabela, sendo o resultado da multiplicação da coluna “Quantidade” pela coluna “Preço unitário”, obtendo-se o custo que cada processo gerou no produto. De forma similar, somando as parcelas da coluna “Custo” obtém-se o custo total de mão de obra envolvida na produção do item.

Tabela 4.5 - Cálculo dos custos de mão de obra para um produto da família 1

Produto: Meta 5x3 FactoryPlay			Família F1	
Atividades do processo	Quantidade	Unidade de medida	Preço unitário	Custo
Desenho/Plano de Corte	0.5	horas	€ 6.00	€ 3.00
Corte	0.333	horas	€ 10.00	€ 3.33
Pintura/Impressão	0.5	horas	€ 7.00	€ 3.50
Costura	0.917	horas	€ 6.00	€ 5.50
Custo total de mão de obra				€ 15.33

Fonte: o próprio autor.

Tabela 4.6 - Cálculo dos custos de mão de obra para um produto da família 2

Produto: Boneco de Neve 2018 T300			Família F2	
Atividades do processo	Quantidade	Unidade de medida	Preço unitário	Custo
Desenho/Plano de Corte	0.75	horas	€ 6.00	€ 4.50
Corte	3.750	horas	€ 10.00	€ 37.50
Pintura/Impressão	0.5	horas	€ 7.00	€ 3.50
Costura	11.250	horas	€ 6.00	€ 67.50
Custo total de mão de obra				€ 113.00

Fonte: o próprio autor.

Tabela 4.7 - Cálculo dos custos de mão de obra para um produto da família 3

Produto: AP100CA750			Família F3		
Atividades do processo	Quantidade	Unidade de medida	Preço unitário	Custo	
Desenho/Plano de Corte	0.33	horas	€ 6.00	€ 1.98	
Corte	0.200	horas	€ 10.00	€ 2.00	
Pintura/Impressão	0.2	horas	€ 7.00	€ 1.40	
Costura	5.250	horas	€ 6.00	€ 31.50	
Colocar ilhós	0.666	horas	€ 6.00	€ 4.00	
Montar	0.500	horas	€ 6.00	€ 3.00	
Custo total de mão de obra				€ 43.88	

Fonte: o próprio autor.

Visualizando as tabelas de cálculo dos custos relativos à mão de obra, salienta-se que a atividade que mais contribui para o custo do produto é a costura por ser uma atividade mais demorada pelas suas características e complexidade, sendo 36% de contribuição para o custo do produto para F1, 60% para a F2 e 72% para a F3. Assim como a costura, o corte também é uma atividade crítica para alguns produtos, que possuem uma geometria mais complexa, com curvas e formas assimétricas. Nesse caso é possível visualizar que para o produto da família 2, o tempo de corte é relativamente elevado, contribuindo em 33% para o custo do produto.

Desse modo, pretende-se identificar as variáveis críticas para o estudo. As bases de dados para estimar as variáveis críticas foram criadas no software *Excel* e salvas em formato “.csv” para facilitar a manipulação no software *Rstudio*.

A variável crítica, nesse caso dependente, considerada na análise de regressão para estimar a quantidade de matéria-prima principal gasta num produto foi o VML. Pode ser visto na **Figura 4.5**, uma parte da base de dados utilizada para criar o futuro modelo capaz de estimar o VML. A primeira coluna representa a família em que o produto está inserido, e na segunda coluna um código criado para identificar cada produto. Na terceira coluna da base de dados desta variável (*ml*), estão presentes os VML reais, gastos realmente no produto, já considerando desperdícios. Nas colunas seguintes estão presentes as variáveis independentes para o futuro modelo de RLM, a área do produto (*area*), a massa do produto (*massa*), o volume estimado do produto (*volume*), o perímetro das peças do produto

(*perimetro*) e a altura média do produto (*alt_media*). Estes dados foram obtidos na base de dados da empresa, relativamente a 83 produtos diferentes, sendo 47 da família 1, 21 da família 2 e 15 da família 3. Desse modo foram calculadas, utilizando a metodologia anteriormente definida, a área, a massa, o volume, o perímetro e a altura média para cada um destes 83 produtos. A base de dados resultou em 83 observações de 6 variáveis, sendo uma dependente (VML) e as restantes independentes.

Para os tempos de mão de obra, as variáveis críticas consideradas foram o tempo de corte e o tempo de costura que se pretendem estimar. Uma parte da base de dados para o tempo de corte pode ser visualizada na **Figura 4.6**. As duas primeiras colunas representam a família e os códigos de identificação dos produtos. A terceira coluna (*corte*) apresenta a variável “tempo de corte” que inclui os tempos reais gastos no corte da matéria-prima principal para as peças do produto. Nas colunas seguintes também estão presentes as variáveis independentes utilizadas no modelo a ser definido. Os dados de tempo de corte foram fornecidos pela empresa, que incluem 6 observações relativas a 67 produtos, sendo 41 da família 1, 14 da família 2 e 12 da família 3. Da mesma forma que para o VML, foram calculadas a área, a massa, o volume, o perímetro e a altura média para cada um dos 67 produtos. A dimensão da base de dados para o tempo de corte foi de 67 observações de 6 variáveis, sendo uma dependente (tempo de corte) e as restantes independentes. Na base de dados para a variável “tempo de costura”, em que parte pode ser visualizada na **Figura 4.7**, também estão presentes, na primeira e segunda coluna, as informações relativas às famílias e códigos dos produtos, assim como na terceira coluna os tempos reais que foram gastos na confecção da costura do produto e nas colunas seguintes as variáveis independentes. Da mesma forma que os outros casos, esses dados de tempos de costura foram fornecidos pela empresa, incluindo 6 variáveis relativas a 67 produtos, sendo 41 da família 1, 15 da família 2 e 11 da família 3, calculando-se as variáveis independentes para todos os 67 produtos. A dimensão da base de dados para o tempo de costura foi de 67 observações de 6 variáveis, sendo uma dependente (tempo de costura) e as restantes independentes. Apesar da base de dados de corte e de costura terem a mesma dimensão, possuem algumas observações diferentes entre elas.

Figura 4.5 – Fragmento da base de dados para estimar o VML

familia	codigo	ml	area	massa	volume	perimetro	alt_media
1	1	5	12.56472	5.654124	6.283185	27.29896	1.5
1	3	90	149.8596	36.38677	111.5796	286.1531	3
1	4	11.686	13.56258	6.103159	7.019789	53.0269	2.473465
1	5	11.436	13.56258	6.103159	7.019789	53.0269	2.473465
1	6	44.867	44.20437	19.89197	35.14234	105.6704	4.05475
1	8	38	32.75965	14.74184	22.13816	77.01512	3.349515
1	9	32.152	38.64443	17.39	30.00983	84.85404	3.244615
1	10	33.271	49.02001	22.059	40.92526	87.44634	3.830375
1	11	33.272	47.45639	21.35537	41.24666	115.9912	4.744615
1	13	45.17	57.55278	25.89875	55.24262	104.7985	4.638
1	15	60	57.55278	25.89875	55.24262	104.7985	4.638
1	16	49.35173	57.55278	25.89875	55.24262	104.7985	4.638
1	18	54.536	49.71689	22.01631	76.99429	225.3421	4
1	19	39.291	48.68578	21.50339	70.86785	281.9443	4.5
1	20	11.5	14.85008	6.682538	11.93788	43.92559	0.834

Fonte: o próprio autor.

Figura 4.6 – Fragmento da base de dados para estimar o tempo de corte

familia	codigo	corte	area	massa	volume	perimetro	alt_media
1	4	0.666667	13.56258	6.103159	7.019789	53.0269	2.473465
1	5	0.333333	13.56258	6.103159	7.019789	53.0269	2.473465
1	6	0.916667	44.20437	19.89197	35.14234	105.6704	4.05475
1	9	0.5	38.64443	17.39	30.00983	84.85404	3.244615
1	10	0.333333	49.02001	22.059	40.92526	87.44634	3.830375
1	11	0.583333	47.45639	21.35537	41.24666	115.9912	4.744615
1	13	0.333333	57.55278	25.89875	55.24262	104.7985	4.638
1	16	0.333333	57.55278	25.89875	55.24262	104.7985	4.638
1	17	0.166667	55.7372	25.08174	31.051	155.7473	3.788018
1	18	0.333333	49.71689	22.01631	76.99429	225.3421	4
1	19	0.5	48.68578	21.50339	70.86785	281.9443	4.5
1	20	0.25	14.85008	6.682538	11.93788	43.92559	0.834
1	21	0.166667	14.85008	6.682538	11.93788	43.92559	0.834
1	23	1.5	40.09532	18.04289	30.87989	107.9256	2.89577
1	24	0.2	40.09532	18.04289	30.87989	107.9256	2.89577

Fonte: o próprio autor.

Figura 4.7 - Fragmento da base de dados para estimar o tempo de costura

familia	codigo	costura	area	massa	volume	perimetro	alt_media
1	4	1.066667	13.56258	6.103159	7.019789	53.0269	2.473465
1	5	0.916667	13.56258	6.103159	7.019789	53.0269	2.473465
1	6	7.75	44.20437	19.89197	35.14234	105.6704	4.05475
1	9	3.333333	38.64443	17.39	30.00983	84.85404	3.244615
1	10	5.666667	49.02001	22.059	40.92526	87.44634	3.830375
1	11	3.666667	47.45639	21.35537	41.24666	115.9912	4.744615
1	13	7.333333	57.55278	25.89875	55.24262	104.7985	4.638
1	16	4.116667	57.55278	25.89875	55.24262	104.7985	4.638
1	18	3.416667	49.71689	22.01631	76.99429	225.3421	4
1	19	6	48.68578	21.50339	70.86785	281.9443	4.5
1	20	3.320833	14.85008	6.682538	11.93788	43.92559	0.834
1	21	3.691667	14.85008	6.682538	11.93788	43.92559	0.834
1	23	8.541667	40.09532	18.04289	30.87989	107.9256	2.89577
1	24	5.583333	40.09532	18.04289	30.87989	107.9256	2.89577
1	25	2.75	21.16139	9.522626	16.67338	59.92559	0.834

Fonte: o próprio autor.

Através de uma estimativa das três variáveis críticas para os produtos será possível calcular seus custos envolvidos no produto, facilitando o processo inicial de orçamentação. Em seguida estão descritos os resultados da análise RLM, assim como a validação dos seus pressupostos e os modelos escolhidos.

