

Analítica Preditiva para Redução de Taxa de Rejeição de Placas de Fibra

Maria Rodrigues Fonseca Pires

Dissertação de Mestrado

Orientador na FEUP: Prof. Mário Amorim Lopes

Orientador no Instituto Kaizen: Eng. Miguel Mourão Teixeira



Mestrado Integrado em Engenharia e Gestão Industrial

2020-06-29

Resumo

Em unidades industriais de produção contínua é crucial ter um elevado nível de controlo sobre o processo. As variáveis envolvidas estabelecem relações entre si, influenciando o resultado obtido. Para aprofundar o conhecimento do processo, é necessário identificar as variáveis significativas e isolar o seu efeito na qualidade do produto final. O facto de haver pouca intervenção dos operadores acresce à importância de gerir todas as variáveis envolvidas.

A empresa alvo de estudo nesta dissertação tem como objetivo assegurar a qualidade dos produtos vendidos, a fim de consolidar a sua relação com os clientes. Para assegurar a obtenção de produtos que cumprem os limites de especificação, a organização recorre a uma estratégia de sobreconsumo de matéria prima, que se traduz em custos adicionais de produção.

O projeto de melhoria contínua desenvolvido pretende conferir aos operadores, que desempenham maioritariamente funções de controlo e monitorização, uma ferramenta de suporte à operação. O desafio foi reduzir a taxa de rejeição dos produtos, através da estabilização do processo de produção, e, simultaneamente, reduzir a taxa de sobreconsumo de matéria prima.

À data de arranque do projeto, a gestão da linha de produção baseava-se no conhecimento empírico dos trabalhadores da empresa. Recorrendo a uma abordagem estruturada, analisou-se, através de ferramentas estatísticas, o impacto das variáveis envolvidas no processo na qualidade do produto vendido aos clientes.

Tendo como objetivo reduzir a variabilidade dos resultados obtidos, desenvolveu-se um modelo que, face às condições da linha produtiva, permite prever o valor da resistência à tração no final do processo. Este permite detetar desvios relativamente ao valor objetivo e efetuar os ajustes necessários mais rapidamente.

Através de regressões lineares múltiplas, definiram-se os parâmetros com um impacto estatisticamente significativo na qualidade das placas. Após identificar as variáveis com um efeito significativo na qualidade do produto final, limitou-se a gama de valores aceitável para as mesmas.

Com o intuito de reduzir a taxa de rejeição e, em simultâneo, a de sobreconsumo, elaborou-se uma árvore de regressão como suporte de decisão para determinar o intervalo de valores ótimo relativamente à quantidade de matéria prima utilizada. Esta medida possibilitou restringir a variabilidade do processo, potenciando a sua otimização e autonomia.

A análise desenvolvida permitiu aumentar o nível de controlo sobre o processo, reduzindo as taxas de rejeição e sobreconsumo de matéria prima. As medidas propostas vão proporcionar um aumento da margem operacional da empresa.

Abstract

Predictive Analytics for Reduction of the Rejection Rate of Fiber Boards

In industrial units of continuous production having a high level of control is crucial. The variables involved in the process are related amongst themselves, which influences the result. In order to increase the knowledge of the process it is required to identify the significant variables and to isolate their effect on the quality of the final product. The fact that there is little intervention of operators raises the importance of managing all the variables involved.

The company being studied in this dissertation has the purpose of ensuring the quality of the products in the interest of consolidating its relationship with its customers. In order to guarantee that the products respect the specification limits, the organization uses a strategy of raw materials' overconsumption which results in additional production costs.

The goal of this continuous improvement project is to provide a tool to support the workers' operation who perform mainly activities of control and monitoring. The challenge was to reduce the rejection rate of the products through the stabilization of the production process and, at the same time, decrease the overconsumption rate.

When the project started, management of the production line was based on the empirical knowledge of the employees. Through a structured approach and using statistical tools, an analysis of the impact of the production variables on the quality of the product sold to clients was performed.

Setting the goal of reducing the variability of the obtained results, a model that according to the conditions of the production line can forecast the value of tensile strength by the end of production was developed. The model enables detection of deviations from the target value and to perform the required adjustments of parameters faster.

Through multiple linear regressions, the parameters with a statistically significant effect on the quality of the boards were defined. After identifying the variables with significant impact on the quality of the finished product the acceptable range of their values was restricted.

With the purpose of reducing the rejection rate and the overconsumption rate at the same time, a regression tree was performed in order to provide support for decision making regarding the optimal range for the amount of raw materials used. This allowed narrowing the variability of the process, enhancing its optimization and autonomy.

This analysis provided an increase in the overall process control while reducing rejection and overconsumption rates. The proposed measures will generate an increment in the company's operational margin.

Agradecimentos

Ao Instituto *Kaizen*, em particular, ao meu orientador Miguel Teixeira e ao meu líder de equipa Joel Queirós pela oportunidade de me incluírem neste projeto, por toda a orientação, motivação e exigência que tiveram para comigo.

À Beatriz Santos que me integrou no projeto e me ajudou em todo o processo, partilhando comigo muito do seu conhecimento. À Fátima Carneiro por me ajudar a manter o foco, pela sua disponibilidade e por todo o apoio e motivação que me deu durante os últimos meses.

Ao Professor Mário Amorim Lopes por todo o acompanhamento e orientação prestados.

Aos meus amigos que estão sempre presentes e disponíveis. Em particular, à Marta, ao Chico e ao Kiko por tudo o que partilhámos durante os últimos cinco anos. À Tété, à Maria, à Kika e à Xica por todo o seu carinho, apoio e amizade.

À minha família, especialmente aos meus pais pelo apoio incondicional e constante motivação e ao meu irmão, que desde pequena me desafia, sendo o principal responsável pela minha resiliência.

"It is not enough to do your best; you must know what to do and then do your best."

William Edwards Deming

Conteúdo

1	Introdução	1
1.1	Instituto Kaizen	1
1.2	Enquadramento do projeto e motivação	2
1.3	O setor dos derivados da madeira e a empresa	2
1.4	Objetivos do projeto	3
1.5	Metodologia adotada	4
1.6	Estrutura da dissertação	4
2	Enquadramento Teórico	6
2.1	Indústria de Processo	6
2.2	<i>Total Quality Management</i>	8
2.3	Variabilidade de um Processo	10
2.4	Six Sigma	11
2.4.1	DMAIC	12
2.4.2	Variabilidade do sistema de medição	13
2.4.3	Controlo Estatístico do Processo	14
2.5	Modelos Preditivos	15
2.5.1	Avaliar a qualidade do modelo	17
2.6	<i>Machine Learning</i>	19
3	Descrição do Problema	21
3.1	Descrição do processo	21
3.2	Seleção do tipo de produto	22
3.3	Controlo estatístico do processo	24
3.4	Seleção do produto	27
4	Métodos	30
4.1	Análise da variabilidade	30
4.1.1	Variabilidade do processo	31
4.1.2	Análise exploratória	32
4.2	Desenvolvimento dos modelos	34
4.2.1	Resistência à Tração	34
4.2.2	Densidade do Núcleo	35
4.2.3	Densidade da Placa	36
4.2.4	Modelo 1 nível	37
4.3	Limitações dos modelos	38
4.3.1	Coefficiente de determinação	38
4.3.2	Multicolinearidade	39
4.4	<i>Principal component analysis</i>	40

5	Resultados	41
5.1	Estabilização do processo produtivo	41
5.2	Impacto financeiro da implementação do modelo	42
6	Conclusão	46
6.1	Plano de ações	46
6.2	Perspetivas de trabalho futuro	46
6.2.1	<i>Improve</i>	46
6.2.2	<i>Control</i>	47
6.3	Conclusões	47
A	Simulação de dados para realizar teste ANOVA	53
B	Gráficos para validar os pressupostos referentes aos erros do teste ANOVA	56
C	Resultados do modelo da resistência à tração	57
C.1	Resultados do modelo	57
D	Resultados do modelo da densidade do núcleo	59
D.1	Resultados do modelo	59
E	Resultados do modelo da densidade da placa	61
E.1	Resultados do modelo	61
F	Diagrama de relações	63
G	Resultados e validação dos pressupostos do modelo de 1 Nível	64
G.1	Resultados do modelo	64
G.2	Validação de pressupostos	66
G.2.1	Normalidade multivariada	66
G.2.2	Influência multivariada dos outliers	66
G.2.3	Independência estatística	67
G.2.4	Multicolinearidade	68
G.2.5	Homocedasticidade	69
H	Resultados e validação dos pressupostos do modelo após PCA	70
H.1	Resultados do modelo	70
H.2	Validação de pressupostos	71
H.2.1	Normalidade multivariada	71
H.2.2	Influência multivariada dos outliers	72
H.2.3	Independência estatística	72
H.2.4	Multicolinearidade	73
H.2.5	Homocedasticidade	73

Acrónimos e Siglas

ANOVA	<i>Analysis of Variance</i>
B2B	<i>Business to Business</i>
C_p	Índice de Capacidade Potencial
C_{pk}	Índice de Capacidade Efetiva
CART	<i>Classification And Regression Trees</i>
CTQ	<i>Critical To Quality</i>
DMAIC	<i>Define - Measure - Analyse - Improve - Control</i>
DOE	<i>Design Of Experiments</i>
KBS	<i>Kaizen Business System</i>
LE	Limites de Especificação
LNT	Limites Naturais de Tolerância
MAPE	<i>Mean Absolute Percentage Error</i>
ML	<i>Machine Learning</i>
MP	Matéria Prima
OEE	<i>Overall Equipment Efficiency</i>
PCA	<i>Principal Component Analysis</i>
PDCA	<i>Plan - Do - Check - Act</i>
P_p	Índice de <i>Performance</i> Potencial
P_{pk}	Índice de <i>Performance</i> Efetiva
R&R	Reprodutibilidade e Repetibilidade
RMSE	<i>Root Mean Square Error</i>
RMSPE	<i>Root Mean Square Percentage Error</i>
RT	Resistência à Tração
SDCA	<i>Standardize - Do - Check - Act</i>
SPC	<i>Statistical Process Control</i>
TQM	<i>Total Quality Management</i>
VIF	<i>Variance Inflation Factor</i>
VOC	<i>Voice Of the Customer</i>
WIP	<i>Work In Progress</i>

Lista de Figuras

2.1	Diferenças entre indústria de montagem e indústria de processo. Adaptado de: Fransoo and Rutten (1993).	7
2.2	Pilares do KBS. Fonte: Kaizen Institute (2017).	8
2.3	Impacto dos ciclos PDCA e SDCA na obtenção de resultados. Fonte: Kaizen Institute (2017).	9
2.4	Melhoria de um processo. Fonte: Kaizen Institute (2017).	11
2.5	Carta de controlo. Fonte: Montgomery and Runger (2013).	14
3.1	Quantidade produzida entre Janeiro de 2019 e Março de 2020, por tipo de produto.	23
3.2	Variação dos valores por ordem de produção e tipo de produto.	23
3.3	Carta de controlo do processo para uma resistência à tração pretendida de 0,55.	25
3.4	Carta de controlo do processo para uma resistência à tração pretendida de 0,70.	26
3.5	Carta de controlo do processo para uma resistência à tração pretendida de 0,75.	27
3.6	Relação entre as três características de qualidade.	28
4.1	Diagrama de causa e efeito.	31
4.2	Direções longitudinal e transversal da placa e secção de corte para teste ANOVA.	32
4.3	Intervalos de confiança de 95% de Tukey relativamente à localização transversal da placa.	33
4.4	Diagrama de relações das variáveis que impactam a resistência à tração.	34
4.5	Diagrama de relações das variáveis que impactam a densidade do núcleo.	35
4.6	Diagrama de relações das variáveis que impactam a densidade da placa.	36
4.7	Diagrama de relações das variáveis do processo que impactam a resistência à tração da placa.	37
5.1	Árvore de regressão para determinar o valor da resistência à tração.	41
5.2	Variação da curva da distribuição normal de acordo com o valor esperado.	43
5.3	Taxas de sobreconsumo associadas à gama de valores definida.	44
5.4	Custos de rejeição de vários cenários.	44
B.1	Teste à normalidade.	56
B.2	Teste à homogeneidade da variância.	56
C.1	Equação da regressão linear múltipla - Resistência à tração.	57
C.2	Sumário do modelo - Resistência à tração.	57
C.3	Gráficos de resíduos - Resistência à tração.	58
C.4	Análise de variância - Resistência à tração.	58
D.1	Equação da regressão linear múltipla - Densidade do núcleo.	59

D.2	Sumário do modelo - Densidade do núcleo.	59
D.3	Gráficos de resíduos - Densidade do núcleo.	59
D.4	Análise de variância - Densidade do núcleo.	60
E.1	Equação da regressão linear múltipla - Densidade da placa.	61
E.2	Sumário do modelo - Densidade da placa.	61
E.3	Gráficos de resíduos - Densidade da placa.	61
E.4	Análise de variância - Densidade da placa.	62
F.1	Diagrama de relações com os três níveis.	63
G.1	Equação da regressão linear múltipla - Resistência à tração (1 nível).	64
G.2	Sumário do modelo - Resistência à tração (1 nível).	64
G.3	Gráficos de resíduos - Resistência à tração (1 nível).	65
G.4	Análise de variância - Resistência à tração (1 nível).	65
G.5	Gráfico normal de resíduos.	66
G.6	Tabela de observações atípicas.	67
G.7	Estatística de <i>Durbin-Watson</i>	67
G.8	Coeficientes do modelo obtido.	68
G.9	Gráfico de resíduos versus ajustes.	69
H.1	Equação da regressão linear múltipla - Resistência à tração (após PCA).	70
H.2	Sumário do modelo - Resistência à tração (após PCA).	70
H.3	Gráficos de resíduos - Resistência à tração (após PCA).	70
H.4	Análise de variância - Resistência à tração (após PCA).	71
H.5	Gráfico normal de resíduos.	71
H.6	Tabela de observações atípicas.	72
H.7	Estatística de <i>Durbin-Watson</i>	72
H.8	Coeficientes do modelo obtido.	73
H.9	Gráfico de resíduos versus ajustes.	73

Lista de Tabelas

3.1	Classificação ABC dos produtos de acordo com a densidade pretendida.	28
4.1	Divisão das variáveis em níveis pela sua gama de valores.	39
4.2	Coeficiente de determinação do modelo com valores dos 3 níveis.	39
4.3	Coeficiente de determinação do modelo com valores do nível baixo.	39
5.1	Custos adicionais de produção.	42
A.1	Dados utilizados na simulação do teste ANOVA.	53

Capítulo 1

Introdução

A presente dissertação, inserida no âmbito do Mestrado Integrado em Engenharia de Gestão Industrial, enquadra-se num projeto de melhoria contínua liderado pelo *Kaizen Institute* numa empresa industrial de produção de placas de fibra. Pretende-se com este projeto o desenvolvimento de um modelo preditivo que, através da análise de características da linha de produção, permita prever o valor de uma característica essencial das placas – a resistência à tração.

1.1 Instituto Kaizen

O *Kaizen Institute Consulting Group* é uma empresa que oferece serviços de consultoria operacional e formação em todos os setores da economia, em mais de 60 países, através da implementação de metodologias *Lean*.