5. Resultados e Discussão

Neste capítulo serão apresentadas e discutidas as etapas para o desenvolvimento dos modelos, os resultados na avaliação dos pressupostos e na análise dos modelos de modo a realizar a seleção final do modelo mais adequado e por fim apresenta-se a sua validação.

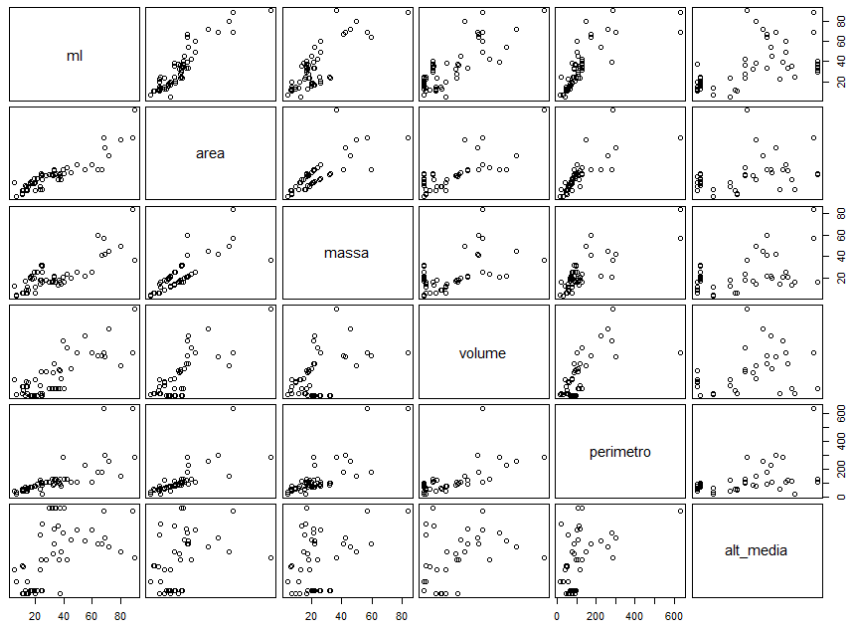
5.1 Desenvolvimento dos modelos

Os modelos foram desenvolvidos em script no *RStudio*. O *script* consiste em um programa ou uma sequência de instruções que são interpretadas ou realizadas por outro programa e não pelo processador do computador como um programa compilado. O primeiro passo para o desenvolvimento do *script* neste trabalho foi a leitura de três base de dados anteriormente criadas em formato “.csv”, relativas às três variáveis críticas a estimar por cada modelo de RLM, isto é, metros lineares de matéria-prima (VML), tempo de corte e tempo de costura, que são consideradas variáveis dependentes para cada modelo de RLM. As variáveis independentes são as mesmas para os três modelos e consistem na área do produto, massa do produto, volume estimado do produto, perímetro das peças do produto e altura média do produto, tal como referido no capítulo anterior. Para visualizar como o modelo de regressão se comporta com dados nunca vistos antes pelo modelo é necessário fazer a separação entre dados de treino e dados de teste. Os dados de treino são utilizados para a criação do modelo, estimar os coeficientes e fazer suas análises, já os de teste são utilizados para validá-lo, já que são novidade para o modelo. Para este estudo, os dados de treino foram de 80% da base de dados total, e conseqüentemente 20% destes foram utilizados para teste. Inicialmente para obter esta separação, é feito a randomização dos dados, a fim de não selecionar apenas dados de uma única família de produtos e garantir uma distribuição coerente entre os dados, sendo necessário que produtos de todas as famílias sejam selecionadas tanto para o treino quanto para o teste. Após feita a randomização é possível fazer a multiplicação da base de dados pelas correspondentes percentagens, de 20% e 80%, para obter os dados de teste e treino, respetivamente.

A fim de pré-visualizar as relações entre as variáveis para ter uma ideia se são lineares ou não, é possível representar um diagrama que fornece uma comparação em pares de todas as variáveis na base de dados de treino. Na **Figura 5.1** é possível pré-visualizar a

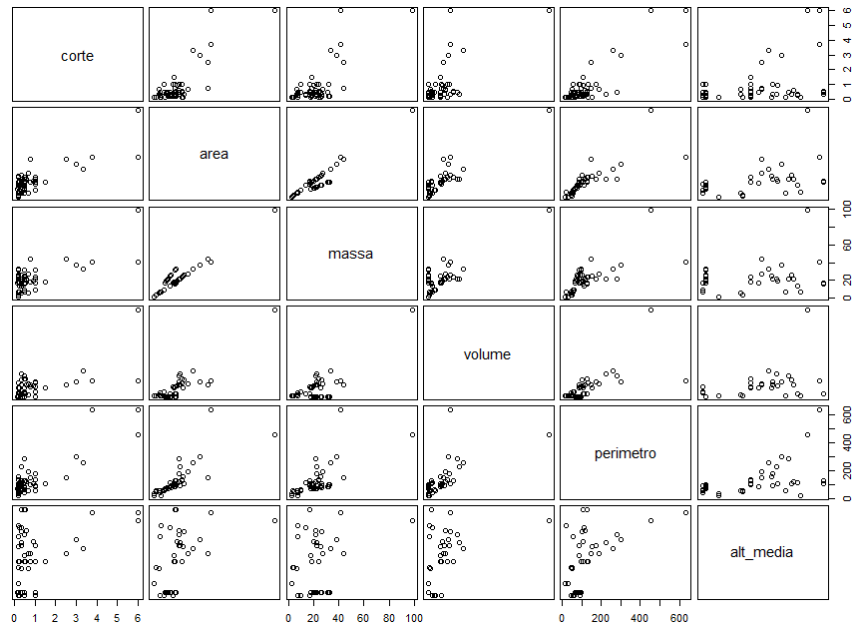
relação entre todas as variáveis, independentes e a crítica (dependente), sendo que neste caso, o interesse é a obtenção de um modelo que estime a variável crítica, o VML, portanto a pré-análise é realizada na linha “ml”, visualizando a relação linear que ela tem com cada uma das variáveis independentes. Comparando com a área do produto, massa do produto, volume estimado do produto e perímetro das peças do produto é possível notar que os pontos seguem uma tendência linear, diferindo apenas para a altura média do produto, com um padrão menos linear.

Figura 5.1 - Diagrama da relação entre as variáveis da base de dados de treino da matéria-prima

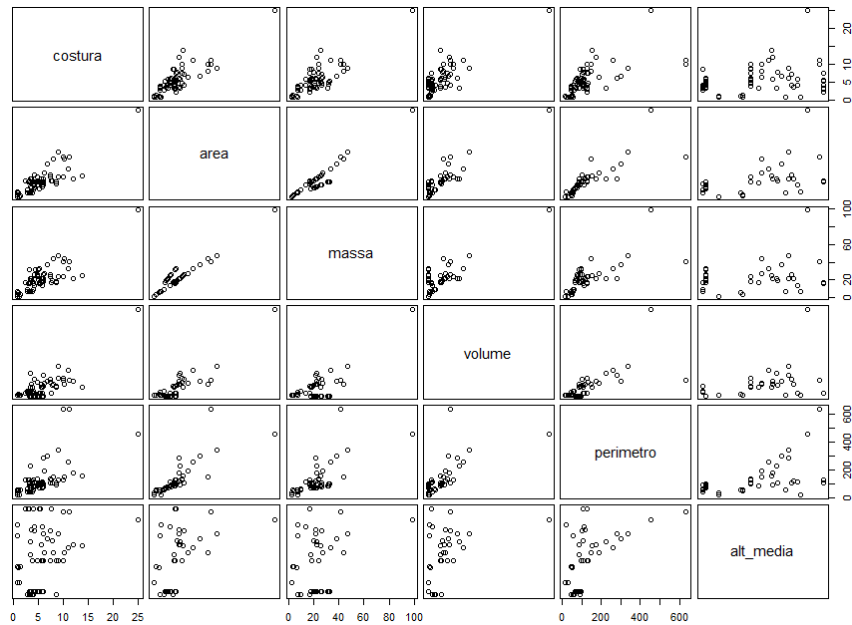


Fonte: o próprio autor.

De forma similar, obtém-se a **Figura 5.2** para pré-visualizar a relação entre todas as variáveis independentes e a crítica (dependente), neste caso o tempo de corte. A pré-análise é realizada na linha “corte”, visualizando a relação linear que ela tem com cada uma das variáveis independentes. Comparando com área, massa, volume, perímetro e altura média não é possível pré-visualizar uma tendência linear tão clara quanto para o modelo do VML, mas ainda assim é visto que existe uma relação com tendência aproximadamente linear. Para realizar também uma pré-visualização da relação entre todas as variáveis independentes e a crítica (dependente), neste caso o tempo de costura, observa-se a linha “costura” da **Figura 5.3**. Verifica-se que há uma relação linear mais perceptível para as variáveis área, massa, volume e perímetro, e não tão definida para a variável altura média.

Figura 5.2 - Diagrama da relação entre as variáveis da base de dados de treino do tempo de corte

Fonte: o próprio autor.

Figura 5.3 – Diagrama da relação entre as variáveis da base de dados de treino do tempo de costura

Fonte: o próprio autor.

Após fazer a observação dos diagramas pode-se inferir que há uma relação linear entre cada variável dependente a estimar e as variáveis independentes, desse modo opta-se pela RLM. Para cada variável dependente, foram criados dois modelos, um considerando o termo independente (*intercept*) no eixo y e o outro sem o termo independente.

5.1.2 Obtenção dos modelos de RLM

Para a geração dos modelos de regressão, é utilizada a função “lm” (*linear model*) dentro do software *RStudio*. Nessa função é relacionada a variável dependente com todas as variáveis independentes disponíveis.

Na **Figura 5.4** é possível verificar os modelos sendo criados na linguagem do *R*, o primeiro modelo “mp1” representa o modelo de RLM, considerando o *intercept*, a fim de estimar a quantidade de matéria-prima, o VML, em função da área do projeto (*area*), massa do produto (*massa*), volume estimado do produto (*volume*), perímetro das componentes (*perimetro*) e da altura média do produto (*alt_media*), o termo final “data” indica que está ser utilizada a base de dados de treino relativa à matéria-prima, que corresponde a 80% dessa base de dados, que inclui 66 observações de 6 variáveis. O segundo modelo “mp2” difere do primeiro na medida em que não considera a *intercept* (pela inclusão do termo “-1”). O modelo sem *intercept* implica que a reta da regressão inicia no ponto $x = 0$ e $y = 0$, este modelo foi considerado para análise pois é expectável que quando todas variáveis independentes são zero, o VML também seja zero, por não utilizar matéria-prima.

Figura 5.4 - Modelos de RLM na linguagem R

```
mp1 <- lm (ml ~ area + massa + volume + perimetro + alt_media, data = treino_mp)
mp2 <- lm (ml ~ -1 + area + massa + volume + perimetro + alt_media, data = treino_mp)
```

Fonte: o próprio autor.

Seguidamente do desenvolvimento do modelo de regressão para cada variável dependente a estimar, é necessário fazer a seleção das variáveis que são estatisticamente significativas para os modelos. Neste sentido, o método adotado foi o “*backward elimination*”, o qual elimina as variáveis que não são estatisticamente significativas para os modelos. A função “stepAIC” realiza esta seleção automaticamente no *software* através de diversas iterações, a cada passo retira uma variável que submetida a um *Teste t*, buscando a melhor combinação de variáveis (melhor modelo), consequentemente com o menor valor de *Akaike Information Criteria* (AIC). Este *Teste t* testa duas hipóteses com base na estatística de teste, a hipótese nula (H_0), que refere que o coeficiente de regressão é igual a zero, e a hipótese alternativa (H_1), que supõe o coeficiente de regressão é diferente de zero. Caso *p-value* seja menor que 0,05 (nível de significância adotado), rejeita-

se H_0 (neste caso a variável mantém-se no modelo), caso contrário não se rejeita H_0 (neste caso a variável é retirada do modelo). A **Figura 5.5** demonstra a utilização do método “*backward elimination*” implementado através da função “stepAIC” nos modelos “mp1” e “mp2”, obtendo-se os modelos “stepmp1” e “stepmp2” respectivamente. Assim, ao utilizar-se a direção *backward* significa que os modelos iniciam com todas as variáveis e algumas são retiradas uma a uma.

Figura 5.5 - Método backward selection com a função stepAIC

```
stepmp1 <- stepAIC ( mp1, direction = "backward" )
stepmp2 <- stepAIC ( mp2, direction = "backward" )
```

Fonte: o próprio autor.

Após a execução do método “*backward elimination*”, as variáveis independentes mais significativas são mantidas nos modelos. Assim, no modelo “stepmp1” obtém-se a equação (5.1) onde é possível visualizar que todas as variáveis anteriormente adicionadas ao modelo foram mantidas, resultando um modelo de RLM com cinco variáveis independentes e com *intercept*. No modelo “stepmp2” obtém-se a equação (5.2), inicialmente também foram mantidas todas as variáveis, sendo elas área, massa, volume, perímetro, altura média, porém desconsiderando a *intercept*.

$$\widehat{VML}_1 = b_0 + b_1 * a + b_2 * m + b_3 * V + b_4 * P + b_5 * h \quad (5.1)$$

$$\widehat{VML}_2 = b_1 * a + b_2 * m + b_3 * V + b_4 * P + b_5 * h \quad (5.2)$$

Sendo,

- \widehat{VML}_1 : valor dos metros lineares a ser estimado pelo modelo “stepmp1”;
- \widehat{VML}_2 : valor dos metros lineares a ser estimado pelo modelo “stepmp2”;
- $b_k, k = 0, \dots, 5$: estimativas dos coeficientes da RLM relativos a $\beta_k, k = 0, \dots, 5$;
- a : área do produto;
- m : massa do produto;
- V : volume estimado do produto;
- P : perímetro das peças do produto;

- h : altura média do produto.

Um importante pressuposto da RLM é a ausência da multicolinearidade, ou seja, uma correlação muito alta entre as variáveis independentes, que gera um impacto na estimativa dos coeficientes, gerando redundâncias e coeficientes que algumas vezes podem não fazer sentido, como por exemplo, quando é expectável que a relação da variável seja positiva, e no modelo se obtém um coeficiente negativo. Uma das formas de medir essa multicolinearidade é o cálculo do fator de inflação da variância (VIF), que avalia justamente a correlação entre as variáveis independentes num modelo de regressão. Quando o VIF calculado for maior que 10, é considerado que há um problema de multicolinearidade. Através da função “vif” do *RStudio* é possível calcular esse fator, sendo que os resultados são apresentados na **Tabela 5.1**. Para o modelo (5.1) é verificado que todos os valores do VIF estão menores que o limite, considerando-se que não há multicolinearidade. No caso do modelo (5.2), as variáveis área e massa apresentam valores de VIF que extrapolam os limites, significando que estão gerando uma correlação alta com as outras variáveis, portanto umas delas devem ser retirada do modelo.

Tabela 5.1 - Valores do VIF para os modelos (5.1) e (5.2)

Modelo	VIF				
	Área	Massa	Volume	Perímetro	Altura média
(5.1)	5.44	4.05	2.63	3.87	1.47
(5.2)	16.69	10.93	4.72	7.83	3.6

Fonte: o próprio autor.

Desse modo, a variável área é retirada do modelo (5.2) uma vez que tem o valor de VIF mais elevado. Após a remoção, o cálculo do VIF foi realizado novamente a fim de verificar se ainda existiam correlações, obtendo-se os valores apresentados na **Tabela 5.2**.

Tabela 5.2 - Valores do VIF para o modelo "stepmp2" após a remoção da variável “Área”

Modelo	VIF			
	Massa	Volume	Perímetro	Altura média
(5.3)	6.87	3.13	7.78	3

Fonte: o próprio autor.

Após verificação nota-se que todos os valores de VIF estão dentro do limite, portanto os modelos podem ser definidos. Como uma variável foi retirada, é necessário verificar

novamente se todas as variáveis mantidas são significativas utilizando o método “*backward elimination*”, para este caso a variável perímetro foi retirada por apresentar um *p-value* maior que 0.05. O modelo (5.1) foi mantido igual e o novo modelo “stepmp2” considera apenas as variáveis massa, volume e altura média, apresentado na equação (5.3).

$$\widehat{VML}_2 = b_1 * m + b_2 * V + b_3 * h \quad (5.3)$$

De forma similar, utilizou-se o mesmo procedimento para obter os modelos de RLM para estimar o tempo de costura e o tempo de corte. Também foram estimados dois modelos para cada variável dependente (o tempo de costura e o tempo de corte), um considerando a *intercept* e o outro desconsiderando. Inicialmente, foram criados modelos com todas as variáveis, área, massa, volume, perímetro e altura média. Foram aplicadas as funções “stepAIC” e “vif” em todos os modelos a fim de obter o modelo com as variáveis mais significativas e sem multicolinearidade.

A base de dados de treino, utilizada para criar o modelo, contava com 80% da base de dados total, resultando em 53 observações de 6 variáveis. Os modelos para estimar o tempo de corte (T_c) foram chamados de “stepc1”, que considera área, perímetro e altura média como variáveis independentes, com *intercept*, e “stepc2”, que considera volume, perímetro e altura média, sem *intercept*. O sufixo “c” vem da palavra corte na língua inglesa, “*cut*”. Os modelos obtidos para estimar T_c são representados pelas equações (5.4) e (5.5).

$$\widehat{T}_{c1} = b_0 + b_1 * a + b_2 * P + b_3 * h \quad (5.4)$$

$$\widehat{T}_{c2} = b_1 * V + b_2 * P + b_3 * h \quad (5.5)$$

Sendo,

- \widehat{T}_{c1} : valor do tempo de corte a ser estimado pelo modelo “stepc1”;
- \widehat{T}_{c2} : valor do tempo de corte a ser estimado pelo modelo “stepc2”;

- $b_k, k = 0, \dots, 3$: estimativas dos coeficientes da RLM relativos a $\beta_k, k = 0, \dots, 3$;
- a : área do produto;
- V : volume estimado do produto;
- P : perímetro das peças do produto;
- h : altura média do produto.

A base de dados de treino para a costura também contou com 53 observações de 6 variáveis. Os modelos para estimar o tempo de costura (T_s) foram designados por “steps1”, que considera a área, o volume e a altura média como variáveis independentes, com *intercept*, e “steps2”, que considera a massa, o volume e a altura média, sem *intercept*. O sufixo “s” vem da palavra costura na língua inglesa, “*sew*”. Os modelos obtidos para estimar T_s são representados pelas equações (5.6) e (5.7).

$$\widehat{T}_{s1} = b_0 + b_1 * a + b_2 * V + b_3 * h \quad (5.6)$$

$$\widehat{T}_{s2} = b_1 * m + b_2 * V + b_3 * h \quad (5.7)$$

Sendo,

- \widehat{T}_{s1} : valor do tempo de costura a ser estimado pelo modelo “steps1”;
- \widehat{T}_{s2} : valor do tempo de costura a ser estimado pelo modelo “steps2”;
- $b_k, k = 0, \dots, 3$: estimativas dos coeficientes da RLM relativos a $\beta_k, k = 0, \dots, 3$;
- a : área do produto;
- m : massa do produto;
- V : volume estimado do produto;
- h : altura média do produto.

Em seguida, serão analisados se os modelos obtidos atendem aos pressupostos relativos aos resíduos, já que o pressuposto da multicolinearidade foi verificado nesta secção.

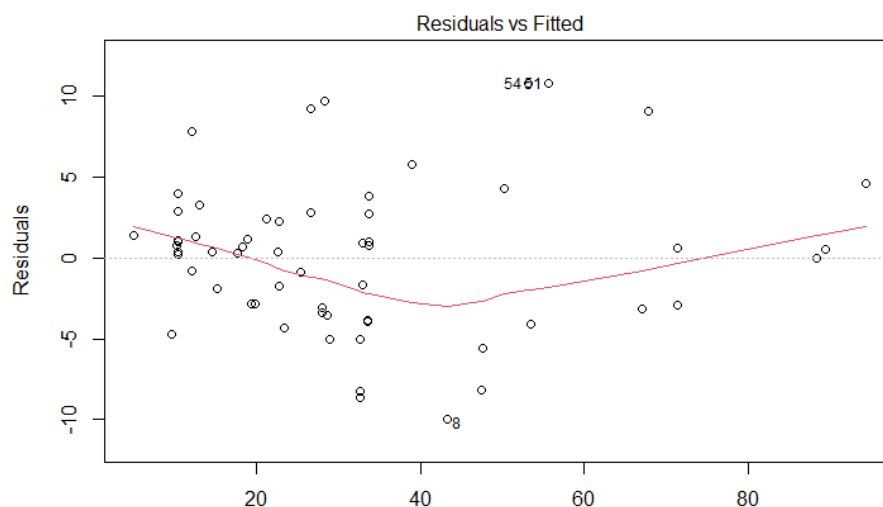
5.2 Avaliação dos pressupostos relativos aos resíduos dos modelos RLM

Pretende-se avaliar os pressupostos dos modelos de RLM obtidos anteriormente. Inicialmente, foi feita a análise gráfica dos resíduos obtidos para cada modelo. Assim, quatro gráficos foram gerados, para verificar a variância constante (homocedasticidade) dos resíduos e se estes apresentam uma distribuição normal com média nula. Pretende-se também verificar se existem *outliers* e pontos influentes.