De origem japonesa, *Kaizen* resulta da junção de dois vocábulos: “*Kai*”, que significa “mudar”, e “*Zen*” “melhor”, podendo esta palavra ser traduzida para “mudar para melhor” ou “melhoria contínua”. Masaaki Imai (1986), o fundador, explica que “Quando aplicado no local de trabalho, *Kaizen* significa melhoria contínua envolvendo todos – administração e restantes colaboradores”. Atualmente, a implementação de metodologias *Kaizen* é reconhecida como um importante pilar da estratégia competitiva de organizações. O foco do *Kaizen* é dotar os seus clientes de ferramentas de cultura e liderança que assegurem a sua vantagem competitiva e que lhes permitam tornar-se autónomos na implementação de melhoria contínua, contribuindo de forma decisiva para a melhoria do seu desempenho.

Os 5 princípios do *Kaizen* são:

1. Criar valor para o cliente;
2. Criar eficiência de fluxo;
3. Ser orientado para o *gemba*;
4. Capacitar pessoas;
5. Ser científico e transparente;

Seguindo estes 5 princípios, o objetivo *Kaizen* é melhorar a qualidade, a produtividade e a gestão do trabalho dos seus clientes, ou seja, melhorar os seus resultados.

1.2 Enquadramento do projeto e motivação

A empresa em que foi desenvolvido o projeto (Empresa A¹) dedica-se à produção de placas de aglomerado numa prensa industrial de larga escala, através de um processo contínuo. É crucial monitorizar o estado do processo produtivo para assegurar a consistência da qualidade dos produtos finais (Roggo et al., 2020).

De acordo com a gama de produtos e a espessura da placa, a empresa determinou um valor objetivo para a resistência à tração. Após a produção de um lote, são testadas, em laboratório, algumas amostras para avaliar se esse lote está apto para ser vendido aos clientes.

Se os valores medidos em laboratório forem inferiores aos parâmetros definidos pela empresa, as placas não estão aptas para serem vendidas. Caso sejam muito superiores aos parâmetros definidos pela empresa, significa que foi utilizada uma maior quantidade de matéria-prima do que necessário. Ambas as situações representam uma redução da rentabilidade da empresa. Logo, uma diminuição na variabilidade do processo vai permitir reduzir a taxa de desqualificação das placas produzidas ao mesmo tempo que minimiza o consumo de matéria prima e, assim, aumentar o retorno da empresa.

Atualmente, a empresa não tem capacidade para, através da análise das características da linha de produção, prever a qualidade da placa final. Assim, para assegurar que se obtém um produto com a qualidade pretendida à primeira, aumentou-se o consumo de matéria-prima e, consequentemente, os custos de produção.

O projeto assegura acrescentar valor para a organização, na medida em que irá proporcionar um aumento da margem operacional. O âmbito restringe-se à afinação das condições de produção, a fim de garantir qualidade à primeira com o menor consumo de reagentes possível. Nesse sentido, após uma análise aprofundada do processo, constituído por centenas de variáveis de controlo, isolaram-se as que tinham maior influência no valor de resistência à tração da placa finalizada e definiu-se uma metodologia para controlar a variabilidade do processo, a fim de reduzir o problema de qualidade da empresa usando uma abordagem normalizada.

1.3 O setor dos derivados da madeira e a empresa

A Empresa A ocupa uma posição de destaque na indústria do setor dos derivados da madeira (madeira, cortiça e papel), em Portugal e globalmente. Destaca-se pelo seu vasto portfólio de produtos, desde a área da construção até à área da decoração, e pelo seu foco e constante investimento em inovação e sustentabilidade.

O setor da madeira, cortiça e papel inclui várias atividades, desde o corte e abate das árvores até à sua transformação e comercialização (AIMMP, 2020).

Em Portugal, 2% das empresas operam neste setor, representando 3% do volume de negócios a nível nacional e do número de pessoas empregadas. A maioria - mais de 80% - são microempresas.

¹Por motivos de confidencialidade, vai ser utilizada esta terminologia para referir a empresa onde decorreu o processo.

No entanto, as de maior dimensão (grupo ao qual pertence a empresa onde foi realizado o projeto da presente dissertação) agregam, sensivelmente, metade do volume de negócios (Portugal, 2016). A crise económica que Portugal enfrentou a partir de 2007 teve repercussões no volume de vendas do setor, afetando também a empresa em estudo. Para contrariar esse cenário, a Empresa A procurou ganhar poder competitivo através da diferenciação. Para tal, inovou no tipo de produtos que oferecia, reduziu o tempo de entrega e apostou num aumento da qualidade do produto final. A Empresa A leva a cabo atividades com as características de uma indústria de processo. As indústrias de processo contínuo correspondem à vanguarda do processo industrial (Toledo et al., 1986). As matérias primas sofrem transformações físicas e químicas ao longo de um processo contínuo e sequencial, sem intervenção direta de mão de obra. Os operadores desempenham funções de controlo e monitorização dos equipamentos e do funcionamento da linha produtiva. Atualmente, a empresa está focada em conseguir reduzir a taxa de desqualificação do produto final, assegurando que se atingem os valores objetivo das características de qualidade à primeira. Deste modo, o projeto desenvolvido teve como foco garantir a qualidade das placas.

1.4 Objetivos do projeto

Utilizou-se como critério de escolha a capacidade do processo para determinar qual a família de produtos a estudar na presente dissertação. Elegeram-se placas do tipo B, com uma densidade pretendida de 700 kg/m^3 e um valor objetivo de resistência à tração de 0,70Pa.

Relativamente a esta família de produtos, em 2019, a empresa A teve taxas de rejeição (requalificação e desqualificação) inferiores a 1%. Uma vez que a organização já tem um nível de maturidade elevado, é necessária uma abordagem rigorosa para melhorar a *performance* produtiva, pelo que se optou por adotar medidas de *Six Sigma* para melhorar a qualidade.

A capacidade de prever o valor da resistência à tração através de variáveis da linha de produção permite detetar desvios mais rápido e aferir se é necessário ajustar a produção, resultando em menor desperdício de capacidade produtiva, matéria prima e energia. Pretende-se obter um modelo que possa ser utilizado pelos operadores como uma ferramenta de controlo da qualidade durante o processo, avaliando se a placa vai cumprir ou não as especificações acordadas com o cliente e garantindo qualidade à primeira.

O principal objetivo deste projeto é reduzir a taxa de rejeição para que tenda para 0%. Traçou-se a meta de, pelo menos, reduzir em 50% o valor da taxa do ano anterior. De forma a assegurar tal propósito, colocam-se as seguintes questões de investigação:

- Quais as principais variáveis que influenciam a variável resposta?
- Quais os limites aceitáveis para a variabilidade das variáveis *input*? (Definição de valores mínimos e máximos para algumas variáveis para reduzir a variabilidade da variável resposta)
- Quais os limites aceitáveis para a variabilidade da variável resposta?
- Como controlar e normalizar o sistema de medição de modo a reduzir a sua variabilidade?

Pretende-se, face aos valores de resistência à tração pretendidos, ter a capacidade de determinar valores-alvo para as variáveis-chave do processo, a fim de aumentar o nível de controlo do processo, reduzir a sua variabilidade e assegurar a qualidade do produto final.

1.5 Metodologia adotada

O projeto focou-se na utilização de metodologias *Six Sigma* para melhoria de qualidade e no desenvolvimento e implementação de modelos preditivos de modo a reduzir variabilidade dos resultados obtidos no processo e assegurar a qualidade final do produto.

Numa fase inicial, realizou-se a caracterização da situação atual e selecionou-se a família de produtos que ia ser alvo de estudo. Consistiu num estudo no terreno para compreender o processo produtivo e todas as variáveis envolvidas (de *input*, *output* e controlo).

Realizou-se um *workshop* no qual vários colaboradores da Empresa A partilharam o seu conhecimento empírico acerca do processo. Apesar de não saberem quais as variáveis relacionadas diretamente com a qualidade da placa, foram levantadas várias questões que permitiram estruturar a abordagem a seguir, nomeadamente no que concerne ao tipo de dados a serem recolhidos. Traçou-se um diagrama de relações, decomposto em três níveis, entre as variáveis envolvidas no processo. A elaboração dos modelos baseou-se em dados recolhidos entre Janeiro de 2019 e Março de 2020.

A fim de identificar quais as variáveis envolvidas no processo que eram estatisticamente significativas para a qualidade do produto final, desenvolveram-se modelos através de regressões lineares múltiplas. Para a análise dos dados e a elaboração dos modelos, recorreu-se ao *software* estatístico Minitab.

Após a obtenção dos modelos, uma vez que não foi possível realizar desenho de experiências, dadas as circunstâncias em que o projeto decorreu², utilizou-se o *software RStudio* para elaborar árvores de decisão, de modo a determinar a amplitude aceitável para as variáveis da linha.

1.6 Estrutura da dissertação

A dissertação divide-se em seis capítulos.

No primeiro capítulo é feita uma breve introdução do projeto desenvolvido, o seu âmbito e a sua relevância para a empresa. Apresenta-se o Instituto *Kaizen* e descreve-se o setor em que se insere a empresa que foi alvo de estudo. Por último, definem-se os objetivos a alcançar com o projeto e qual a metodologia adotada para o conseguir.

O capítulo dois corresponde ao enquadramento teórico, onde se apresenta os fundamentos teóricos que sustentam as decisões tomadas ao longo do projeto. Abordam-se as metodologias utilizadas e explica-se qual a finalidade e o método de implementação das ferramentas usadas.

²No dia 18 de Março de 2020, foi decretado o estado de emergência em Portugal, através do Decreto do Presidente da República n.º 14-A/2020, devido à proliferação do vírus de COVID-19.

No capítulo seguinte, descreve-se com maior grau de detalhe o problema e o estado inicial (*AS IS*). É apresentado o processo produtivo e o racional relativo à seleção da família de produtos analisada na presente dissertação.

O quarto capítulo descreve a análise estatística realizada ao processo e a identificação das variáveis significativas.

No quinto capítulo, pretende-se determinar o estado alvo (*TO BE*) e o impacto financeiro da implementação das soluções propostas.

No capítulo seis, traçou-se um plano de ações a implementar quando for possível retomar o projeto presencialmente. Por último, listaram-se as conclusões do projeto e propostas de melhoria para possível implementação futura.

Capítulo 2

Enquadramento Teórico

Em qualquer setor, o principal objetivo de um fornecedor é oferecer um serviço ou produto de qualidade recorrendo a uma abordagem que lhe permita ser o mais eficiente possível. O método mais comum para atingir este objetivo é identificar e melhorar os elementos do negócio mais críticos para o cliente (Johnston et al., 2008). Tal pode ser feito através de algumas metodologias como *Total Quality Management*, *Six Sigma* e *Lean Manufacturing*.

2.1 Indústria de Processo

Na literatura, os processos de produção industrial são divididos em duas subcategorias: indústria de processo e indústria de montagem ou montagem.

Fransoo and Rutten (1993) partilharam a definição da APICS (*American Production and Inventory Control Society*) de indústria de processo:

“As indústrias de processo são negócios que acrescentam valor às matérias primas misturando, separando, formando ou através de reações químicas. Os processos podem ser contínuos ou em lotes e, geralmente, requerem um controlo do processo rígido e elevado investimento de capital.”

Numa indústria de processo não é possível distinguir as matérias primas no produto final devido ao conjunto de reações químicas e físicas a que são submetidas durante o processo produtivo. No entanto, a seleção das matérias primas é essencial, dado que vai influenciar a qualidade do produto final, representando um fator de variabilidade do processo. A sequência do processo requer uma forte interligação entre os equipamentos, uma vez que as transformações estão interligadas e são interdependentes.

A produção em lotes deve-se à necessidade de minimização de *setups*, pelo que se torna mais vantajoso definir períodos com uma duração fixa para produzir cada tipo de produto (Kadlec et al., 2009).

Os custos mais significativos são o investimento em equipamento. Os gastos relativos à mão de obra podem ser considerados fixos, uma vez que o volume de trabalhadores não se altera em

função do volume de produção. Logo, o objetivo é ocupar ao máximo a capacidade das máquinas (Borges and Dalcol, 2002).

O processo envolve pouca mão de obra, apenas em atividades de suporte, manutenção e monitorização da linha. Por este motivo, a produtividade está relacionada com o desempenho dos equipamentos e não com o ritmo de trabalho da mão de obra.

O estado da operação pode ser monitorizado através de sensores presentes nos equipamentos. A capacidade de centralizar o controlo do processo permite detetar desvios de parâmetros mais rápido e proceder à sua correção. Assim, neste ramo, a automação é de elevada importância, já que se traduz num aumento da produtividade.

Uma das principais características diferenciadoras entre indústria de processo e indústria de montagem é a relação entre a quantidade de diferentes matérias primas utilizadas e a variedade de produtos finais obtidos. Na primeira, vários produtos são obtidos a partir de poucos tipos de matéria prima (processo divergente). Em indústria de montagem, os componentes iniciais são mais diversos que os finais (processo convergente), como se pode observar na Figura 2.1.

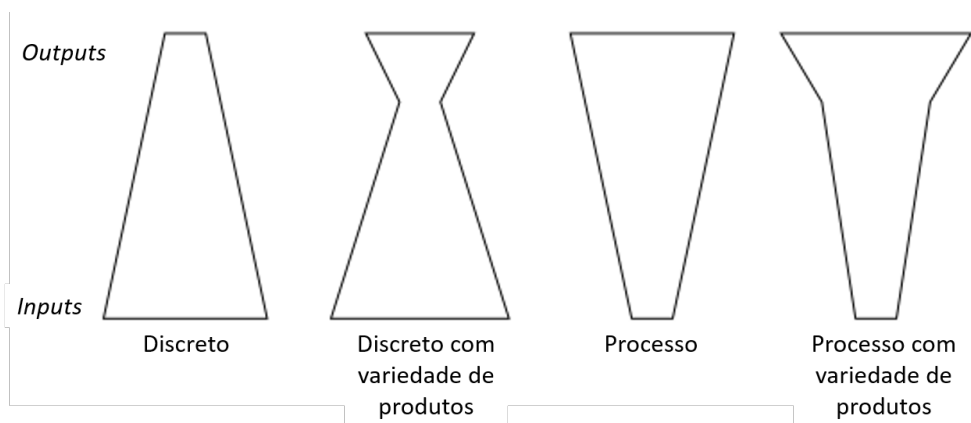


Figura 2.1: Diferenças entre indústria de montagem e indústria de processo. Adaptado de: Fransoo and Rutten (1993).

De um modo geral, os produtos provenientes da indústria de processo são mais diferenciados. Já os originados através de montagem e montagem apresentam menor diferenciação.

O tipo de indústria reflete-se, também, no ambiente de produção. Na indústria de montagem, funcionam como *Job Shop*, ou seja, os materiais deslocam-se ao longo da fábrica de acordo com o tipo de trabalho que está ser levado a cabo. No *Flow Shop*, os materiais seguem um trajeto fixo (indústria de processo). Os equipamentos utilizados também são uma característica que permite distinguir o tipo de indústria. Em montagem, utilizam-se máquinas mais flexíveis, que permitam obter uma maior variedade de produtos. Num processo contínuo, os equipamentos são mais especializados e com um alto grau de automação, a sua função é produzir produtos semelhantes, em grandes quantidades e a baixos custos (Borges and Dalcol, 2002).

As diferenças entre indústria de processo e indústria de montagem traduzem-se em várias vertentes, sendo de destacar os fatores competitivos. Na primeira, que geralmente se trata de um

negócio *Business to Business* (B2B) e cujos produtos finais são matérias-primas para outra indústria, o principal foco é o preço de venda. Em contrapartida, numa indústria de montagem, a inovação, a diferenciação e a customização são os elementos mais diferenciadores.

2.2 Total Quality Management

O *Kaizen Business System* (KBS) é um modelo que procura a excelência operacional, auxiliando as organizações a tornarem-se mais competitivas e rentáveis, através de melhorias em todas as áreas do negócio. Divide-se em cinco pilares compostos por ferramentas que permitem melhorar a qualidade, a eficiência e o serviço de entrega (Figura 2.2).



Figura 2.2: Pilares do KBS. Fonte: Kaizen Institute (2017).

A implementação do KBS ajuda as empresas a pensar de um modo orientado aos processos e a desenvolver estratégias que asseguram melhoria contínua envolvendo todas as pessoas em todos os níveis da organização. Estar orientado aos processos significa confirmar com os resultados, ou seja, as melhorias implementadas vão refletir-se nos resultados, mas não serão despoletadas pelos resultados.

O foco deste modelo são todas as áreas que podem ser melhoradas. Não apenas os serviços e produtos oferecidos pela empresa, mas a forma como as pessoas trabalham, o modo como os equipamentos são operados ou como se lida com os sistemas e procedimentos.

Um dos pilares do KBS é a gestão da qualidade total (*Total Quality Management* - TQM), que se trata de uma abordagem holística para melhorar a qualidade. Esta engloba vários níveis: desde a criação de *standards* e normas até à implementação de ferramentas de *Six Sigma*. Implementar metodologias TQM vai contribuir para um aumento da satisfação dos clientes e do sucesso do negócio (Imai, 2012).

Tradicionalmente, qualidade pode ser definida como “estar de acordo com as especificações”. No entanto, o foco deve ser em que o processo seja o mais “*on-target*” possível, reduzindo a variabilidade ao máximo, a fim de minimizar os custos (Gygi et al., 2005).

De acordo com os *standards* industriais japoneses (Japanese Industrial Standards – JIS Z8101-1981), o controlo de qualidade é “um sistema com meios para produzir bens ou serviços que satisfaçam os requisitos dos clientes” (Imai, 1986).

A qualidade pode ser dividida em duas vertentes: a qualidade dos artigos, isto é, o que o cliente vê, sente e mede, e a qualidade dos processos, ou seja, eliminação do desperdício (*Muda*), da variabilidade (*Mura*) e da sobrecarga (*Muri*).

Há três regras na implementação de projetos de melhoria da qualidade:

- Não se pode melhorar aquilo que não se controla;
- Não se pode controlar aquilo que não se mede;
- Não se pode medir aquilo que não se define.

Masaaki Imai (1986) defende que se devem intercalar ciclos PDCA (*Plan – Do – Check – Act*) com ciclos SDCA (*Standardize – Do – Check – Act*), uma vez que estes são complementares. O primeiro gera melhorias que vão ser estabilizadas e sustentadas pelo segundo, como se pode observar na Figura 2.3.

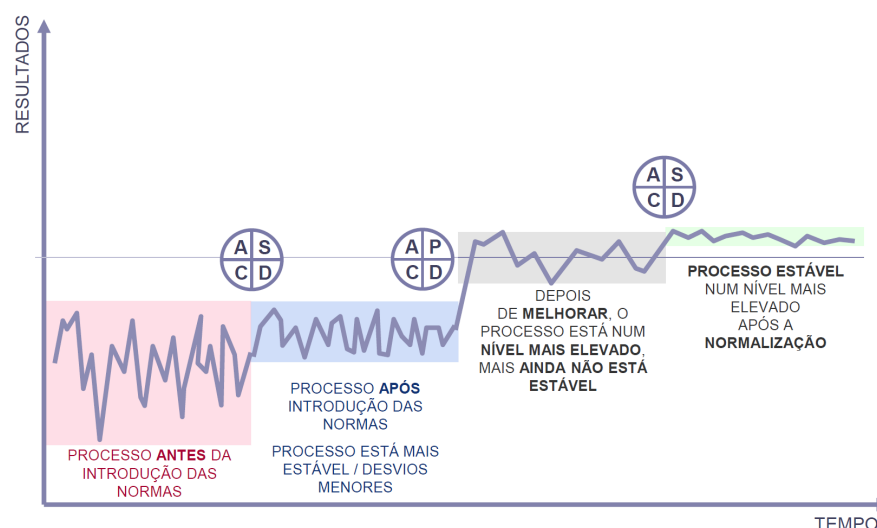


Figura 2.3: Impacto dos ciclos PDCA e SDCA na obtenção de resultados. Fonte: Kaizen Institute (2017).

A criação de normas/*standards* é, frequentemente, o primeiro passo da melhoria. Um *standard* é a forma mais eficiente de realizar uma tarefa. As normas garantem a estabilidade do processo, a prevenção de erros repetitivos e a polivalência da equipa (Kaizen Institute, 2017).

Um dos principais elementos do TQM é a gestão do processo, que passa por melhorar a qualidade do produto durante o processo produtivo. Em paralelo, a gestão da liderança permite incutir nos trabalhadores uma cultura direcionada à qualidade, conferindo às organizações uma vantagem competitiva. A tomada de decisões com base numa abordagem factual é outra vertente do TQM com um impacto positivo na qualidade. Zehir et al. (2012) concluíram que a implementação de metodologias de TQM tem um impacto significativo na qualidade.

Um problema pode ser definido como o desvio entre a situação atual e a desejada. Numa primeira fase, define-se o problema a resolver. De seguida, investiga-se a origem desse problema para que possa ser solucionado na terceira etapa.

Atualmente, as organizações focam-se cada vez mais em processos e metodologias de controlo e melhoria da qualidade porque esse investimento é muito vantajoso face aos custos de não qualidade. Para além de, tal como referido anteriormente, permitir ter uma melhor relação com o cliente.

2.3 Variabilidade de um Processo

A variabilidade pode ser caracterizada como o desvio face à expectativa (valor alvo) e tem um impacto significativo no consumidor. Como defende Montgomery (2009), a qualidade é inversamente proporcional à variabilidade. Um melhor controlo da qualidade vai permitir uma redução dos custos, devendo estar a cargo das pessoas que desempenham cargos de liderança (Deming, 2000).

Para poder controlar a variabilidade de um processo é necessário identificar a sua origem (causa-efeito). Pelo facto de a variabilidade só poder ser descrita em termos estatísticos, os métodos estatísticos de controlo de um processo revelam-se particularmente relevantes. A presença de variabilidade num processo industrial tem um impacto significativo na monitorização, através de métodos estatísticos, do mesmo (Gins et al., 2013).

A variabilidade de um processo pode ser dividida em duas categorias:

- Variabilidade controlada: proveniente de causas naturais ou comuns, ou seja, inerente ao processo. É passível de ser atenuada, mas é impossível eliminá-la;
- Variabilidade descontrolada: proveniente de causas especiais/assinaláveis. Estas devem ser identificadas e eliminadas.

O principal objetivo de engenharia da qualidade é a redução da variabilidade das características CTQ. É crucial identificar as fontes de variabilidade para poder minimizá-la (Aguayo, 2010).

Numa primeira fase, deve-se resolver as causas assinaláveis de variabilidade, uma vez que são fonte de instabilidade para o processo. Geralmente, passa por eliminar um problema em específico. Quando o processo já está controlado, pode ser melhorado através da implementação de estratégias para diminuir o impacto das causas naturais de variabilidade. No entanto, remover uma causa comum de variabilidade normalmente significa alterar o processo. A eliminação da variabilidade vai permitir obter um processo mais uniforme no que toca à qualidade do produto final (Shewhart, 1931).

A variabilidade de um processo afeta economicamente os resultados de uma empresa, traduzindo-se em qualidade mais baixa e insatisfação dos clientes (Aba, 2013). A atenuação da variabilidade de um processo permite reduzir custos, incrementar a produtividade e melhorar a qualidade (Shahriari and Ding, 2011).

2.4 Six Sigma

Six Sigma é uma metodologia associada à gestão da qualidade que visa melhorar os negócios e o desempenho das empresas. É um método inovador de gestão de qualidade (Smętkowska and Mrugalska, 2018). Surgiu na indústria eletrônica (Motorola & Texas Instruments), mas, ao longo das últimas três décadas, foi introduzido em várias outras áreas, como indústrias de serviço, hospitais e setor público.

Surgiu como uma metodologia estruturada do TQM que permite desenhar e melhorar processos, recorrendo a ferramentas estatísticas. A principal motivação para o desenvolvimento do *Six Sigma* foi a necessidade de melhorar a qualidade em contextos de produção complexos, com elevado número de componentes, resultando numa alta probabilidade de obter produtos com defeitos (Arnheiter and Maleyeff, 2005).

Trata-se de um *framework* que assegura o envolvimento de várias áreas, de forma bem-sucedida, para obter qualidade, sendo utilizado para encontrar e eliminar as causas raiz dos problemas. Identifica as variáveis chave de um processo e controla o seu valor para assegurar que o *output* do processo também estará sob controlo. Assim, permite reduzir a variabilidade dos resultados obtidos, de modo que os limites de especificação distem, pelo menos, 6 desvios padrão (σ) da média (Montgomery, 2009).

A melhoria de um processo passa por torná-lo aceitável, ou seja, obter resultados que estejam dentro dos limites de especificação e, de seguida, controlá-lo, isto é, reduzir a sua variabilidade (Figura 2.4).

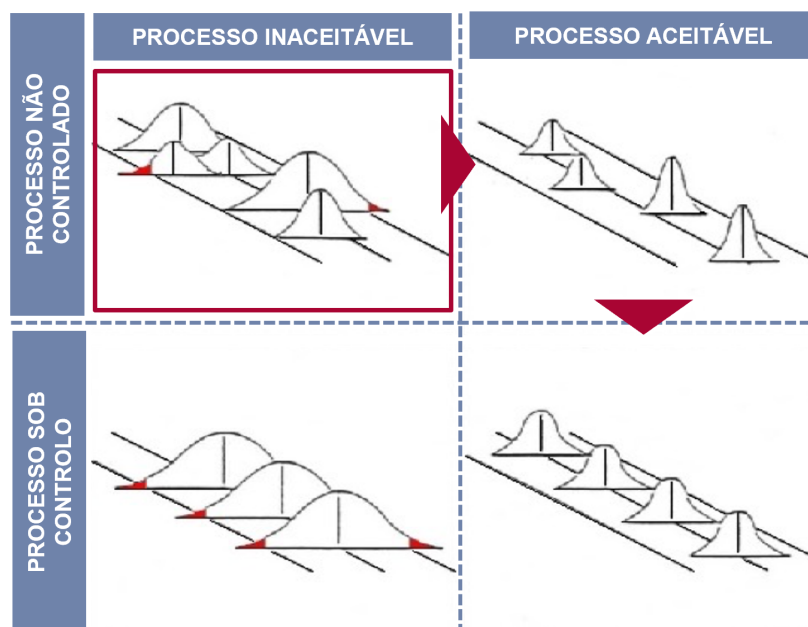


Figura 2.4: Melhoria de um processo. Fonte: Kaizen Institute (2017).

O foco no consumidor é uma das características fundamentais desta metodologia. Os defeitos são

definidos como qualquer tentativa falhada de corresponder às expectativas dos clientes. Estas são definidas como requisitos CTQ (*Critical To Quality*), *i.e.*, as características que os consumidores mais valorizam em termos de impacto na qualidade (Green, 2006).

A escala *sigma* é utilizada para aferir se a característica de qualidade cumpre os requisitos. Quanto maior o valor de *sigma*, maior a capacidade da característica (Gygi et al., 2005). Num processo que opere ao nível *Six Sigma*, o risco de uma característica individual estar incorreta é muito baixo, mais especificamente, 3,4 defeitos por milhão.

A adoção de metodologias *Six Sigma* permite que os trabalhadores controlem melhor o processo produtivo, em particular as tarefas que desempenham. Consequentemente, a sua motivação aumenta e o número de defeitos diminui, já que estão mais capazes para resolver problemas.

Além disso, através da análise de dados anteriores, é possível criar modelos preditivos para corrigir problemas identificados no processo. Assim, representa uma estratégia que permite aumentar a rentabilidade de uma empresa, na medida em que os defeitos são detetados mais cedo, prevenindo gastos em retrabalho (Tjahjono et al., 2010).

2.4.1 DMAIC

O DMAIC é uma ferramenta utilizada na maioria das implementações de *Six Sigma*, com vista à gestão de projetos e melhoria de processos. As letras do acrónimo correspondem às cinco fases do processo: *Define – Measure – Analyse – Improve – Control*.

Na primeira etapa (D) o objetivo é definir o âmbito e as oportunidades. Deve-se acordar o propósito do projeto, identificar os clientes e registar os seus requerimentos (CTQ), mapear os processos, e definir a equipa que vai participar no projeto. É essencial garantir que o trabalho a desenvolver está alinhado com as prioridades da empresa, garantindo o apoio dos líderes e o acesso aos recursos necessários. Deve-se estimar o impacto que o projeto vai ter (face aos custos atuais) e assegurar que está de acordo com as expectativas dos líderes (George et al., 2005).

Na fase seguinte (M) pretende-se quantificar o problema. Para tal, é efetuada a recolha de dados, necessários para compreender o processo, e são determinadas as métricas e os sistemas de medição a adotar. A fim de assegurar a qualidade dos resultados obtidos, é fundamental avaliar a qualidade do sistema de medição.

No terceiro passo (A) procede-se à análise do processo. Nesta etapa, estudam-se os dados recolhidos anteriormente e procura-se encontrar as causas dos problemas identificados. Define-se a capacidade do processo, que expressa a capacidade de satisfazer as especificações do cliente, e são identificadas oportunidades de melhoria, através da comparação do estado atual e do estado alvo.

Na penúltima fase (I) pretende-se melhorar o desempenho. São geradas e quantificadas potenciais soluções para os problemas identificados. Estas são avaliadas e seleciona-se a solução ou conjunto de soluções que se vai implementar. Inicialmente, a solução deve ser testada num grupo piloto para confirmar a sua eficácia e permitir efetuar correções antes da implementação a larga escala.

A última etapa (C) consiste em controlar a *performance* para assegurar que os resultados obtidos são mantidos. As fases mais críticas do processo são monitorizadas continuamente, criam-se

normas, desenvolvem-se metodologias anti-erro e elabora-se um plano de controlo a longo prazo (Montgomery, 2009). Pretende-se verificar que as soluções implementadas na fase anterior são suficientes e que não há desvios face ao plano traçado (Smętkowska and Mrugalska, 2018).

Esta metodologia, utilizada no decorrer de todo o projeto, permite estruturar a abordagem a problemas, uma vez que só se deve passar à etapa seguinte depois de terminada a que está a decorrer.

2.4.2 Variabilidade do sistema de medição

Na segunda etapa do DMAIC, *Measure*, devem-se realizar os testes R&R, isto é, testes de repetibilidade e reprodutibilidade. A análise destas duas características permite quantificar a variabilidade com origem no sistema de medição, como representado na Equação 2.1 (Mgbere et al., 2015).

$$\sigma_{sm}^2 = \sigma_{repe}^2 + \sigma_{repro}^2 \quad (2.1)$$

Em que:

σ_{repe}^2 representa a variabilidade devido à repetibilidade;

σ_{repro}^2 representa a variabilidade devido à reprodutibilidade.