Com finalidade de verificar a homocedasticidade, variância constante dos resíduos, são representados os gráficos dos resíduos em função dos valores previstos para os modelos. Para existir homocedasticidade é expectável que o gráfico também seja uma nuvem de pontos, em torno do eixo $y = 0$, linha pontilhada, sem tendência, verificada pela linha vermelha, possuindo uma variabilidade (distância entre os pontos, mínimos e máximos) praticamente constante.

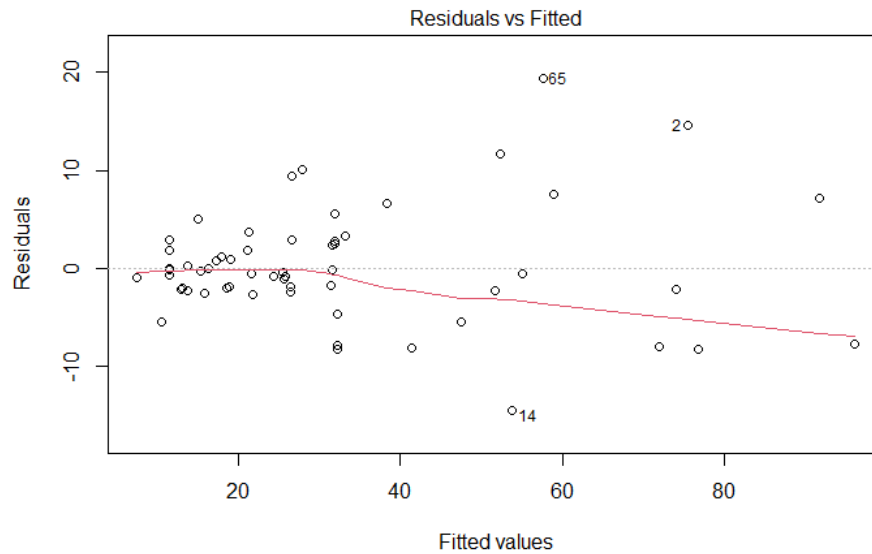
No caso do gráfico da **Figura 5.6** obtido para o modelo (5.1), é possível visualizar que a linha vermelha segue um padrão horizontal muito próximo da pontilhada, e que constitui praticamente uma nuvem de pontos em torno do eixo $y = 0$. Na **Figura 5.7** apresenta-se o gráfico para o modelo (5.3), que também possui uma linha vermelha horizontal muito próxima da pontilhada, no entanto, podemos notar uma variabilidade maior quando os valores previstos são maiores.

Figura 5.6 - Gráfico dos resíduos em função dos valores previstos para o modelo (5.1)



Fonte: o próprio autor.

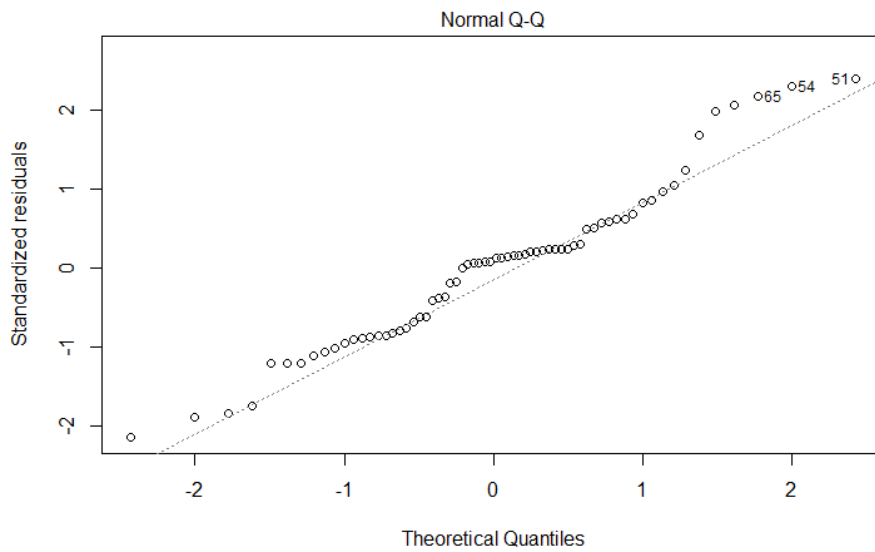
Figura 5.7 - Gráfico dos resíduos em função dos valores previstos para o modelo (5.3)



Fonte: o próprio autor.

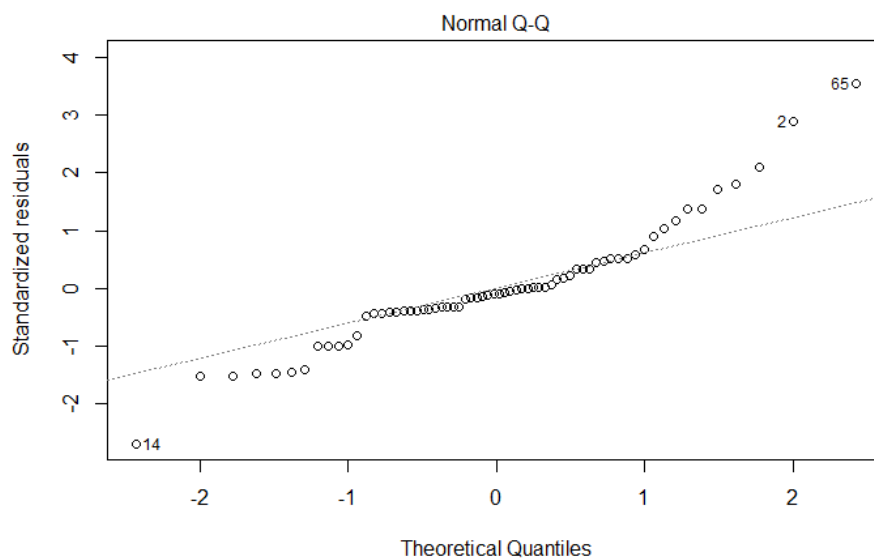
Para confirmar a existência de homocedasticidade, foi realizado o teste de Breusch-Pagan (Silva et al., 2009). Esse teste tem como hipótese nula (H_0) que há homocedasticidade e como hipótese alternativa que há heterodasticidade (H_1). O *p-value* obtido para o modelo (5.1) foi de 0.1616, não se rejeitando a hipótese nula, concluindo-se que há homocedasticidade. Já para o modelo (5.3), o *p-value* determinado foi de 0.01217, assim rejeita-se H_0 e verifica-se H_1 , ou seja, existe heterodasticidade.

Para investigar se os resíduos seguem uma distribuição normal, o gráfico Q-Q (QQ *plot*) é analisado. Nesse gráfico, o eixo y representa os resíduos padronizados e o eixo x os resíduos teóricos, esperados caso a distribuição fosse efetivamente normal. Caso os resíduos tenham uma distribuição normal, os pontos do gráfico devem estar sobre a linha pontilhada, com um padrão linear. Na **Figura 5.8** é visualizado o gráfico Q-Q para o modelo (5.1), no qual é observado que os pontos não estão perfeitamente alocados na linha pontilhada mas muito próximos, permitindo considerar uma normalidade nos resíduos do modelo.

Figura 5.8 - Gráfico Q-Q para o modelo (5.1)

Fonte: o próprio autor.

Por outro lado, na **Figura 5.9** representa-se o gráfico Q-Q para o modelo (5.3), sendo observado que para valores menores e maiores dos resíduos padronizados, os pontos estão muito afastados da linha pontilhada, aproximando-se apenas na zona central do gráfico. Portanto através dessa pré-visualização dos gráficos é possível considerar que o modelo (5.1) possui normalidade nos resíduos, mas para o modelo (5.3) é difícil confirmar este pressuposto.

Figura 5.9 - Gráfico Q-Q para o modelo (5.3)

Fonte: o próprio autor.

De forma a confirmar a normalidade dos resíduos, foi realizado o teste de *Shapiro-Wilk* para os dois modelos. Este teste tem como hipótese nula (H_0) que os resíduos seguem uma distribuição normal e como hipótese alternativa (H_1) que os resíduos não seguem uma distribuição normal. Primeiramente para o modelo (5.1) foi determinado um *p-value* de 0.08111, como este valor é maior que 0.05, não se rejeitou H_0 , desse modo a distribuição dos resíduos pode ser considerada normal. Seguidamente para o modelo (5.3), o *p-value* foi de 0.002068, assim rejeita-se a H_0 , sendo que a distribuição dos resíduos não pode ser considerada normal, confirmando a primeira consideração feita analisando os gráficos Q-Q.

Alguns *outliers*, pontos discrepantes, foram eliminados diretamente no desenvolvimento da base de dados e já não foram incluídos no *dataset* de treino, durante a análise dos dados juntamente a empresa. Para a identificação de outliers não facilmente vistos é possível utilizar a função *summary* para observar os resíduos padronizados, os quais devem estar contidos em um intervalo de - 3 a + 3. A **Tabela 5.3** ilustra os resultados da função, nela podemos visualizar que os resíduos mantiveram-se nos limites para o modelo (5.1), não existindo *outliers* influentes. Para o modelo (5.3) existe um resíduo padronizado máximo que ultrapassa o limite + 3, desse modo esse modelo gera alguns pontos influentes. A mediana próxima de zero também é um bom indicativo para o modelo.

Tabela 5.3 - Resíduos padronizados para os modelos (5.1) e (5.3)

Resíduos Padronizados						
Modelo	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
(5.1)	-2.139	-0.807	0.110	-0.002	0.512	2.405
(5.3)	-2.690	-0.400	-0.102	0.001	0.415	3.551

Fonte: o próprio autor.

Para a análise do pressuposto da independência dos resíduos foi realizado o teste de Durbin-Watson (Faraway, 2000). É considerado que não há correlação nos resíduos quando o valor da estatística Durbin-Watson fica em torno do valor 2, considerando-se aceitável na faixa de valores de 1.5 a 2.5. Verifica-se que para os modelos (5.1) e (5.3) os valores da estatística Durbin-Watson são de 2.16 e 2.15 respectivamente, muito próximo de 2. O teste de Durbin-Watson tem como hipótese nula (H_0) que não há correlação, e com hipótese alternativa (H_1) que há correlação entre os resíduos. O *p-value* é de 0,532

para o modelo (5.1) e de 0.49 para o modelo (5.3), logo H_0 não é rejeitada, considerando-se que não há correlação entre os resíduos.

Após feitas as verificações dos pressupostos, conclui-se que o modelo (5.1) atende a todos os pressupostos da RLM, portanto é possível manter o modelo de regressão e analisá-lo. O modelo (5.3) deixa de atender alguns pressupostos básicos da RLM. Logo no gráfico dos resíduos, **Figura 5.7**, é possível notar que há uma variabilidade maior à medida que os valores previstos aumentam. Verifica-se também que existe um problema de homocedasticidade, que é confirmada pelo teste de Breusch-Pagan. Outro pressuposto não atendido é a falta da normalidade dos resíduos, o que é confirmado pelo gráfico Q-Q do modelo e pelo teste de Shapiro-Wilk. Posteriormente dessa verificação, é possível selecionar o modelo (5.1) para estimar a variável crítica VML, sendo que este modelo atendeu a todos os pressupostos.

As mesmas verificações para validação dos pressupostos foram feitas nos modelos de RLM para estimativa do tempo de costura e para estimativa do tempo de corte, sendo que os resultados obtidos relativos aos testes de Shapiro-Wilk, Breusch-Pagan, Durbin-Watson e a tabela dos resíduos padronizados estão incluídos no anexo A do trabalho. No caso dos modelos definidos para o tempo de corte, apenas o modelo (5.4) atendeu aos pressupostos, sendo mantido para análise. Para o tempo de costura, como os dois modelos definidos, (5.6) e (5.7), atenderam a todos os pressupostos da regressão mantiveram-se para a análise.

5.3 Análise dos resultados dos modelos de RLM

Após a validação dos pressupostos dos modelos é possível realizar a análise dos resultados obtidos para os vários modelos de RLM.

5.3.1 Modelo para estimar o VML

Na **Tabela 5.4** são visualizados os resultados obtidos no *RStudio* para o modelo (5.1).

Tabela 5.4 - Resultados para o modelo (5.1)

Variáveis	Coefficientes	Erro Padrão	t-value	p-value
Intercept	-1.75941	1.28062	-1.374	0.17459
a	0.2446	0.050077	4.817	1.03E-05
m	0.62836	0.07739	8.12	3.06E-11
V	0.29114	0.03495	8.33	1.34E-11
P	-0.02959	0.01036	-2.856	0.00588
h	2.56415	0.38602	6.642	1.02E-08
	RSE	4.711	em 60 graus de liberdade	
	R²	0.9565	R² ajustado	0.9528
	F-Statistic	263.7	p-value	< 2.2E-16

Fonte: o próprio autor.

Na primeira coluna da tabela encontram-se as variáveis independentes presentes no modelo, juntamente com a intercept, valor onde a reta cruza o eixo y. Em seguida, na segunda coluna, apresentam-se os coeficientes de regressão, para todas variáveis independentes e o valor do intercept. Na terceira coluna apresenta-se o erro padrão para cada um dos coeficientes. A quarta coluna apresenta a estatística de teste, “*t-value*”, que é a razão entre a estimativa do coeficiente e o respectivo erro padrão. Na última coluna de cada coeficiente associado a cada variável independente, obtêm-se o *p-value* associado à estatística de teste “*t-value*”, obtido para o Teste t, que tem como hipótese nula (H_0) que o coeficiente é igual a zero, e como hipótese alternativa (H_1) que o coeficiente é diferente de zero. Como a seleção das variáveis já foi realizada, verifica-se que os valores de *p-value* para as variáveis independentes são todos menores que 0.05, portanto as variáveis são estatisticamente significativas no modelo, com exceção da *intercept*. Segundo Batista (2004), o *intercept* é mantido mesmo tendo um *p-value* acima do limite, nesse caso não é influente pois apresenta um valor bem reduzido.

As últimas três linhas dos resultados da **Tabela 5.4** demonstram algumas informações estatísticas sobre a qualidade do modelo de regressão em relação à adequação aos dados (Lilja, 2016). O modelo possui um “*residual standard error*” (RSE), desvio padrão dos resíduos de 0.4888. O R^2 é uma medida estatística que representa a porcentagem da variação da variável dependente explicada pelo modelo linear. O coeficiente R^2 ajustado, é o valor do R^2 corrigido tendo em conta o número de preditores utilizados no modelo, portanto essa é uma das medidas utilizadas para fazer uma pré-comparação entre modelos, não importando o número de preditores. No caso do modelo (5.1), o valor de R^2 foi

de 0,9565 e o valor do R^2 corrigido foi de 0,9528, muito próximo de 1, significando que o modelo se ajusta muito bem aos dados. Nem sempre o modelo com o maior R^2 ajustado é o melhor, pois não determina se as estimativas dos coeficientes são tendenciosas (Onyutha, 2020).

Verifica-se que o modelo (5.1) apresenta um valor da estatística de teste (“*F-statistic*”) igual a 263.7 no *Teste F*. Este teste faz a comparação do modelo criado, das cinco variáveis independentes, com o modelo nulo. O modelo nulo é conhecido como o modelo sem variáveis independentes, apenas incluindo a *intercept*, ou seja, o modelo vai prever que todos os produtos terão o mesmo VML, que é o valor do *intercept*. No *Teste F*, a hipótese nula define que todos os coeficientes das variáveis independentes são nulos, e como hipótese alternativa define-se que o modelo é explicado por pelo menos uma variável independente. No *Teste F*, obteve-se o *p-value* muito menor que 0,05, rejeitando a hipótese nula e considerando que o modelo criado tem uma melhor capacidade de predição comparado com o modelo nulo.

5.3.2 Modelo para estimar o tempo de corte

Os resultados do modelo (5.4) definido para estimativa dos tempos de corte está representado na **Tabela 5.5**. Pode-se observar que todas as variáveis independentes do modelo são estatisticamente significativas e que o modelo possui uma capacidade de predição melhor do que a do modelo nulo. Analisando o *p-value* obtido para o *Teste F*, esse valor é muito menor que 0,05, rejeitando-se a hipótese nula e concluindo-se que o modelo é estatisticamente significativo (hipótese alternativa). O valor do R^2 ajustado indica que o modelo linear explica 79% da variação da variável dependente, sendo um valor bem expressivo.

Tabela 5.5 - Resultados para o modelo (5.4)

Variáveis	Coefficientes	Erro Padrão	t-value	p-value
<i>Intercept</i>	-0.350196	0.133626	-2.621	0.011652
a	0.019721	0.003287	6	2.34E-07
P	0.00418	0.001095	3.819	0.000378
h	-0.091292	0.043923	-2.078	0.042928
	RSE	0.4888	em 49 graus de liberdade	
	R²	0.8031	R² ajustado	0.7911

<i>F-Statistic</i>	66.63	<i>p-value</i>	< 2.2E-16
---------------------------	-------	-----------------------	-----------

Fonte: o próprio autor.

Verifica-se que o *p-value* do *Teste T* para todas as variáveis independentes é menor que 0.05, rejeitando-se a hipótese nula e concluindo-se que todas as variáveis independentes são significativas no modelo.

5.3.2 Modelo para estimar o tempo de costura

Para a estimativa do tempo de costura dos produtos foram considerados dois modelos (5.6) e (5.7), sendo que os resultados se apresentam na **Tabela 5.6** e **Tabela 5.7**, respectivamente.

Tabela 5.6 - Resultados para o modelo (5.6)

Variáveis	Coefficientes	Erro Padrão	<i>t-value</i>	<i>p-value</i>
<i>Intercept</i>	2.313278	0.432582	5.348	2.33E-06
a	0.074298	0.013403	5.543	1.17E-06
V	0.028768	0.009866	2.916	0.00534
h	-0.285469	0.127299	-2.243	0.02949
	RSE	1.502	em 49 graus de liberdade	
	R²	0.8573	R² ajustado	0.8485
	<i>F-Statistic</i>	98.09	<i>p-value</i>	< 2.2E-16

Fonte: o próprio autor.

Tabela 5.7 - Resultados para o modelo (5.7)

Variáveis	Coefficientes	Erro Padrão	<i>t-value</i>	<i>p-value</i>
m	0.168299	0.020356	8.268	6.42E-11
V	0.030152	0.008144	3.702	0.000533
h	0.259079	0.106985	2.422	0.01912
	RSE	1.733	em 50 graus de liberdade	
	R²	0.9368	R² ajustado	0.933
	<i>F-Statistic</i>	246.9	<i>p-value</i>	< 2.2E-16

Fonte: o próprio autor.

Inicialmente, analisando os *p-value* para o *Teste F* para os dois modelos da estimativa do tempo de costura, pode-se inferir que são menores do que o nível de significância igual

a 0,05, rejeitando-se a hipótese nula e considerando a hipótese alternativa, de que o modelo é estatisticamente significativo. Observando o *p-value* para o *Teste T* nos dois modelos, verifica-se que todos os *p-value* são inferiores a 0,05 o que implica rejeitar a hipótese nula, tomando em consideração a hipótese alternativa, de que os coeficientes das variáveis independentes são estatisticamente significativos. Como os dois modelos atenderam a todos os pressupostos e possuem um R^2 ajustado bastante elevado, será feita uma comparação entre os modelos mais profunda, utilizando o *Akaike Information Criterion* (AIC), para selecionar o modelo que melhor representa os dados.

A comparação entre modelos apenas foi realizada para o caso da estimativa do tempo de costura, pois foi o único caso em que ambos modelos atenderam os pressupostos. Essa comparação foi feita através de uma função do *RStudio*, a função “AIC”, que retorna justamente este valor. Como mencionado na revisão da literatura, esse indicador representa a qualidade do modelo, sendo utilizado na comparação entre modelos, sendo que o modelo com o menor valor de AIC é considerado o melhor. O valor de AIC obtido foi de 199,4 para o modelo (5.6) e 213,6 para o modelo (5.7). Desse modo, é possível concluir que o (5.6) possui maior qualidade no ajuste, segundo o valor de AIC.