O objetivo é determinar quanta da variabilidade observada se deve ao sistema de medição adotado, isolar as componentes de variabilidade do sistema de medição e avaliar se o equipamento utilizado é capaz, ou seja, se consegue cumprir o que se pretende (Montgomery, 2009).

A repetibilidade de um sistema de medição afere a consistência da sua variabilidade. O mesmo operador, utilizando o mesmo equipamento de medição, vai medir várias vezes a mesma característica do produto. A consistência dos resultados obtidos traduz a repetibilidade.

A reprodutibilidade de um sistema de medição traduz a capacidade de diferentes operadores obterem resultados consistentes. Para avaliar este parâmetro é crucial manter todos os restantes constantes, ou seja, os operadores devem utilizar o mesmo equipamento de medição e medir os mesmos objetos. Pode ser estimada comparando a média dos valores medidos por cada um dos operadores (Pyzdek, 2003).

Quando o processo de medição danifica a amostra (por exemplo, num teste de resistência à tração, o provete é submetido a forças de tração até partir), inviabilizando a utilização da mesma amostra para realizar os testes R&R. Logo, tem de se realizar um teste R&R limitado.

Neste tipo de testes, uma vez que não se pode usar a mesma amostra duas vezes, é necessário utilizar amostras o mais semelhantes possível. Assim, é essencial garantir que as amostras provenientes do mesmo lote são homogéneas, isto é, que podem ser tratadas como se fossem a mesma. Se houver partes suficientes de cada lote para todos os operadores medirem, pode-se realizar o teste R&R da forma tradicional: cruzado. Caso não seja possível, tem de se recorrer a um teste R&R limitado (Excel, 2013).

2.4.3 Controlo Estatístico do Processo

Um processo diz-se “sob controlo” quando apenas é afetado pela variabilidade inerente ao processo – causas naturais. No entanto, há outras causas de variabilidade que podem influenciar o *output* de um processo, como má calibração dos equipamentos, erros dos operadores ou falta de qualidade da matéria prima.

Um dos principais objetivos do controlo estatístico de um processo (*Statistical Process Control* - SPC) é detetar rapidamente a ocorrência de causas assinaláveis de variabilidade e proceder à sua correção, para evitar a produção de unidades não conformes. Pode dividir-se em três fases:

- Identificar os fatores que contribuem para a variabilidade do processo;
- Analisar estes fatores para eliminar causas de anormalidade;
- Controlar estes fatores para estabilizar o processo.

A melhoria da qualidade, para além de pretender eliminar os defeitos, passa também por eliminar os desperdícios, tais como testes, inspeções e os custos e o tempo gasto em corrigir produtos com defeitos. Os métodos estatísticos são cruciais para medir, monitorizar, controlar e melhorar a qualidade.

As cartas de controlo são uma ferramenta de SPC que permite compreender a variabilidade de um processo e determinar, de um ponto de vista estatístico, se este está controlado. Permitem também avaliar a capacidade de um processo ter resultados que estejam de acordo com as especificações do cliente (*Voice Of the Customer* - VOC). São, deste modo, relevantes enquanto fonte de informação para melhoria de processos.

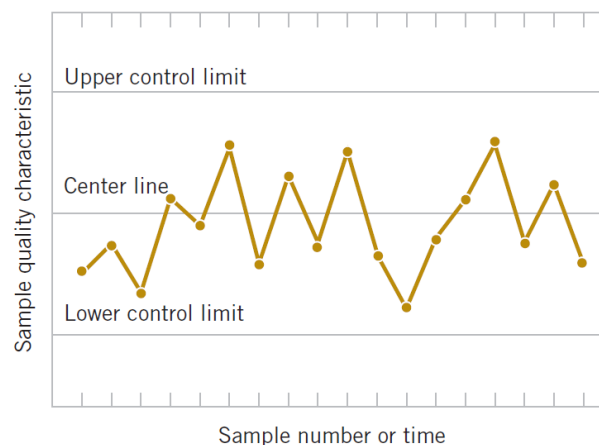


Figura 2.5: Carta de controlo. Fonte: Montgomery and Runger (2013).

Se, como na Figura 2.5, os pontos do gráfico se encontrarem entre os limites de controlo definidos e seguirem um padrão aleatório, é possível concluir que não há evidências estatísticas de que o processo esteja fora de controlo.

Caso o processo não esteja sob controle, significa que está sujeito a causas assinaláveis de variabilidade. Deve-se identificá-las, determinar as suas causas raiz e eliminá-las, através da implementação de ações corretivas. Este método permite reduzir a variabilidade e, conseqüentemente, melhorar o processo (Montgomery and Runger, 2013).

A capacidade de um processo traduz a sua aptidão para produzir produtos conforme as especificações do cliente. Antes de avaliar a capacidade de um processo, é necessário assegurar que está estável. Um processo fora de controle é imprevisível. No entanto, se estiver controlado, é possível prever o seu desempenho e melhorar a sua capacidade.

A capacidade de um processo é aferida através de 2 índices: C_p (Índice de Capacidade) e P_p (Índice de *Performance*). O primeiro caracteriza a capacidade do processo a curto prazo, o segundo fá-lo a longo prazo. Os índices de capacidade de um processo são utilizados para medir a *performance* de um processo e o seu potencial (Chen et al., 2017).

Considera-se que os Limites Naturais de Tolerância (LNT) de um processo são 3σ acima e abaixo da média. O C_p compara os Limites de Especificação (LE) com os LNT. É o melhor indicador da potencial capacidade de um processo, uma vez que este está estável e centrado. O P_p compara os LE com a variabilidade do processo a longo prazo, permitindo medir a sua capacidade atual. Quanto maior for o valor destes indicadores, mais capaz será o processo de ir de encontro às especificações do cliente.

A análise da capacidade de um processo é utilizada para alcançar e manter um elevado nível de qualidade em processos industriais. Para melhorar um processo, é fundamental compreendê-lo e quantificar o seu desempenho. Como defende Yang (2019), os índices de capacidade são uma forma eficaz de avaliar a *performance* de um processo e identificar se deve ou não ser alvo de melhorias.

2.5 Modelos Preditivos

A aplicação de técnicas de modelação preditiva inclui várias metodologias: regressão linear múltipla e regressões não lineares, como logarítmicas, quadráticas ou parabólicas. Apesar de, na maioria dos casos, o comportamento das variáveis não ser linear, é comum a modelação recorrendo à regressão linear múltipla, uma vez que é mais intuitiva e simples de desenvolver. A sua finalidade é extrair conhecimento dos dados recolhidos.

A regressão linear múltipla é uma forma de análise preditiva muito utilizada (Chen et al., 2018). Para que seja válida é necessário que se verifiquem alguns pressupostos:

- Normalidade multivariada - a regressão múltipla assume que os resíduos estão normalmente distribuídos;
- Influência multivariada dos *outliers* - os coeficientes estimados da regressão múltipla podem estar enviesados se existirem pontos com uma influência indevida no modelo;
- Independência estatística - a regressão múltipla assume que os resíduos são estatisticamente independentes;

- Multicolinearidade - a regressão múltipla assume que não existe multicolinearidade nos dados;
- Homocedasticidade - a regressão múltipla assume que a variância do erro é similar ao longo dos valores das variáveis independentes.

Se forem garantidos os pressupostos listados, é possível concluir que o modelo obtido produz estimativas de coeficientes não viciadas com variância mínima.

Dentro da abordagem *Six Sigma*, a regressão linear múltipla é uma ferramenta muito utilizada, já que permite identificar relações entre as variáveis em estudo (vários *inputs* e um *output*) e traduzir essas relações em modelos matemáticos, como demonstrado na Equação 2.2.

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + e \quad (2.2)$$

Y representa a variável dependente, ou seja, a variável cujo valor se pretende estimar. As variáveis X são as variáveis independentes, isto é, que vão influenciar a variável resposta. Os β são os coeficientes da regressão: β_0 é a ordenada na origem e os restantes traduzem o impacto das variáveis X no valor da variável Y .

O coeficiente de erro que se espera em cada observação está representado por e . Quanto maior o seu valor, pior será o modelo obtido, uma vez que a relação entre as variáveis que compõem o modelo será mais fraca (Johnston et al., 2008). Esta característica é mensurável através do valor do Coeficiente de Determinação (R^2), que indica qual a percentagem de variação dos dados é explicada pelo modelo. Assim, permite avaliar se o modelo elaborado se ajusta ou não aos valores medidos (Montgomery and Runger, 2013).

Há vários métodos para determinar as variáveis estatisticamente significativas a incluir no modelo. A eliminação *backward* consiste em introduzir todas as variáveis no modelo inicial. As variáveis são testadas uma a uma para determinar a sua significância estatística. Removem-se as que não são significantes, ou seja, aquelas com um *p-value* inferior ao valor de α (Telmo et al., 2010). Em cada iteração é removida a menos significativa (a que tem um maior *p-value*), até obter um modelo composto apenas por regressores estatisticamente significativos.

Na literatura, a multicolinearidade é identificada como um problema que pode ter um impacto significativo na aplicabilidade do modelo. Quando as variáveis independentes estão correlacionadas, o seu coeficiente de regressão depende de quais outras variáveis são incluídas no modelo. Há três fontes primárias de multicolinearidade:

- O método utilizado na recolha de dados;
- Restrições do modelo ou da população;
- Especificações do modelo.

Os *Variance Inflation Factors* (VIF) são um indicador que permite identificar a presença de multicolinearidade no modelo. É aceite entre vários autores (Salmerón et al., 2017) que um VIF superior a 10 significa que os coeficientes da regressão estão mal calculados devido ao elevado valor de colinearidade, pondo em causa a qualidade do modelo.

A fim de resolver ou, pelo menos, de minimizar o efeito da multicolinearidade no modelo, é possível recolher dados adicionais, reajustar o modelo ou combinar as variáveis que apresentam multicolinearidade - realizar uma análise dos principais componentes (*Principal Component Analysis* – PCA).

A primeira opção nem sempre é viável por questões financeiras ou pelo facto de o processo em estudo já não estar disponível. A segunda hipótese pode significar redefinir os regressores, ou seja, encontrar uma fórmula que relacione os regressores (variáveis independentes incluídas na regressão) colineares e substituí-los por essa relação ou eliminar as variáveis colineares que é, habitualmente, mais eficaz, mas que pode implicar que o modelo perca capacidade de prever a resposta da variável dependente. A última solução é adotada principalmente quando há um número elevado de regressores (Montgomery et al., 2012).

O método de *principal component analysis* foi desenvolvido com o intuito de resolver os problemas de multicolinearidade, através da eliminação da instabilidade do modelo e reduzindo a variabilidade dos coeficientes de regressão.

Numa primeira fase, obtêm-se os componentes principais. Estes são combinações lineares das variáveis independentes incluídas no modelo. Na etapa seguinte, utiliza-se os principais componentes para elaborar um modelo que preveja o valor da variável resposta (Adnan et al., 2006).

A existência de colinearidade entre as variáveis independentes afeta a aplicabilidade prática do modelo. Ainda que, conceptualmente, seja possível variar apenas um dos regressores mantendo os restantes fixos, se estes estiverem correlacionados, na prática, uma alteração de uma das variáveis vai refletir-se nas restantes.

Numa outra perspetiva, defendida por Neter et al. (1983), diferencia-se o efeito da presença de multicolinearidade num modelo. É prejudicial para determinar o melhor conjunto de regressores a considerar. No entanto, se o objetivo for elaborar um modelo preditivo, o seu efeito pode ter pouca significância se as variáveis independentes seguirem padrões semelhantes de multicolinearidade. Um estudo realizado para avaliar o efeito da multicolinearidade concluiu que esta parece influenciar significativamente as previsões realizadas pelo modelo. Concluiu, também, que a utilização de *datasets* com um número elevado de observações permite mitigar o efeito da multicolinearidade no modelo (Mundfrom et al., 2018).

2.5.1 Avaliar a qualidade do modelo

Após a elaboração do modelo preditivo, é necessário aferir quão bem o modelo se ajusta aos dados (*goodness-of-fit*). De um modo simples, é possível afirmar que um modelo se ajusta aos dados quando a diferença entre os valores observados e os valores previstos pelo modelo é pequena e não enviesada.

Para assegurar a qualidade de um modelo, é fulcral dividir corretamente o *dataset*. A divisão dos dados em amostra de treino e amostra de teste é fundamental para avaliar a qualidade dos modelos obtidos. O objetivo de um modelo preditivo é que tenha capacidade de prever o valor da variável dependente a partir de novos valores das variáveis independentes. Logo, para aferir a qualidade

do modelo, este deve ser testado num conjunto de dados diferente daquele que foi utilizado para o elaborar.

Um dos indicadores que se pode analisar e cuja interpretação é intuitiva é o Coeficiente de Determinação (Giorgi, 2018). No entanto, um valor elevado de R^2 não implica necessariamente um impacto causal real (característica relevante numa perspetiva preditiva). O coeficiente tem em consideração três fatores: o impacto dos preditores (variáveis independentes incluídas no modelo preditivo) na variável dependente, o grau de variabilidade desta variável e o grau de dispersão em torno da linha de regressão (Moksony, 2014).

Num modelo elaborado através de uma regressão múltipla, a introdução de mais preditores pode contribuir para um aumento do R^2 . No entanto, é crucial perceber se esse aumento se deve a uma melhoria na qualidade do modelo ou ao aumento do número de variáveis independentes presentes no modelo. Nesse sentido, é importante analisar o valor do R^2 ajustado, já que tem em conta o número de preditores envolvidos no modelo. Se a adição de uma nova variável independente melhorar o modelo significativamente (mais do que seria esperado no caso de acrescentar uma variável aleatória), o valor do R^2 ajustado vai aumentar (Saunders et al., 2012). Caso contrário, vai diminuir.

Uma vez que não considera a variabilidade inerente ao processo, apenas analisar o valor do R^2 pode levar à introdução de variáveis a mais no modelo, já que contribuem para um aumento do Coeficiente de Determinação, mas não necessariamente para melhorar a qualidade do modelo. Quando se pretende elaborar um modelo preditivo, como na presente dissertação, é crucial o valor do R^2 previsto. Este indicador reflete quão bem um modelo consegue prever respostas para novas observações. É calculado retirando cada observação do *dataset*, elabora-se um modelo e avalia-se a capacidade de o modelo prever a observação retirada.

Para além do valor do Coeficiente de Determinação, é relevante analisar os *p-values* dos preditores incluídos no modelo. O facto de as variáveis que constituem o modelo serem estatisticamente significativas (*p-value* < 0,05) permite concluir que estas influenciam significativamente a variável resposta.

A fim de avaliar a capacidade dos modelos obterem previsões próximas da realidade, deve-se analisar os erros (resíduos) obtidos. Estes calculam-se através da diferença entre o valor real o valor previsto relativos a um determinado *input*. A precisão da previsão pode ser um critério de desempate quando se pretende eleger o melhor modelo.

Quando se pretende comparar modelos preditivos para variáveis com ordens de grandeza diferentes, recomenda-se calcular valores percentuais, como o MAPE (*Mean Absolute Percentage Error*). Em casos em que se pretende penalizar erros maiores, é aconselhável recorrer ao RMSE (*Root Mean Square Error*).