5.5 Modelos finais

Após as etapas anteriores, são definidos os modelos de RLM para cada variável a estimar. A estimativa do VML será feita a partir do modelo (5.8), designado por modelo 1. A estimativa do tempo de corte será realizada a partir do modelo (5.9) designado por modelo 2. A estimativa do tempo de costura será feita a partir do modelo (5.10) designado por modelo 3.

$$\widehat{VML} = -1,75941 + 0,2446 * a + 0,62836 * m + 0,29114 * V - 0,02959 * P + 2,56415 * h \quad (5.8)$$

$$\widehat{T}_c = -0,350196 + 0,019721 * a + 0,00418 * P - 0,02959 * h \quad (5.9)$$

$$\hat{T}_s = 2,313278 + 0,074298 * a + 0,028768 * V - 0,285469 * h \quad (5.10)$$

Sendo,

- \widehat{VML} : valor dos metros lineares a ser estimado;
- \hat{T}_c : tempo de corte a ser estimado;
- \hat{T}_s : tempo de costura a ser estimado para o modelo;
- a : área do produto;
- m : massa do produto;
- V : volume estimado do produto;
- P : perímetro das peças do produto;
- h : altura média do produto.

Adicionalmente, é possível obter o intervalo de confiança a 95% para a *intercept* e cada coeficiente do modelo. O intervalo de confiança para cada coeficiente do modelo permite que a empresa decida entre uma estimativa do coeficiente da respetiva variável que pode ser mais baixa, mais próxima à média ou uma mais conservadora, considerando valores mais elevados. Podem-se visualizar os intervalos de confiança obtidos para os modelos 1, 2 e 3, respetivamente na **Tabela 5.8**, **Tabela 5.9** e **Tabela 5.10**

Tabela 5.8 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 1

Modelo 1 - VML			
Coefficientes	Limite inferior a 95%	Limite superior a 95%	Valor estimado
Intercepto	-4.32103	0.80220	-1.75941
Área	0.14304	0.34617	0.24460
Massa	0.47357	0.78316	0.62836
Volume	0.22122	0.36105	0.29114
Perímetro	-0.05031	-0.00886	-0.02959
Altura Média	1.79199	3.33631	2.56415

Fonte: o próprio autor.

Tabela 5.9 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 2

Modelo 2 - Tempo de corte			
Coeficientes	Limite inferior a 95%	Limite superior a 95%	Valor estimado
Intercepto	-0.61873	-0.08166	-0.35020
Área	0.01312	0.02633	0.01972
Perímetro	0.00198	0.00638	0.00418
Altura Média	-0.17956	-0.00302	-0.091292

Fonte: o próprio autor.

Tabela 5.10 - Intervalos de confiança a 95% para os coeficientes das variáveis do modelo 3

Modelo 3 - Tempo de costura			
Coeficientes	Limite inferior a 95%	Limite superior a 95%	Valor estimado
Intercepto	1.44397	3.18258	2.31328
Área	0.04736	0.10123	0.07430
Volume	0.00894	0.04859	0.02877
Altura Média	-0.54129	-0.02965	-0.28547

Fonte: o próprio autor.

Após o desenvolvimento e seleção de modelos, utilizando os fundamentos estatísticos subjacentes à RLM, a última etapa corresponde à validação destes modelos para apoio à orçamentação de produtos novos da empresa utilizada como caso de estudo.

5.6 Validação dos modelos

A validação é um passo imprescindível na metodologia da RLM, pois é necessário verificar, para este caso de estudo, se as estimativas obtidas com os modelos realmente fazem sentido para fins estratégicos e de tomada de decisão.

Essa validação foi realizada utilizando dados de teste que o modelo não utilizou para ser parametrizado e portanto “nunca viu”. Esses dados são compostos pela base de dados de teste anteriormente criada para cada variável, que engloba 20% da base de dados total, obtida de forma aleatória e considerando produtos das três diferentes famílias

apresentadas no estudo. Aplicando o modelo da RLM definido para cada caso, foram realizadas previsões para as variáveis críticas dos diferentes produtos.

Na **Tabela 5.11** estão listados todos os produtos da base de dados de teste do VML. A coluna “*VML*” apresenta os VML reais utilizados na fabricação do produto e na coluna “ \widehat{VML} ” estão os valores de saída, ou seja, os valores estimados pelo modelo 1. As duas colunas definidas como “Erro das previsões”, apresentam justamente o erro de cada previsão e a porcentagem do erro quando comparado ao valor real dos VML, quando positivos significam que o valor real é maior que o previsto, e quando negativos, menor que previsto. Os resultados do MAE, MAPE, MSE e RMSE para o modelo 1 foram apresentados nas últimas duas linhas da **Tabela 5.11**.

Na **Tabela 5.12** estão apresentados os resultados da comparação entre os valores estimados e os reais para o modelo 2. Pode-se visualizar a coluna “ T_c ” que representa os valores reais de tempo de corte e a coluna “ \widehat{T}_c ” que demonstra os valores estimados para o tempo de corte. Nas duas últimas colunas encontram-se os erros para cada previsão e nas duas linhas abaixo da tabela os valores para MAE, MAPE, MSE e RMSE.

Para o modelo 3, os resultados da comparação entre os valores estimados e os reais podem ser observados na **Tabela 5.13**. A coluna “ T_s ” inclui os tempos de costura reais, e “ \widehat{T}_s ” para os valores estimados de tempo de corte. Nas duas últimas colunas também incluem-se os erros para cada previsão e nas duas linhas abaixo da tabela os valores para MAE, MAPE, MSE e RMSE.

Tabela 5.11 - Comparação entre os valores estimados e os reais para o Modelo 1

Família	Código	<i>VML</i>	\widehat{VML}	Erro das previsões	
1	36	30.3	43.2	-13.0	-43%
2	10	23.1	21.9	1.2	5%
1	9	32.2	33.2	-1.0	-3%
2	18	80.0	69.1	10.9	14%
2	13	69.1	62.6	6.5	9%
3	3	19.0	18.0	1.0	5%
3	15	16.0	17.8	-1.8	-11%
1	38	37.0	26.6	10.4	28%
1	30	25.4	32.6	-7.2	-28%
1	4	11.7	12.2	-0.5	-4%
1	13	45.2	53.5	-8.3	-18%

3	13	12.5	15.6	-3.1	-25%
1	11	33.3	44.0	-10.7	-32%
1	43	40.0	33.7	6.3	16%
1	15	60.0	53.5	6.5	11%
2	7	64.0	67.1	-3.1	-5%
2	15	6.8	5.8	1.1	16%
		MAE	MAPE	MSE	RMSE
		5.462	16%	46.196	6.797

Fonte: o próprio autor.

Tabela 5.12 - Comparação entre os valores estimados e os reais para o Modelo 2

Família	Código	T_c [h]	\widehat{T}_c [h]	Erro das previsões	
1	21	0.17	0.05	0.12	70%
2	5	1.00	0.99	0.01	1%
1	37	1.00	0.98	0.02	2%
1	31	0.25	0.05	0.20	80%
2	10	0.17	0.22	-0.06	-34%
1	36	1.00	0.98	0.02	2%
1	48	0.25	0.36	-0.11	-45%
1	4	0.67	0.09	0.58	87%
1	17	0.17	1.05	-0.89	-533%
1	16	0.33	0.80	-0.47	-140%
2	21	0.17	0.42	-0.25	-152%
1	27	0.50	0.05	0.45	90%
1	49	0.33	0.49	-0.16	-47%
1	35	0.17	0.23	-0.06	-36%
		MAE	MAPE	MSE	RMSE
		0.241	94%	0.122	0.349

Fonte: o próprio autor.

Tabela 5.13 - Comparação entre os valores estimados e os reais para o Modelo 3

Família	Código	T_s [h]	\widehat{T}_s [h]	Erro das previsões	
1	23	8.5	5.4	3.2	37%
2	8	6.3	7.3	-1.0	-16%
1	37	4.8	6.8	-2.0	-42%
1	31	2.8	3.5	-0.7	-24%
2	10	0.8	2.3	-1.5	-201%
1	36	7.4	6.8	0.6	8%
1	48	5.4	3.9	1.5	27%

1	4	1.1	2.8	-1.8	-164%
1	18	3.4	7.1	-3.7	-107%
1	16	4.1	6.9	-2.7	-67%
2	19	8.0	9.8	-1.8	-23%
1	29	3.5	3.5	-0.1	-2%
1	49	3.2	4.1	-0.9	-29%
1	35	4.0	4.2	-0.2	-4%
		MAE	MAPE	MSE	RMSE
		1.546	54%	3.485	1.867

Fonte: o próprio autor.

O MAPE possui uma desvantagem significativa, pelo fato de sua fórmula possuir uma divisão pelo valor real da variável crítica. Quando os valores reais são muito pequenos ou se aproximam de zero, a fórmula produz valores infinitos ou inflacionados, ou seja, o valor de MAPE aumenta muito (Kim & Kim, 2016). Por esse motivo, é possível visualizar um valor elevado do MAPE para o modelo 2, nos casos em que alguns valores reais são mais baixos. Para o modelo 3, o valor do MAPE é de quase 100%, pelo fato de que a maioria dos valores reais de tempo de corte se aproximam de zero. Devido a esse motivo o MAPE é uma estatística que não deve ser levada em consideração para estes casos. Quanto às outras métricas de erro, MAE, MSE e RMSE apresentaram valores reduzidos, ou seja, não tão significativos quando considerados em termos de custo para a empresa.

Sendo assim, os modelos de previsão para estimar as três variáveis críticas foram apresentados à empresa. Tendo a empresa um total conhecimento sobre seus produtos, processos produtivos e matérias-primas, o seu *feedback* sobre os modelos constitui a principal ferramenta para verificar se os resultados estimados fazem sentido e são úteis. Os resultados foram analisados para cada caso e verificados para todas as famílias de produto. O *feedback* obtido mostrou que os resultados da RLM realmente faziam sentido para a empresa na estimativa da quantidade de matéria-prima utilizada e dos tempos de corte e de costura, constituindo uma estimativa muito boa para etapas iniciais de projeto. Desse modo, as estimativas obtidas poderiam ser utilizadas para calcular estes três principais custos críticos envolvidos no produto no âmbito da realização do orçamento inicial entregue ao cliente. Em seguida são feitos os cálculos para as variáveis críticas estimadas e assim feita uma comparação com o custo real, utilizando um produto da base de dados de teste.

5.7 Cálculo dos custos

Após a validação do modelo já é possível utilizar o modelo para estimar as variáveis e calcular seus respectivos custos. Na **Tabela 5.14** é apresentado o cálculo dos custos, com as variáveis estimadas e as reais, para um dos produtos que estava presente na base de dados de teste dos três modelos, da família 1, código 36.

Tabela 5.14 - Cálculo dos custos críticos, estimados e reais, para um produto da família 1

	Variáveis críticas			Custo unitário			Custo das variáveis críticas do produto		
	VML	T_c [h]	T_s [h]	MP	Corte	Costura	VML	Corte	Costura
Estimado	43.2	0,98	6.8	3.0 € / ml	10 € /h	6 € /h	128 €	9.8 €	40.6 €
Real	30.3	1.0	7.4				89.5 €	10.0 €	44.3 €
Diferenças							38.5 €	-0.2€	-3.7 €

Fonte: o próprio autor.

Os custos foram calculados multiplicando os valores das variáveis críticas, estimadas e reais, pelo seu custo unitário. Após estes cálculos é possível visualizar a diferença entre o custo estimado e o real, sendo que o custo para o VML estimado foi de 38,5 euros a mais do que o valor real, o custo de mão de obra relativo ao corte foi de 0,2 euros a menos do que o valor real, e o custo de mão de obra de costura foi de 3,7 euros a menos do que o valor real.

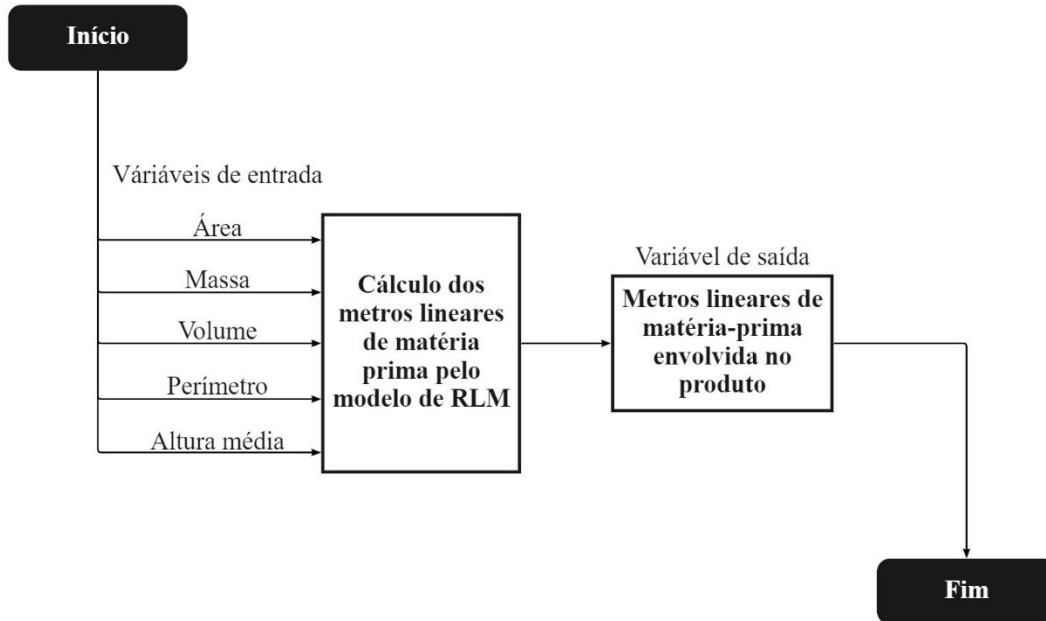
5.8 Utilização dos modelos pela empresa

Foi criada uma ferramenta para auxiliar a empresa no software *Excel*, por ser familiar e de simples para manuseio pela empresa. A ferramenta integra os modelos de RLM (5.8), (5.9) e (5.10) que permitem estimar as variáveis críticas do custo do produto. As únicas etapas que a empresa deve realizar são definir as variáveis independentes do sistema (variáveis de entrada), recolhidas no desenho 3D do produto, e apresentar os custos unitários do VML e dos processos de tempo de corte e costura. Deste modo, serão obtidas as estimativas das variáveis críticas e respetivos custos. Os intervalos de confiança também foram disponibilizados, caso a empresa julgue necessário fazer reajustes nos cálculos. De seguida, ilustra-se o procedimento.

A primeira etapa para a realização das estimativas é a inserção de variáveis de entrada pela empresa. Após essa inserção, é realizado o cálculo substituindo as variáveis

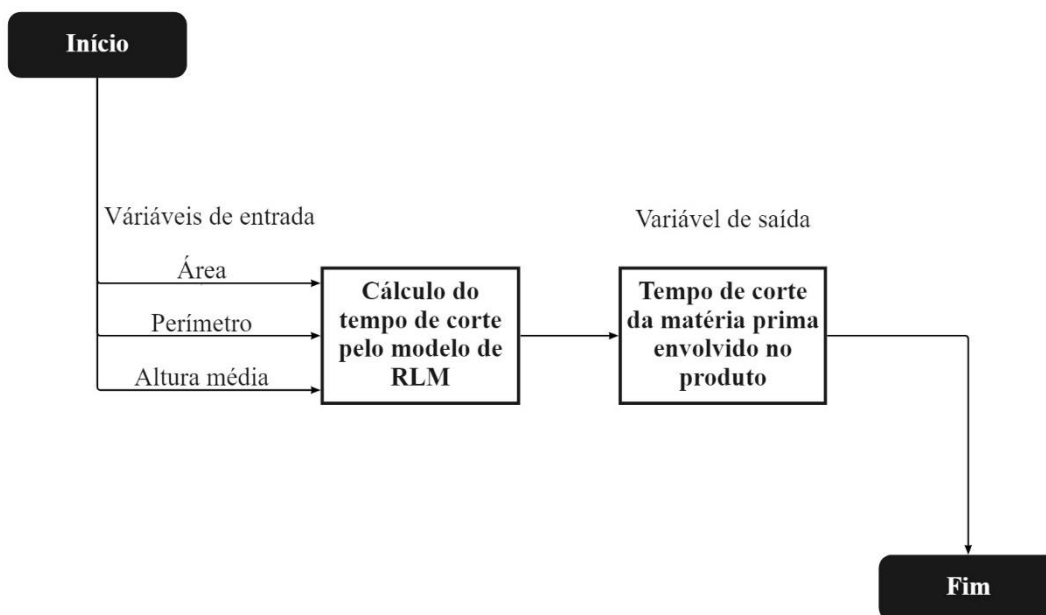
independentes pelas variáveis de entrada, gerando a estimativa da variável dependente, considerada variável de saída. Para melhor visualização dessas etapas e dos três modelos desenvolvidos, foram realizados diagramas com as respectivas variáveis de entrada e variáveis de saída. As variáveis de entrada e de saída dos modelos 1, 2 e 3 estão ilustradas na **Figura 5.10** , **Figura 5.11** e **Figura 5.12**, respectivamente.

Figura 5.10 - Diagrama para estimar o VML através do modelo 1

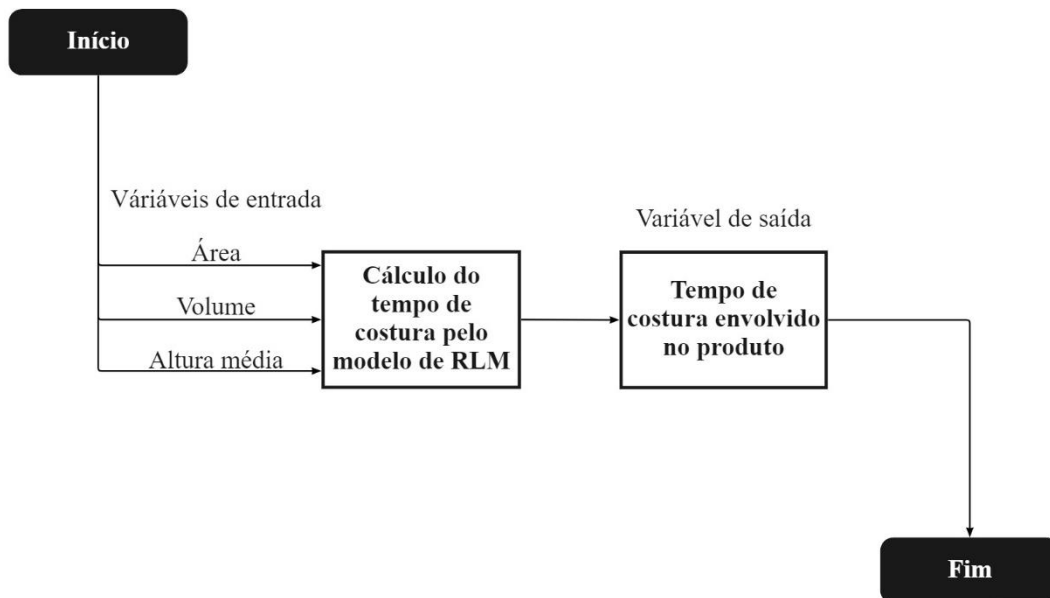


Fonte: o próprio autor.

Figura 5.11 - Diagrama para estimar o tempo de corte através do modelo 2



Fonte: o próprio autor.

Figura 5.12 - Diagrama para estimar o tempo de costura através do modelo 3

Fonte: o próprio autor.

6. Conclusão

O problema proposto no âmbito do estágio realizado numa empresa foi o desenvolvimento de um sistema de apoio à decisão na orçamentação de produtos por encomenda que seja capaz de estimar os custos críticos, ou seja, mais difíceis, na etapa embrionária do projeto. Como forma de demonstrar o funcionamento do sistema, o estudo de caso incluiu os resultados e a validação para exemplos de produtos novos, a fim de apresentar à empresa.

Entendida a importância da necessidade de uma orçamentação mais precisa, aproximando a estimativa de custo ao valor real do custo do produto, de forma a gerar mais competitividade no mercado e a maximizar lucros com manufatura de produtos com menores custos e maior produtividade, desenvolveu-se uma ferramenta para estimar os custos críticos para apoiar a decisão no processo de orçamentação de novos produtos.