A interpretação do RMSE é mais complexa que a do MAPE. De acordo com Saigal and Mehrotra (2012), se a diferença do RMSE entre modelos for superior a 30%, será muito significativa. Caso seja cerca de 10%, provavelmente, será significativa, se for 2%, não será significativa.

O RMSPE (*Root Mean Square Percentage Error*) calcula o erro percentualmente, penalizando desvios maiores. Pode ser utilizado quando o intuito é melhorar a *performance* preditiva de um

modelo (Alfons et al., 2013). É calculado através da fórmula representada na Equação 2.3:

$$RMSPE = \sqrt{\frac{1}{n} \sum \left(\frac{y_i - \hat{y}}{y_i} \right)^2} \quad (2.3)$$

Onde:

- n é o número de previsões feitas;
- y_i é o valor real;
- $\hat{y}$ é o valor previsto.

Se dois modelos tiverem estatísticas de erro semelhantes, deve-se eleger o que for mais simples e/ou mais facilmente interpretável, já que será aceite com maior facilidade.

2.6 Machine Learning

Atualmente, em ambiente de produção industrial, há cada vez mais informação e dados disponíveis. Se, por um lado, uma maior quantidade de informação permite melhorar os processos e a qualidade dos produtos, por outro, pode ser um desafio e ter um impacto negativo nas organizações. Isto deve-se ao facto de desviar a atenção dos principais problemas ou porque complexifica o processo de análise, podendo levar a conclusões erradas ou mais demoradas (Wuest et al., 2016). A implementação de técnicas de *Machine Learning* (ML) facilita o tratamento e a interpretação dos dados, permitindo extrair relações entre as variáveis. Para além de ter capacidade de analisar diferentes tipos de variáveis, permite encontrar relações entre variáveis que não têm um comportamento linear. Assim, ML é uma ferramenta que pode ser utilizada a fim de melhorar a qualidade dos produtos e a gestão de recursos (Wuest et al., 2016).

Machine Learning pode ser aplicado em diversas áreas, englobando várias metodologias que se podem dividir em três categorias:

- *Reinforcement Learning*: O algoritmo recebe *feedback* que o premia ou castiga se acertar ou errar, respetivamente;
- *Unsupervised Learning*: O algoritmo utiliza um conjunto de dados em que as variáveis não estão identificadas;
- *Supervised Learning*: O algoritmo utiliza dados classificados, ou seja, as variáveis estão identificadas.

Dentro da categoria *Supervised Learning* encontram-se, entre outros, as árvores de regressão. Estas inserem-se dentro do CART (*Classification And Regression Trees*). Se a variável objetivo for categórica, trata-se de uma árvore de classificação. Caso a variável objetivo seja numérica, está-se perante uma árvore de regressão (Choi et al., 2013). A utilização de árvores tem a vantagem de ser intuitiva e de a informação ser facilmente interpretada visualmente (Naghbi et al., 2016).

O processo de elaboração de uma árvore de regressão passa por dividir os dados sucessivamente até atingir os nódulos finais. Consideram-se as variáveis significativas e dividem-se os dados da forma que permitir reduzir mais os desvios da variável independente. Cada "caminho" é dado

por uma série de decisões dicotómicas, que representa as condições que vão levar ao resultado (Lawrence and Wright, 2001). O valor representado no nó final é a média das observações da amostra que pertencem a esse nó.

As árvores são maioritariamente utilizadas com fins preditivos. No entanto, outra finalidade das árvores de regressão é otimizar o valor de parâmetros. No caso de ter um valor objetivo para a variável dependente, é possível recorrer a uma árvore de regressão para determinar qual o “caminho” para obter esse valor.

Hayes et al. (2016) utilizaram uma metodologia em que, numa primeira fase, recolheram dados e desenvolveram uma árvore de decisão com base nessas observações. De seguida, face aos resultados obtidos, determinaram quais os valores a definir para as variáveis de modo a obter um determinado valor para a variável dependente.

Capítulo 3

Descrição do Problema

Neste capítulo descreve-se o processo produtivo e o processo de análise de qualidade do produto final. Utilizou-se o método DMAIC para estruturar a abordagem ao problema. As duas primeiras etapas: *Define* e *Measure* estão contidas neste capítulo. Definiu-se o objetivo do projeto, tendo em conta as expectativas e exigências dos clientes, estudou-se o processo e foram recolhidos dados para validar e quantificar o problema.

3.1 Descrição do processo

Na Empresa A, devido à vasta oferta de produtos, o processo produtivo divide-se em várias vertentes. A presente dissertação tem como foco a operação *core* da organização: formação de placas de fibra numa prensa contínua.

A produção de placas de fibra é uma operação com várias etapas e à qual está associado um elevado número de variáveis. Estas eram geridas de forma empírica antes do arranque do projeto. A fim de garantir a qualidade da cadeia produtiva, várias características do produto final são avaliadas antes da sua venda.

Face aos resultados dos testes efetuados no laboratório, as placas produzidas podem ser caracterizadas de três formas. Se as especificações do cliente forem cumpridas, estão aptas para serem vendidas. Caso contrário, as placas serão requalificadas ou desqualificadas.

Se a categoria inferior tiver no portfólio placas com a mesma espessura das que estão a ser produzidas, as placas vão ser requalificadas. Este processo consiste em vender, mas ao preço de um produto de uma categoria inferior. Caso a espessura produzida não pertença à gama de produtos da categoria inferior, as placas vão ser desqualificadas. Nesse caso, serão vendidas como um produto de segunda categoria. A desqualificação reduz mais o preço do que a requalificação, representando uma maior diminuição da margem operacional.

Um dos valores críticos para determinar a qualidade do produto final é o valor da resistência à tração, o qual permite aferir a resistência da placa.

O processo utiliza como matéria-prima troncos de árvores, que vão ser cortados e destrocados e aos quais vão ser adicionados outros reagentes, como resina e químicos (componente A, componente

B, componente C, componente D ¹), formando uma pasta que se designa por fibra. O *core* do processo é a prensagem. A fibra é utilizada para constituir o colchão que vai ser sujeito a diferentes níveis de pressão e temperatura ao longo dos 60 metros da prensa contínua.

Numa primeira fase, o colchão é alisado por uns rolos, com o objetivo de o uniformizar. De seguida, o colchão vai passando ao longo das várias arcadas da prensa onde é sujeito a diferentes valores de pressão e temperatura. Na extremidade final da prensa, as placas são cortadas conforme o comprimento acordado com o cliente.

No final deste percurso, obtém-se uma placa de aglomerado que, após arrefecer, será sujeita a testes de laboratório que vão permitir aferir a sua qualidade.

No laboratório, a placa é cortada transversalmente em amostras mais pequenas. O conhecimento empírico dos operadores leva-os a crer que as características da placa se mantêm constantes longitudinalmente, pelo que, caso seja garantida transversalmente, a qualidade da placa manter-se-á longitudinalmente.

As amostras são sujeitas a vários testes no laboratório. O seu comprimento, largura e espessura são medidos para calcular o volume. De seguida, são pesadas. No passo seguinte, a amostra é colocada num aparelho que mede a sua densidade ao longo da espessura da placa, obtendo também um valor para a densidade do núcleo.

O último teste é o da resistência à tração. É realizado no final, porque consiste em colocar a amostra numa máquina que aplica uma força de tração, alongando gradualmente a amostra até à sua fratura.

Dado que a duração do teste de resistência à tração em laboratório é de, no mínimo, 30 minutos (15 minutos para preparar a amostra, se não houver fila de espera, e 15 minutos para realizar o teste), caso se detete que as especificações não estão a ser cumpridas, haverá uma quantidade significativa de placas que não estarão aptas para ser vendidas.

A linha de produção contínua funciona 24 horas por dia, em três turnos, 7 dias por semana. O processo é contínuo e sequencial, dependendo maioritariamente dos equipamentos. No entanto, ressalva-se o facto de os operadores terem a função de controlar e monitorizar os parâmetros das máquinas e o seu funcionamento.

3.2 Seleção do tipo de produto

A Empresa A oferece um vasto portfólio de produtos aos seus clientes, organizado em tipos de produtos com características diferentes, obtidos através de processos produtivos diferentes. Assim, tornou-se pertinente selecionar um tipo de produto para ser alvo de análise, de modo a centrar a análise num grupo mais restrito, a fim de obter resultados mais concretos.

A seleção do tipo de produto dividiu-se em duas fases. Primeiro, eliminaram-se os tipos de produto com menor expressão, isto é, que são produzidos menos frequentemente. Por um lado, porque não haveria tantos dados para analisar, o que poderia tornar-se uma limitação do projeto. Por outro,

¹ Por uma questão de confidencialidade, não se especificam quais os químicos utilizados no processo produtivo

porque um dos objetivos do projeto é aumentar a margem operacional da organização, tendo mais impacto se a melhoria for efetuada num produto com muita procura.

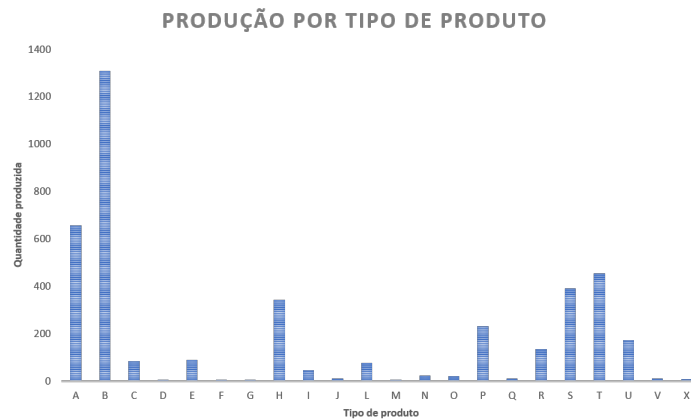
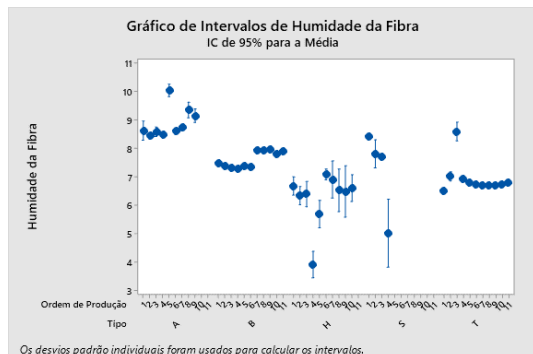
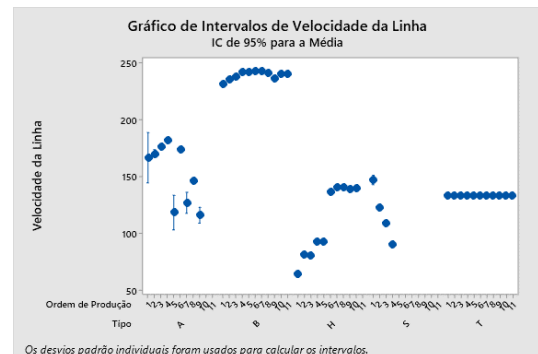


Figura 3.1: Quantidade produzida entre Janeiro de 2019 e Março de 2020, por tipo de produto.

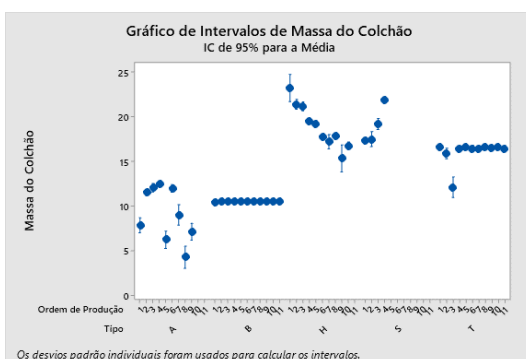
A análise do gráfico da Figura 3.1 permitiu identificar o *Top 5* dos tipos de produtos mais produzidos: A (16%), B (32%), H (8%), S (10%) e T (11%). Excluíram-se os restantes dados, uma vez que se pretende selecionar uma amostra representativa - os cinco tipos de produtos escolhidos representam 77% da produção total.



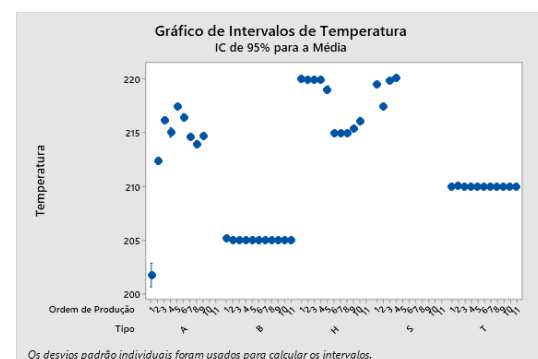
(a) Humidade da Fibra



(b) Velocidade da Linha



(c) Massa do Colchão



(d) Temperatura da Prensa

Figura 3.2: Variação dos valores por ordem de produção e tipo de produto.

De seguida, analisaram-se as variáveis do processo dentro de cada ordem de produção. Escolheu-se analisar a humidade da fibra, a velocidade da linha, a massa do colchão e a temperatura da prensa (Figura 3.2).

Optou-se pela massa do colchão e pela humidade da fibra porque, pelo conhecimento empírico dos trabalhadores, se destacaram como muito significativas para a qualidade do produto (hipótese que foi analisada no decorrer do projeto). A velocidade da linha e a temperatura da prensa foram escolhidas pelo facto de serem variáveis que os operadores conseguem alterar ao longo do processo. Assim, é possível perceber se se trata de um processo estável ou se é sujeito a ajustes regulares. Após analisar a variabilidade dos valores registados para cada variável por ordem de produção e por tipo de produto, é possível concluir que as variáveis do processo estão mais controladas durante a produção do tipo B e do tipo T.

Assim, conjugando a quantidade produzida e a estabilidade das variáveis de processo, optou-se por analisar a produção do tipo B.

3.3 Controlo estatístico do processo

Para compreender melhor o processo e o comportamento da variável de qualidade em estudo, realizou-se uma análise à variável resposta Resistência à Tração. No caso de se estar a produzir placas do tipo B, há três valores objetivo para a resistência à tração - 0,55Pa, 0,70Pa e 0,75Pa. Realizaram-se cartas de controlo para cada valor objetivo a fim de avaliar a capacidade de o processo produzir placas conforme as especificações acordadas com o cliente.

De acordo com a espessura pretendida, define-se um valor objetivo para a resistência à tração das placas. A empresa tem de assegurar que as placas têm, no mínimo, esse valor de resistência à tração. Este é o Limite Inferior de Especificação (LIE), ou seja, o menor valor aceitável para esta característica. Não há um Limite Superior de Especificação (LSE), no entanto, aumentar a resistência à tração passa por utilizar uma quantidade de matéria prima superior à necessária, o que se traduz numa redução da rentabilidade da empresa. Como as especificações são unilaterais, o valor referência para avaliar a capacidade do processo é 1,25 (Montgomery, 2009).

Idealmente, os grupos deveriam ser constituídos por amostras pertencentes ao mesmo lote. No entanto, tal não foi possível devido à interrupção temporária da atividade na empresa. A fim de não atrasar o projeto, a análise realizou-se com base em dados históricos, tendo sido respeitada a sequência temporal dos mesmos. Isto é, os dados foram retirados em fluxo, sendo que os primeiros subgrupos englobam os dados mais antigos e os últimos os dados mais recentes.

A carta de controlo $\bar{x}$ -R para um valor objetivo de 0,55Pa encontra-se na Figura 3.3. O número de amostras disponível para esta categoria é reduzido. Utilizaram-se dados históricos recolhidos entre Janeiro de 2019 e Março de 2020, organizados em 3 grupos de 3 elementos.

A carta R avalia a variabilidade dos dados de cada subgrupo. Como os pontos se encontram dentro dos limites de controlo, é possível afirmar que o processo está sob controlo.

A carta $\bar{x}$ avalia a variabilidade do valor esperado do processo. Uma vez que os pontos estão dentro dos limites de controlo estabelecidos, pode concluir-se que o processo se encontra estável.

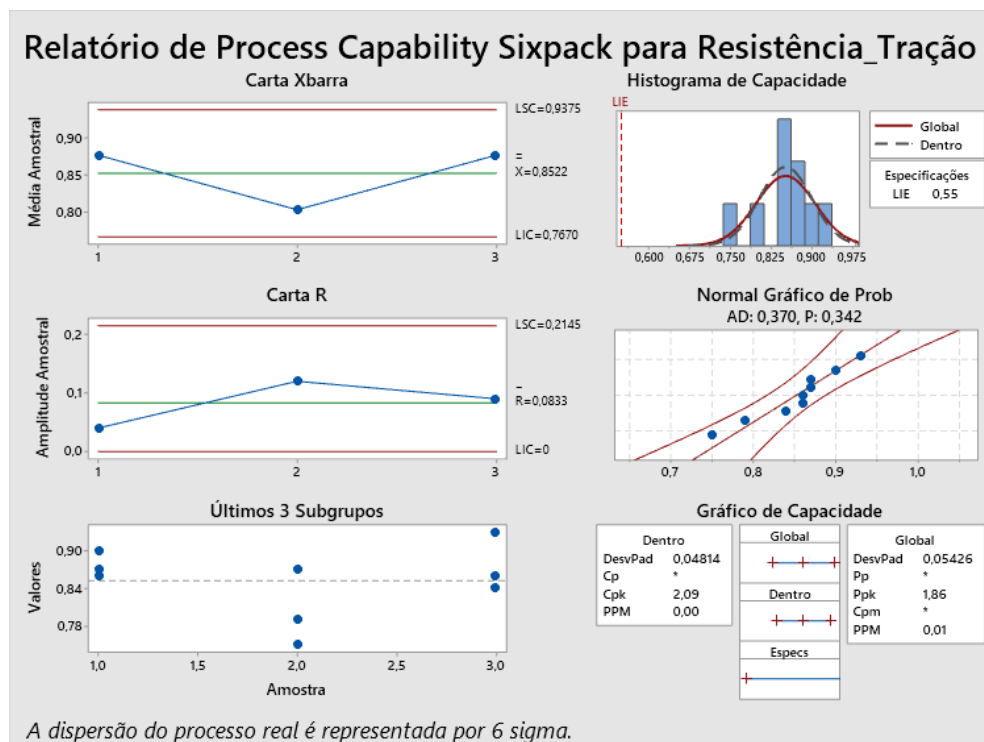


Figura 3.3: Carta de controlo do processo para uma resistência à tração pretendida de 0,55.

Como o processo se encontra em controlo estatístico, é possível avaliar a sua capacidade. Uma vez que não foi definido o LSE, os valores a analisar são o C_{pk} e o P_{pk} . Como ambos são superiores ao valor de referência (1,25), pode-se concluir que o processo é capaz.

Através do teste de *Anderson-Darling* é possível afirmar que a característica de qualidade segue uma Distribuição Normal. De acordo com o Teorema do Limite Central, à medida que o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se de uma distribuição normal. Logo, é possível generalizar os resultados obtidos e concluir que o processo se encontra sob controlo e que é capaz de produzir produtos que satisfazem os requisitos definidos (resistência à tração mínima de 0,55Pa). Por conseguinte, para este valor objetivo, não há necessidade de realizar um projeto de *Six Sigma*.

Na Figura 3.4 encontra-se a carta de controlo do processo, quando o valor de resistência à tração pretendido é 0,70Pa. Os dados da amostra foram agrupados em 62 subgrupos de 8.

Tanto na carta $\bar{x}$, que avalia a centragem dos dados da amostra, como na carta R, que mede a amplitude da amostra, há valores fora dos limites de controlo definidos. Deste modo, é possível concluir que há indícios de que o processo está sujeito a causas assinaláveis de variação, não estando sob controlo.

Esta análise estatística e a identificação de pontos fora dos limites de controlo permite concluir que, para além da variabilidade inerente ao processo, há variabilidade associada à medição da característica de qualidade – Resistência à Tração – e aos testes realizados em laboratório (Jones et al., 2005).

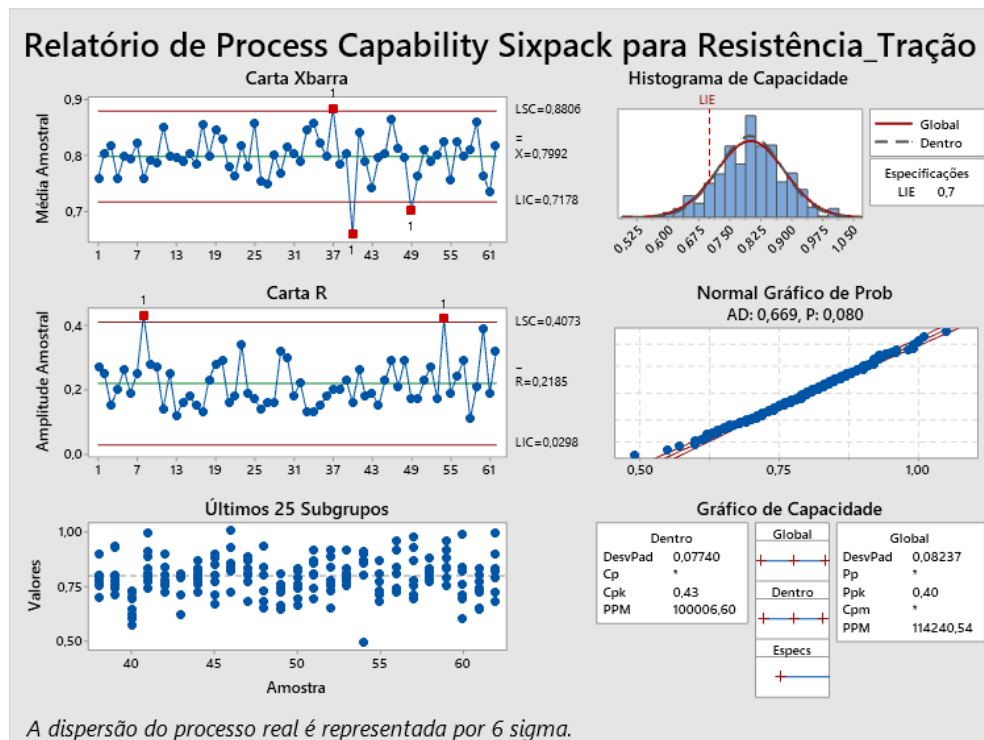


Figura 3.4: Carta de controlo do processo para uma resistência à tração pretendida de 0,70.

Um dos pressupostos para poder avaliar a capacidade de um processo é que este se encontre estável estatisticamente. Tal não se verifica, no entanto, esta análise foi levada a cabo com o objetivo de compreender melhor a capacidade do processo. A capacidade atual do processo (P_{pk}) apresenta um valor muito baixo, logo, o processo não é capaz. Reduzindo a variabilidade entre amostras será possível alcançar uma capacidade do nível da indicada por C_{pk} (medida da capacidade potencial). O valor de C_{pk} (0,43) é inferior ao valor de referência, pelo que resolver as causas de variação assinaláveis não será suficiente para tornar este processo capaz. Assim, o foco do projeto vai ser esta referência.

A carta de controlo do processo para um valor de resistência à tração objetivo de 0,75Pa encontra-se na Figura 3.5. Os dados da amostra foram agrupados em 19 subgrupos de 3.

A análise da carta R permite afirmar que o processo não está sob controlo relativamente à variabilidade dentro de cada grupo amostral, devido à existência de pontos fora dos limites.

O facto de haver pontos fora dos limites de controlo da carta $\bar{x}$ parece indicar que o processo está fora de controlo. No entanto, os limites são calculados com base na centragem, mas também na dispersão do processo. Se a carta R estiver fora de controlo, os limites calculados poderão ser imprecisos e indicar incorretamente uma situação fora de controlo (falso alarme) ou falhar a deteção de uma situação anómala.

Dado que foram registados pontos fora dos limites em ambas as cartas, é possível afirmar que o processo não se encontra controlado estatisticamente. No entanto, como se trata de apenas um ponto em cada carta, procedeu-se à análise da capacidade do processo.

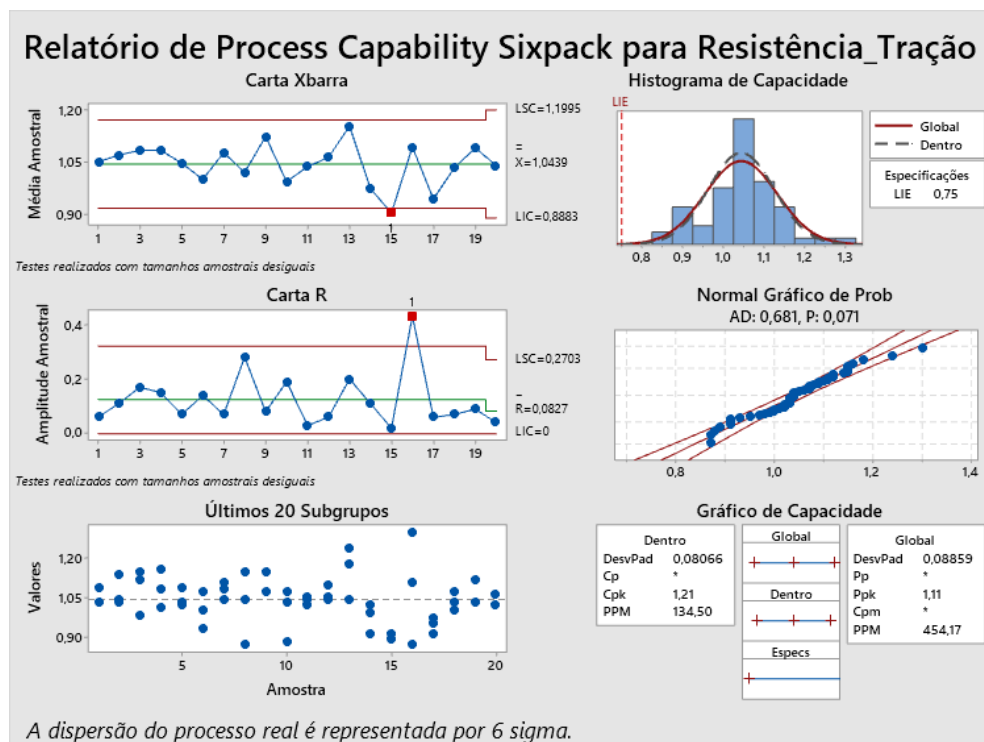


Figura 3.5: Carta de controlo do processo para uma resistência à tração pretendida de 0,75.

Como o valor da variabilidade global é muito próximo do valor da variabilidade dentro do grupo, é possível avaliar a capacidade do processo através do C_{pk} . Uma vez que se encontra muito próximo do valor de referência (1,25), vai-se considerar que o processo é capaz de corresponder às especificações definidas.

Assim, decidiu-se que o foco do projeto seria, numa primeira fase, os produtos com valor de resistência à tração objetivo de 0,70Pa por se tratar do caso mais crítico, isto é, processo fora de controlo e incapaz de cumprir as especificações determinadas. No futuro, recomenda-se que a empresa melhore também o processo cujo valor objetivo é 0,75Pa.

3.4 Seleção do produto

Quando um cliente faz uma encomenda e especifica o produto que quer comprar, há três características que são referência do produto e que têm de ser asseguradas: a resistência à tração, a densidade e a espessura da placa. Estas três características são consideradas CTQ no processo produtivo em estudo.

Para um valor objetivo de resistência à tração, pode haver diferentes valores de densidade pretendida, ao qual correspondem determinados valores de espessura de placa, como explicitado na Figura 3.6.

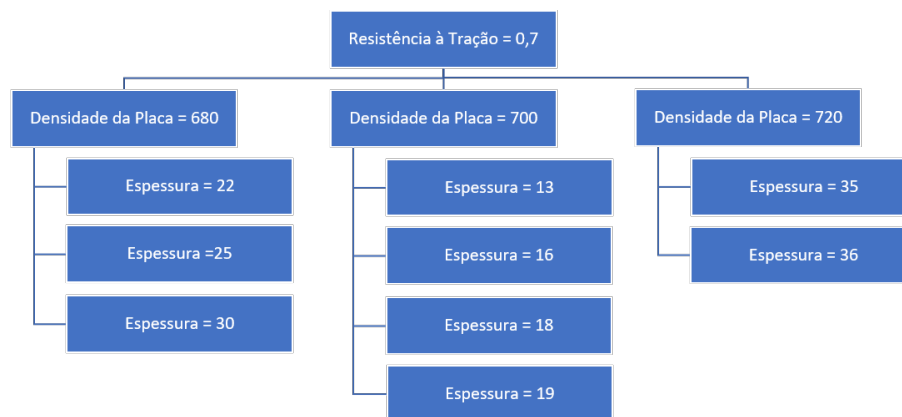


Figura 3.6: Relação entre as três características de qualidade.

De acordo com a placa que está a ser produzida, é necessário fazer ajustes à linha de produção para poder atingir os valores objetivo das características de qualidade. A fim de compreender melhor o processo e de identificar as variáveis significativas para a qualidade do produto final, decidiu-se analisar placas com a mesma densidade pretendida.

Como critério de decisão, recorreu-se à Curva ABC, já que, tal como referido anteriormente, se pretende focar a análise num segmento de produtos com representatividade para a empresa. A classificação ABC permite ordenar os produtos, tendo em conta a percentagem de volume de negócios que lhes corresponde (Teunter et al., 2010). Por motivos de confidencialidade, apresenta-se a classificação ABC sem indicar os valores de volume de vendas correspondentes a cada subgrupo de produtos (Tabela 3.1).

Tabela 3.1: Classificação ABC dos produtos de acordo com a densidade pretendida.

Densidade Pretendida (kg/m^3)	Classificação
700	A
720	B
680	C

Uma vez que as placas com densidade pretendida de 700 kg/m^3 foram classificadas como produtos A (os que geram maior volume de vendas), elegeu-se este segmento de produtos, que representa, sensivelmente, 25% da produção anual da Empresa A. De seguida, revelou-se pertinente estudar a variável de qualidade crítica para o projeto - Resistência à Tração.

O objetivo do projeto é estabilizar e manter o processo sob controlo, ou seja, com um comportamento constante e cujo resultado obtido se encontre dentro das especificações. Pretende-se que o valor da resistência à tração das placas seja o mais próximo possível do objetivo ($0,70\text{Pa}$) para reduzir os custos operacionais, assegurando que nunca tem um valor inferior à especificação para que não haja desqualificação.

Analisando os dados da amostra ($N= 279$), verificou-se que todas as placas cumprem os requisitos de espessura, ou seja, a sua espessura no final do processo produtivo está de acordo com as especificações acordadas com o cliente.