Desenvolveram-se três modelos baseados em RLM, utilizando uma base de dados de treino, para realizar a estimativa das variáveis críticas do custo do produto, isto é, do VML e dos tempos de mão de obra de corte e costura a fim de realizar o cálculo dos respectivos custos na etapa inicial de desenvolvimento. A validação dessas estimativas foi feita utilizando uma base de dados de teste, na qual os modelos apresentaram um bom desempenho, avaliado pela empresa. Desse modo, foi criada uma planilha de cálculo no *Excel* como ferramenta, para otimizar e facilitar o trabalho da empresa.

Vale destacar que a literatura acadêmica enfatiza os EGF no cálculo dos custos, o que é imprescindível para um orçamento mais exato. No presente trabalho esses custos não são ressaltados, pois o objetivo principal era a determinação dos custos críticos ligados diretamente ao produto, sendo que os EGF são custos fixos e de certa forma simples de serem determinados, pela fácil identificação de todos estes encargos envolvidos na produção.

Futuros trabalhos podem incrementar o modelo com um enriquecimento da base de dados, a fim de englobar todas as famílias de produtos produzidas pela empresa, até as mais complexas que são constituídas por insufláveis para diversão e para competições esportivas. Outra maneira para otimizar ainda mais o trabalho é a criação de uma

aplicação através da programação, a fim de gerar uma ferramenta mais integrada, entre a base de dados, o modelo de regressão e a interface gráfica do usuário.

Por fim, a comparação dos valores retornados pelos modelos com os valores reais de VML e tempos de mão de obra gastos reforçam que os modelos devem ser utilizados nas etapas iniciais do projeto, tanto para ter um orçamento mais preciso quanto para poder desenvolver projetos, alternativas de *designs* diferentes, alterando formatos, dimensões, buscando reduzir a quantidade de matéria-prima utilizada e o tempo de processo envolvido no produto, a fim de comparar os custos e selecionar o mais rentável para a empresa.

7. Referências bibliográficas

- Adalier, O., Uğur, A., Korukoğlu, S., & Ertaş, K. (2007). A New Regression Based Software Cost Estimation Model Using Power Values. In *LNCS* (Vol. 4881).
Software Cost Estimation Model Using Power Values.
- Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370–374. <https://doi.org/10.1002/wics.84>
- Association for the Advancement of Cost Engineering - *Cost Engineering Terminology (CET)*, 2010. *AACE International Recommended Practice No. 10S-90*.
- Batista, D., Universidade, M., Campina, F. De, & Brasil, G. (2005). *A importância dos custos e da formação do preço básico do produto na tomada de decisões em micros, pequenas e médias empresas : um estudo de caso numa indústria de confecções em Sanharó-PE*.
- Batista, J. (2004). *Análise de Regressão Aplicada*. 265.
- Ben-Arieh, D., & Qian, L. (2003). Activity-based cost management for design and development stage. *International Journal of Production Economics*, 83(2), 169–183. [https://doi.org/10.1016/S0925-5273\(02\)00323-7](https://doi.org/10.1016/S0925-5273(02)00323-7)
- Ching, H. Y. (1997). *Gestão Baseada em Custeio por Atividade* (2nd ed.).
- Cooper, R., & Kaplan, R. (1988). *Measure costs right make the right decisions*. (pp. 96–103).
- Cooper, R., & Kaplan, R. (1991). *Profit priorities from activity-based costing* (pp. 130–135).
- D.G. Ullman. (2008). *The Mechanical Design Process*.
- Daoud, J. I. (2018). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949(1). <https://doi.org/10.1088/1742-6596/949/1/012009>
- Diaconis, P., & Efron, B. (2018). *Computer-Intensive Methods in Statistics*. 248(5), 116–131.
- Douglas C. Montgomery, G. C. R. (2011). *Applied Statistics and Probability for Engineers* (Fifth Edit).

- Efroymson, M. A. (1966). Stepwise regression-a backward and forward look. *Presented at the Eastern Regional Meetings of the Inst. of Math. Statist., Florham Park, New Jersey.*
- Faraway, J. (2000). "Linear Regression in R. In *New York.*
- Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, 49(1), 92.
<https://doi.org/10.2307/1937887>
- Harel, O. (2009). The estimation of R2 and adjusted R2 in incomplete data sets using multiple imputation. *Journal of Applied Statistics*, 36(10), 1109–1118.
<https://doi.org/10.1080/02664760802553000>
- Hocking, R. R. (1976). A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1), 1.
<https://doi.org/10.2307/2529336>
- Jagannath, A., & Tsuchido, T. (2003). Validation of a polynomial regression model: The thermal inactivation of *Bacillus subtilis* spores in milk. *Letters in Applied Microbiology*, 37(5), 399–404. <https://doi.org/10.1046/j.1472-765X.2003.01416.x>
- Katholieke. (1981). Production planning: a review Ludo F. GELDERS and Luk N. VAN Production planning decides upon the use of pro-WASSENHOVE duction resources in order to satisfy forecast demand. In *European Journal of Operational Research* (Vol. 7).
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679.
<https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Kuonen, D. (2004). Book Review: Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. In *Statistical Methods in Medical Research* (Vol. 13, Issue 5). <https://doi.org/10.1177/096228020401300512>
- Lilja, D. (2016). Linear Regression Using R: An Introduction to Data Modelling. In *Linear Regression Using R: An Introduction to Data Modelling.*

<https://doi.org/10.24926/8668/1301>

Lucambio, F. (2020). *Seleção de modelos Critério de Informação de Akaike* (p. 28).

Marchionni, V., Cabral, M., Amado, C., & Covas, D. (2016). Estimating Water Supply Infrastructure Cost Using Regression Techniques. *Journal of Water Resources Planning and Management*, 142(4), 04016003.

[https://doi.org/10.1061/\(asce\)wr.1943-5452.0000627](https://doi.org/10.1061/(asce)wr.1943-5452.0000627)

Marill, K. A. (2004). Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression. *Academic Emergency Medicine*, 11(1), 94–102.

<https://doi.org/10.1197/j.aem.2003.09.006>

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.).

Miguel, P., & Pires, V. (2017). *Controlo e Gestão de Custos de Produção – O Caso da Nestlé Portugal, S.A.* Instituto Politécnico de Coimbra.

Mileski, A. (2011). *Processos Produtivos* (p. 120).

Morgan, M. J., & Bork, H. P. (1993). Is ABC really a need, not an option?.

MANAGEMENT ACCOUNTING-LONDON, 71, 26–26.

Niazi, A., Dai, J. S., Balabani, S., & Seneviratne, L. (2006). Product cost estimation: Technique classification and methodology review. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 128(2), 563–575.

<https://doi.org/10.1115/1.2137750>

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>

Onyutha, C. (2020). From R-squared to coefficient of model accuracy for assessing goodness-of-fits; *Geoscientific Model Development Discussions*, May, 1–25.

Pak, S. II, & Oh, T. H. (2010). Correlation and simple linear regression. *Journal of Veterinary Clinics*, 27(4), 427–434. https://doi.org/10.1007/978-3-319-89993-0_6

Park, J.-H., Seo', K.-K., Wallace', D., & Lee3, K. 4. (n.d.). *Approximate Product Life*

Cycle Costing Method for the Conceptual Product Design.

- Paul, R. K. (2006). *Multicollinearity: causes, effects and remedies*. 4405, 14.
- Profillidis, V. A., & Botzoris, G. N. (2019). *Trend Projection and Time Series Methods*.
- Sayadi, A. R., Lashgari, A., & Paraszczak, J. J. (2012). Hard-rock LHD cost estimation using single and multiple regressions based on principal component analysis. *Tunnelling and Underground Space Technology*, 27(1), 133–141.
<https://doi.org/10.1016/j.tust.2011.08.006>
- Scott, A. J., Hosmer, D. W., & Lemeshow, S. (1991). Applied Logistic Regression. In *Biometrics* (Vol. 47, Issue 4). <https://doi.org/10.2307/2532419>
- Silva, V. L., Reis, E. A., & Amaral, G. D. Do. (2009). Análise de Regressão Linear no Pacote R. *File:///C:/Users/Administrador/Documents/AU/Biblio/AZAVEDO et Al 2008 - Principais Ecto e Endoparasitas Que Acometem Bovinos Leiteiros.Pdf*, 51.
<http://plutao.est.ufmg.br/arquivos/rts/RT-SE-2009.pdf>
- Smith, A., & Mason, A. K. (1997). Cost estimation predictive modeling: Regression versus neural network. *Engineering Economist*, 42(2), 137–161.
<https://doi.org/10.1080/00137919708903174>
- Taddy, M. (2014). Business Data Science Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. In *Paper Knowledge . Toward a Media History of Documents* (1st ed.).
- Wears, R., & Lewis, R. (1999). *Statistical Models and Occam's Razor*. 6(2), 93–94.
- Weustink, I. F., Ten Brinke, E., Streppel, A. H., & Kals, H. J. J. (2000). *A generic framework for cost estimation and cost control in product design*.

Anexo A

Anexo 1 - Valores do VIF para os modelos (5.4), (5.5), (5.6) e (5.7)

Modelo	VIF		
	Área	Perímetro	Altura média
(5.4)	2.61	2.83	1.4
(5.5)	Volume	Perímetro	Altura média
	2.5	4.88	3.17
(5.6)	Área	Volume	Altura média
	5.11	4.47	1.34
(5.7)	Massa	Volume	Altura média
	4.97	3.21	2.49

Fonte: o próprio autor.

Anexo 2 - Testes estatísticos para os modelos (5.4), (5.5), (5.6) e (5.7)

Modelo	Teste de Shapiro-Wilk		Teste de Breusch-Pagan		Teste de Durbin-Watson	
	<i>W</i>	<i>p-value</i>	<i>BP</i>	<i>p-value</i>	<i>D-W</i>	<i>p-value</i>
(5.4)	0.98318	0.6568	6.9319	0.0741	2.0804	0.778
(5.5)	0.90371	0.0004	6.1977	0.0451	1.9747	0.97
(5.6)	0.98228	0.6146	2.6768	0.4442	1.8711	0.652
(5.7)	0.98763	0.8556	2.6666	0.2636	1.9637	0.908

Fonte: o próprio autor.

Anexo 3 - Resíduos padronizados para os modelos (5.4), (5.5), (5.6) e (5.7)

Resíduos Padronizados						
Modelo	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
(5.4)	-2.594	-0.453	0.024	0.010	0.552	3.135
(5.5)	-3.004	-0.420	-0.022	-0.015	0.333	2.883
(5.6)	-2.437	-0.667	0.031	-0.007	0.428	2.393
(5.7)	-1.922	-0.524	0.184	0.173	0.936	2.669

Fonte: o próprio autor.