De seguida, verificou-se que o valor médio da densidade das placas é superior ao valor de referência. Este valor é obtido através de um sobreconsumo de matéria prima. Como todas as amostras estão de acordo com as especificações de espessura e todas têm o mesmo comprimento e largura, é possível assumir que o volume das amostras tem um valor constante. Logo, se as amostras tiverem uma densidade superior ao objetivo, significa que se utilizou uma quantidade de matéria prima superior à necessária.

Apesar da especificação acordada com o cliente ser uma densidade de 700 kg/m^3 , as placas têm uma tolerância de 30 kg/m^3 , ou seja, uma placa com uma densidade de 670 kg/m^3 é aceite pelo cliente. Assim, para a empresa, o objetivo é produzir placas com uma densidade de 670 kg/m^3 , uma vez que lhe permite cumprir os requisitos de qualidade acordados com o cliente, utilizando menos matéria prima e, consequentemente, reduzindo os custos operacionais.

Face a este valor objetivo, a densidade média das placas representa uma taxa de densidade extra de 6,47%. Dado que a densidade da placa se calcula através da razão entre a massa da amostra e o seu volume, o sobreconsumo de matéria prima tem o mesmo valor que a taxa de excesso de densidade - 6,47%.

O tipo de produto que está a ser objeto de estudo na presente dissertação - placas do tipo B, com uma densidade pretendida de 700 kg/m^3 e com um valor objetivo de resistência à tração de $0,70 \text{ Pa}$ - representa 23,46% da produção anual da empresa.

A Empresa A produz 500 m^3 de placas diariamente, 7 dias por semana, 40 semanas por ano, o que se traduz numa produção anual de $140\,000 \text{ m}^3$.

Das espessuras que foram alvo de estudo, se não cumprirem os requisitos de qualidade (valor da resistência à tração mínimo de $0,70 \text{ Pa}$), duas são requalificadas e duas são desqualificadas. No ano de 2019, o segmento de produtos em estudo teve uma taxa de desqualificação de 0,18% e uma taxa de requalificação de 0,31%.

Tendo em conta as taxas de sobreconsumo, desqualificação e requalificação identificadas, calcularam-se os custos adicionais de produção correspondentes a 2019 (referentes à gama de produtos em estudo). Os gastos, cujos cálculos estão detalhados no capítulo 5, são superiores a 300 k€ .

Assim, o projeto visa aumentar o nível de controlo e estabilidade do processo produtivo, reduzindo as taxas de requalificação e desqualificação dos produtos. Por outro lado, pretende dotar a empresa de um modelo que lhe permita controlar a qualidade do produto final, ajustando as variáveis do processo, diminuindo o consumo excessivo de matéria prima e aumentando a exatidão (acertar no valor objetivo) do resultado final.

Capítulo 4

Métodos

Neste capítulo, continua-se a fase *Measure* e segue-se para a fase *Analyse*. Os testes referentes à capacidade do sistema de medição estão contidos na segunda etapa do DMAIC. Já a análise às variáveis do processo estatisticamente significativas para determinar a característica de qualidade em estudo está compreendida na terceira fase da metodologia. Devido ao grau de complexidade do processo produtivo e do elevado número de variáveis envolvidas, decidiu-se dividir a análise em três níveis (Resistência à Tração - Densidade do Núcleo - Densidade da Placa).

4.1 Análise da variabilidade

Quantificar as fontes de variação é uma etapa fulcral de qualquer projeto de melhoria contínua que visa reduzir a variabilidade de um processo (Clapp et al., 2008).

A variabilidade total do processo (σ_{total}^2) pode ter várias fontes, como representado na Equação 4.1.

$$\sigma_{total}^2 = \sigma_{met}^2 + \sigma_{mo}^2 + \sigma_{maq}^2 + \sigma_{sm}^2 + \sigma_{ma}^2 + \sigma_{mp}^2 \quad (4.1)$$

As fontes de variabilidade dividem-se em seis categorias:

- variabilidade devido ao método de trabalho (σ_{met}^2);
- variabilidade devido à mão de obra (σ_{mo}^2);
- variabilidade devido às máquinas (σ_{maq}^2);
- variabilidade devido ao sistema de medição (σ_{sm}^2);
- variabilidade devido ao meio ambiente (σ_{ma}^2);
- variabilidade devido à matéria prima (σ_{mp}^2).

Se se englobar a variabilidade do processo produtivo ($\sigma_{processo}^2$) segundo a Equação 4.2, é possível simplificar a Equação 4.1, obtendo a Equação 4.3.

$$\sigma_{processo}^2 = \sigma_{met}^2 + \sigma_{mo}^2 + \sigma_{maq}^2 + \sigma_{ma}^2 + \sigma_{mp}^2 \quad (4.2)$$

$$\sigma_{total}^2 = \sigma_{processo}^2 + \sigma_{sm}^2 \quad (4.3)$$

Estes potenciais fatores de variabilidade provocam variações que podem ser originadas por causas naturais ou causas assinaláveis, sendo constituídos por variáveis passíveis de se ajustar e outras incontrolláveis. A identificação das fontes de variabilidade e o conhecimento das variáveis que lhe estão associadas permite tomar ações que atenuem a variabilidade do processo.

Devido à interrupção na atividade da empresa, não foi possível realizar os testes R&R, pelo que o projeto se focou em minimizar a variação inerente ao processo. Numa fase seguinte, será vantajoso avaliar a variação proveniente do sistema de medição e, caso se justifique¹, levar a cabo ações de melhoria.

4.1.1 Variabilidade do processo

O diagrama de *Ishikawa* permite estruturar as causas de variabilidade organizadas por categorias, organizando-as de uma forma mais visual.

Em ambiente de *workshop*, com a equipa, elaborou-se o diagrama de causa e efeito para determinar através de uma abordagem estruturada potenciais fontes de variabilidade presentes no processo (Figura 4.1).

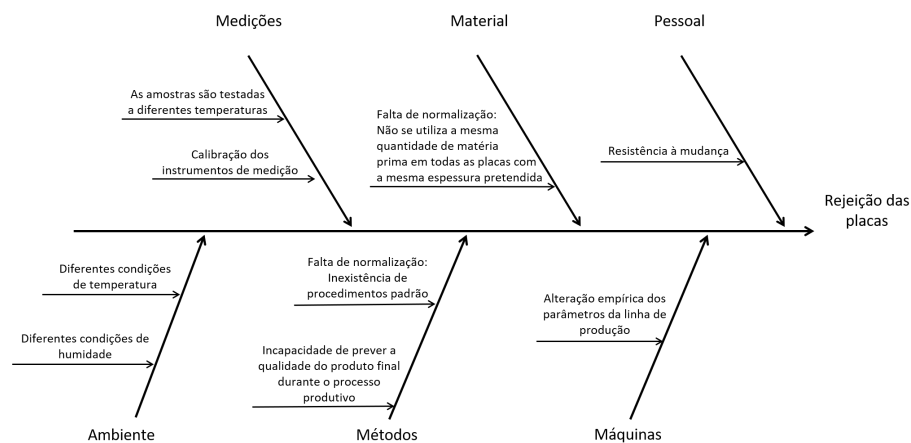


Figura 4.1: Diagrama de causa e efeito.

Uma vez que, na indústria de processo, os operadores desempenham funções de controlo, monitorização e manutenção, não estando diretamente envolvidos no decorrer do processo produtivo, os fatores associados à mão de obra foram desprezados.

Como aspetos mais relevantes destacam-se a incapacidade de prever a qualidade do produto final durante o processo produtivo e a alteração empírica dos parâmetros da máquina.

Assim, revelou-se pertinente desenvolver um modelo preditivo que, através dos valores dos parâmetros da linha de produção, permitisse prever se as placas estariam de acordo com as especificações do cliente ou não. Deste modo, seria possível ajustar o processo mal se detetasse um desvio, sem ser necessário esperar pela análise realizada em laboratório.

¹De acordo com o AIAG (*Automotive Industry Action Group*), se a variabilidade do sistema de medição representar até 10% da variabilidade total, pode-se concluir que o sistema de medição é aceitável

4.1.2 Análise exploratória

A fim de analisar estatisticamente algumas variáveis do processo e de validar os pressupostos assumidos, revelou-se pertinente fazer testes exploratórios. Neste âmbito, realizou-se uma análise à variância através de um teste ANOVA.

Nos testes realizados em laboratório, apenas se avalia a qualidade da placa transversalmente, assumindo-se que não há variações na qualidade da placa longitudinalmente (Figura 4.2). Este é um paradigma da organização que é crucial validar, na medida em que vai determinar se os testes de qualidade levados a cabo no laboratório são suficientes para garantir a conformidade com as especificações do cliente,

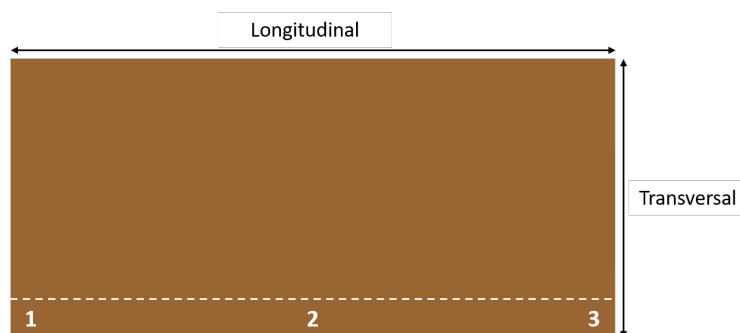


Figura 4.2: Direções longitudinal e transversal da placa e secção de corte para teste ANOVA.

ANOVA é um teste de hipótese estatístico que pode ser utilizado para comparar a média de diferentes populações. A Hipótese Nula (H_0) assume que estas não são significativamente distintas. A rejeição de H_0 permite concluir que há evidências estatísticas de que as médias diferem significativamente.

Para validar os resultados obtidos no teste ANOVA é necessário garantir três pressupostos relativamente aos erros: a normalidade da distribuição, a independência mútua e a homogeneidade da variância.

A análise dos erros sucede a realização do teste ANOVA, uma vez que os resíduos analisados têm origem nos resultados do teste (Montgomery, 2009).

Devido à interrupção temporária da Empresa A, não foi possível realizar os testes para confirmar que a qualidade da placa apenas varia transversalmente, mantendo-se constante na direção longitudinal. No entanto, simularam-se dados para explicar o procedimento a desenvolver e as conclusões a tirar em cada etapa, a fim de agilizar esta análise quando for possível voltar à empresa.

Para efetuar as medições, devem ser utilizadas placas que pertençam ao objeto em estudo na presente dissertação, ou seja, do tipo B, com um valor objetivo de resistência à tração de 0,7 e com uma densidade pretendida de 700.

Deve ser retirada uma secção longitudinal da placa. Dessa secção vão ser retiradas três amostras (uma de cada extremidade e uma do centro) que vão ser alvo de estudo. As localizações 1 e 3 representam as extremidades da placa e a localização 2 representa o centro da placa, como representado na Figura 4.2. Para cada amostra, deve-se medir o valor da densidade da placa e

o valor da resistência à tração, utilizando o equipamento específico para essa finalidade, como descrito no Capítulo 3.

Simularam-se valores para 25 placas, ou seja, 25 amostras para cada localização. Os dados utilizados estão registrados no Apêndice A.

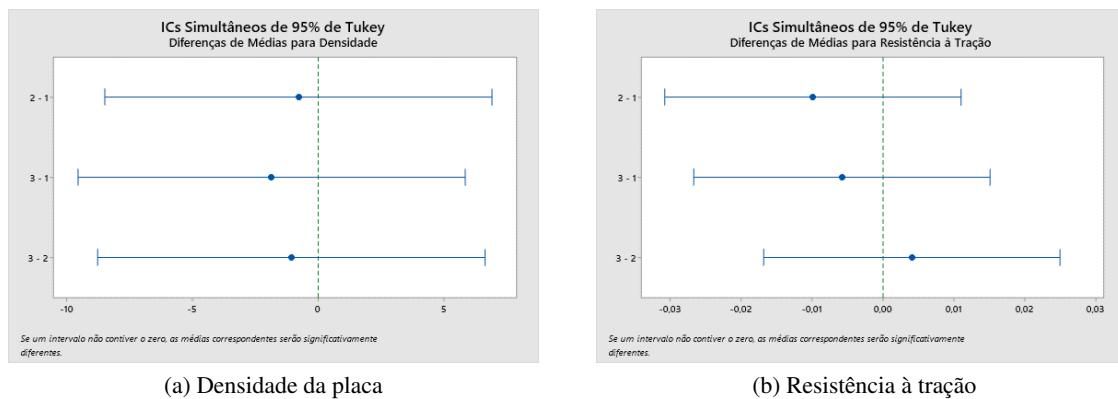


Figura 4.3: Intervalos de confiança de 95% de Tukey relativamente à localização transversal da placa.

Através da análise dos resultados registrados na Figura 4.3 é possível concluir que não há evidências estatísticas para rejeitar H_0 . Esta conclusão é suportada pelo valor do p -value (densidade da placa: p -value = 0,847; resistência à tração: p -value = 0,528) ser superior ao nível de significância do teste (0,05). Uma vez que todos os intervalos contêm o zero, pode-se assumir que o pressuposto de que a qualidade da placa se mantém constante longitudinalmente é válido.

A fim de garantir a validade dos resultados obtidos, procedeu-se à análise dos resíduos, relativamente aos três pressupostos referidos anteriormente.

Para garantir a independência mútua, os dados devem ser recolhidos num momento temporal específico e em condições semelhantes.

Relativamente à normalidade, realizou-se um teste de *Anderson-Darling*, cujos resultados estão registrados no Apêndice B. Uma vez que, quer para a densidade da placa quer para a resistência à tração, o p -value do teste é inferior a 0,05, é possível concluir que se garante o pressuposto da normalidade da distribuição.

Como os erros seguem uma distribuição normal, recorre-se ao teste de *Bartlett* para avaliar a homogeneidade da variância. Os resultados deste teste estão registrados no Apêndice B. Para as duas características de qualidade, verificou-se que o p -value é superior a 0,05. Logo, não há evidências estatísticas que levem a rejeitar a hipótese nula.

Assim, como as observações são independentes e normalmente distribuídas e não apresentam diferenças significativas na variância para cada zona da placa, os resultados obtidos pelo teste de Tukey são válidos e é possível utilizá-los para tirar conclusões. Logo, o teste ANOVA valida o pressuposto da Empresa A de que é suficiente avaliar a qualidade da placa transversalmente, uma vez que esta não varia longitudinalmente.

4.2 Desenvolvimento dos modelos

O princípio de Pareto, também designado por regra dos 80/20, defende que, ainda que haja um elevado número de fatores relacionados com o resultado de um processo, apenas alguns (poucos) o afetam significativamente (Gygi et al., 2005). Assim, pretende-se isolar estas variáveis e fixar valores standard para as restantes, a fim de haver menos variáveis a influenciar o processo e, consequentemente, reduzir a sua variabilidade.

Os dados foram divididos usando a regra $\frac{2}{3}$ - $\frac{1}{3}$. 67% constituiu a amostra de treino, ou seja, foram utilizados para desenvolver o modelo e os restantes 33% foram selecionados para validar o modelo elaborado. Pelo facto de estarem ordenados temporalmente, os dados foram divididos de forma aleatória, de forma a que os dois subgrupos contivessem informação representativa do período analisado. A decisão de dividir a amostra segundo este rácio foi suportada pela literatura. Como a dimensão da amostra é superior a 100, utilizar $\frac{2}{3}$ dos dados para elaborar o modelo permite obter bons resultados (Dobbin and Simon, 2011).

Há duas vantagens significativas que justificam o porquê de optar pela regressão linear múltipla. Por um lado, é simples desenvolver um modelo de regressão linear múltipla recorrendo ao Minitab. Por outro lado, é fácil de interpretar e compreender, o que facilita a sua implementação numa fase mais avançada do projeto.

Para elaborar os modelos, recorreu-se ao Minitab e calculou-se uma regressão linear múltipla, utilizando o método de eliminação *Backward* com um nível de significância de 95%, ou seja, com $\alpha = 0,05$.

4.2.1 Resistência à Tração

Numa primeira fase, estudaram-se quais as variáveis significativas para a resistência à tração (RT). Em ambiente de *workshop* com a equipa envolvida no projeto, procurou-se, através do seu conhecimento empírico, determinar as variáveis do processo que influenciam a característica de qualidade em estudo.

Traçou-se um diagrama de relações, registando as variáveis que poderiam ter impacto na qualidade da placa final (Figura 4.4). Para validar as hipóteses levantadas, realizou-se uma análise estatística. Desenvolveu-se um modelo, através de uma regressão linear múltipla, para determinar quais as variáveis significativas ($p\text{-value} < 0,05$). As variáveis a verde na Figura 4.4 são as que o modelo considerou significativas.



Figura 4.4: Diagrama de relações das variáveis que impactam a resistência à tração.

De acordo com o modelo obtido, as variáveis com um efeito significativo na RT são a densidade do núcleo, a pressão e a distância da prensa, a quantidade de resina, a humidade do secador e a densidade do *MatFormer*.

Por uma questão de simplificação do diagrama, representou-se a pressão da prensa como uma variável única. No entanto, para a elaboração do modelo utilizaram-se os valores de pressão medidos em cada arcada da prensa. Repetiu-se este procedimento para a distância da prensa.

Para avaliar a qualidade do modelo, recorreu-se ao coeficiente de determinação (R^2) para aferir se se está a incluir no modelo variáveis suficientes para explicar o comportamento da variável resposta. O modelo obtido tem um R^2 de 53,17%, um R^2 ajustado de 47,59% e um R^2 previsto de 39,20%.

Através da análise da coluna "Contribuição" da Figura C.4 é possível concluir que o regressor com maior impacto no valor da resistência à tração é a densidade do núcleo. O modelo tem capacidade de explicar 53,17% do comportamento da variável dependente (RT). Dessa percentagem, 15,62% são explicados através do valor da densidade do núcleo. No entanto, esta variável é medida no laboratório, pelo que se revelou pertinente explorar quais as variáveis do processo que tinham influência na densidade do núcleo.

Para garantir a validade do modelo obtido, realizou-se uma análise para verificar se são cumpridos os pressupostos da regressão linear múltipla. Todos os pressupostos, à exceção do da ausência de multicolinearidade, foram assegurados ².

A fim de analisar a precisão do modelo preditivo, utilizou-se o MAPE (*Mean Absolute Percentage Error*) e o RMSPE (*Root Mean Square Percentage Error*).

O MAPE é a média do quociente entre o desvio absoluto (módulo da diferença entre o valor previsto e o valor medido no laboratório - valor real) e o valor medido no laboratório para cada observação. O valor de MAPE do modelo é 7,81%. O modelo obtido tem um valor de RMSPE de 10,24%.

4.2.2 Densidade do Núcleo

Definiu-se com a equipa quais as variáveis que poderiam afetar o valor final da densidade do núcleo. Assim, o passo seguinte foi elaborar um modelo, através de uma regressão linear múltipla, que permitisse identificar quais as variáveis do processo significativas para prever o valor da densidade do núcleo - representadas a verde na Figura 4.5.

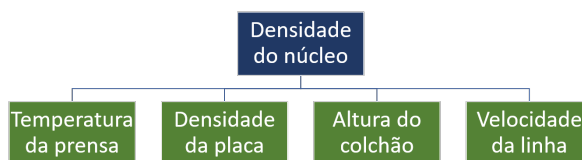


Figura 4.5: Diagrama de relações das variáveis que impactam a densidade do núcleo.

²Uma vez que a validação dos pressupostos segue sempre os mesmos passos, apenas se demonstra esse processo para os modelos de 1 nível e após PCA. No entanto, para todos os modelos desenvolvidos foi feita uma análise para averiguar se os pressupostos foram cumpridos.

Todas as variáveis identificadas pela equipa se revelaram significativas para prever o valor da densidade do núcleo. A temperatura da prensa, apesar de representada como uma única variável, é medida em cinco pontos da prensa.

Os pressupostos da regressão foram assegurados. A variável com maior influência na densidade do núcleo é a densidade da placa, como suportado pela análise da Figura D.4. 35,49% da variação do valor da densidade do núcleo é explicada pela densidade da placa.

Como o coeficiente de determinação do modelo obtido é 39,71%, pode-se concluir que esta variável é a mais significativa do modelo. O modelo tem um R^2 ajustado de 38,11% e um R^2 previsto de 35,96%.

Tal como a densidade do núcleo, o valor da densidade da placa é medido em laboratório. Logo, não é possível, através de uma análise das variáveis da linha de produção, obter diretamente o seu valor. Assim, decidiu-se estudar quais as variáveis significativas para prever o valor da densidade da placa.

O modelo obtido tem um MAPE de 1,87% e um RMSPE de 2,48%.

4.2.3 Densidade da Placa

Seguiu-se a mesma abordagem que se utilizou nos dois passos anteriores. A equipa, através da sua experiência e conhecimento, identificou as variáveis que poderiam influenciar o valor da densidade da placa e validaram-se essas hipóteses através de métodos estatísticos.

Realizou-se uma regressão linear múltipla com um nível de significância de 95%. Isto é, o modelo foi constituído por regressores com um p -value inferior a 0,05. Os resultados obtidos encontram-se no Apêndice E. Destacaram-se a verde, na Figura 4.6, as variáveis significativas para o modelo.

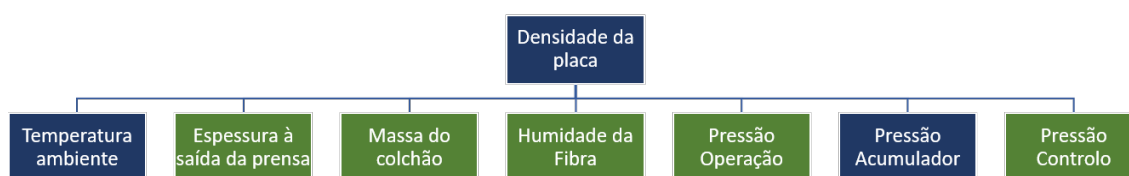


Figura 4.6: Diagrama de relações das variáveis que impactam a densidade da placa.

A validação dos pressupostos da regressão linear múltipla foi realizada e assegurada.

O modelo obtido tem um coeficiente de determinação de 13,48%, um R^2 ajustado de 11,23% e um R^2 previsto de 6,06%. Apesar de os coeficientes de determinação terem valores baixos, os erros suportam a qualidade do modelo obtido. Os valores de MAPE e RMSPE são 0,98% e 1,20%, respetivamente.

Uma vez que a densidade da placa depende apenas de variáveis mensuráveis durante o processo produtivo, não é necessário decompor o processo em mais níveis. Assim, o passo seguinte foi consolidar os três níveis num só, obtendo um único modelo preditivo para calcular o valor da resistência à tração através das variáveis do processo que podem ser medidas na linha.

4.2.4 Modelo 1 nível

A fim de elaborar um modelo que permitisse prever o valor da resistência à tração da placa durante o processo produtivo e fazer ajustes às variáveis do processo caso necessário, consolidaram-se os três modelos desenvolvidos anteriormente num só (Figura 4.7). No Apêndice F, apresenta-se com maior detalhe a junção dos três níveis de variáveis em que se dividiu a análise.

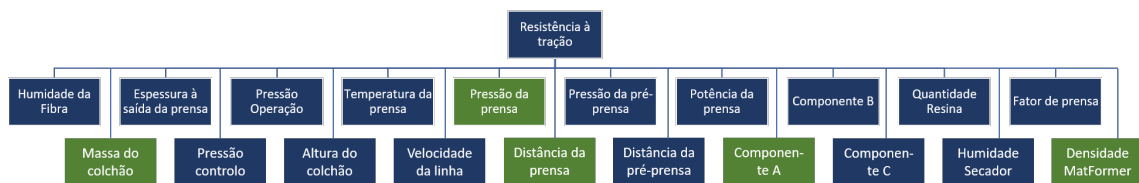


Figura 4.7: Diagrama de relações das variáveis do processo que impactam a resistência à tração da placa.

O resultado do modelo e a validação dos pressupostos referentes à regressão linear múltipla estão no Apêndice G. Não foi assegurado o pressuposto relativo à presença de multicolinearidade no modelo.

Verificou-se que algumas das variáveis significativas no primeiro modelo preditivo da resistência à tração (Figura 4.4) não o são quando se decompõe o modelo e se introduzem as variáveis do processo que influenciam a densidade do núcleo e a densidade da placa.

Das variáveis do processo relevantes nos modelos de densidade do núcleo e densidade da placa, apenas a massa do colchão se revelou significativa para o modelo elaborado. Assim, apesar de se ter introduzido um maior número de regressores para elaborar o modelo, o número de variáveis independentes significativas é inferior.

Ainda assim, obteve-se um modelo com um coeficiente de determinação mais elevado: 55,79%. Os valores de R^2 ajustado e previsto são 50,02% e 42,61%, respetivamente. Deste modo, é possível concluir que o modelo obtido tem uma maior capacidade de explicar o valor da resistência à tração.

O modelo tem um MAPE de 8,55% e um RMSPE de 10,87% (pior que nos 3 níveis). No entanto, não se pode comparar diretamente este valor com o obtido no primeiro modelo relativo à RT, uma vez que os três modelos (resistência à tração - densidade do núcleo - densidade da placa) estão encadeados.

Logo, o erro dos modelos da densidade da placa e da densidade no núcleo propagar-se-ia, penalizando o erro obtido no modelo da resistência à tração.

Considerou-se que, dado que o MAPE do modelo referente à resistência à tração é 7,81%, se pode calcular a precisão do modelo através da Equação 4.4

$$P_{RT} = 100\% - 7,81\% = 92,19\% \quad (4.4)$$

Replicou-se este raciocínio para os modelos da densidade do núcleo (Equação 4.5) e da densidade da placa (Equação 4.6).

$$P_{Nucleo} = 100\% - 1,97\% = 98,13\% \quad (4.5)$$

$$P_{Placa} = 100\% - 0,98\% = 99,02\% \quad (4.6)$$

Assim, dado que os três modelos estão encadeados, considerou-se que a melhor forma de determinar a sua precisão seria multiplicando os valores de precisão de cada nível, como na Equação 4.7.

$$P_{3Niveis} = 92,19\% * 98,13\% * 99,02\% = 89,58\% \quad (4.7)$$

A precisão do modelo de 1 nível foi calculada na Equação 4.8.

$$P_{Placa} = 100\% - 8,55\% = 91,45\% \quad (4.8)$$

Assim, comparando os valores obtidos na Equação 4.7 e na Equação 4.8, é possível concluir que o modelo de 1 nível, em que se prevê o valor da resistência à tração diretamente através das variáveis do processo, tem uma maior precisão. Isto é, obtém valores mais próximos do valor real.

Esta conclusão é suportada pela análise dos valores de RMSPE. O modelo que calcula o valor da resistência à tração elaborado com as variáveis representadas na Figura 4.4 tem um RMSPE de 10,24%. O modelo de 1 nível tem um RMSPE de 10,87%. Para além de não se tratar de uma diferença significativa (Saigal and Mehrotra, 2012), tendo em conta que o primeiro modelo tem mais dois níveis (que acrescem ao erro da previsão), pode-se concluir que a qualidade do modelo de 1 nível é superior.

4.3 Limitações dos modelos

4.3.1 Coeficiente de determinação

Na literatura, refere-se como valor alvo para o R^2 , no mínimo, 80% (Suryanarayana and Mistry, 2016). Logo, os modelos obtidos ficam aquém do desejado.

O valor baixo do Coeficiente de Determinação pode ser justificado pelo facto de se ter usado uma gama de valores operacional, isto é, dados históricos do processo. Uma vez que os operadores já têm conhecimento empírico acerca do funcionamento da linha produtiva, sabem em que gama de valores manter os parâmetros em estudo. Assim, as variáveis não estão a ser levadas ao extremo, tendo consequências negativas na qualidade do modelo.

Para demonstrar o que foi explicado no parágrafo anterior, realizou-se uma simulação recorrendo ao *dataset* utilizado para elaborar o modelo preditivo da densidade da placa. Na primeira etapa, os dados foram divididos em três níveis, como representado na Tabela 4.1.

Tabela 4.1: Divisão das variáveis em níveis pela sua gama de valores.

Nível	Espessura à Saída	Massa Colchão	Humidade Fibra
Baixo	13,5 a 15	9,8 a 11,5	7,1 a 8,9
Médio	15 a 17,5	12,1 a 14,2	6,6 a 8,8
Alto	17,5 a 20	10,1 a 15,9	6 a 8,5

Retiraram-se aleatoriamente 25 amostras de cada nível, constituindo um dataset com 75 observações. Este foi introduzido no Minitab e realizou-se uma regressão linear múltipla. Obteve-se um modelo com as características da Tabela 4.2:

Tabela 4.2: Coeficiente de determinação do modelo com valores dos 3 níveis.

S	R^2	R^2_{aj}	R^2_{pred}
7,41327	28,77%	25,77%	20,55%

De seguida, seleccionaram-se 75 observações pertencentes ao nível baixo. Assim, restringiu-se a gama de valores representada na amostra para estudar a sua influência no R^2 do modelo obtido, que se pode observar na Tabela 4.3.

Tabela 4.3: Coeficiente de determinação do modelo com valores do nível baixo.

S	R^2	R^2_{aj}	R^2_{pred}
8,09598	13,48%	11,07%	4,96%

Através da análise dos resultados obtidos para os dois modelos é possível concluir que, de facto, é relevante utilizar uma gama de valores mais abrangente para poder analisar o comportamento do processo quando levado ao seu limite. Para tal, o recomendado é usar uma gama de valores experimental, ou seja, realizar experiências em que se levam as variáveis em estudo ao limite. No entanto, não foi possível fazê-lo devido à interrupção na atividade da fábrica.

4.3.2 Multicolinearidade

Nos modelos obtidos, identificou-se a existência de multicolinearidade entre regressores. Esta relação entre as variáveis deve-se a uma restrição do modelo, o que é comum acontecer quando se estão a estudar processos produtivos. As variáveis do processo estão relacionadas, por exemplo, quanto maior a massa do colchão, maior a espessura do colchão à saída da prensa. Logo, é natural que haja multicolinearidade no modelo (Montgomery et al., 2012).

A presença de multicolinearidade no modelo significa que este tem informação redundante. Este fenómeno vai impactar a determinação dos coeficientes de cada regressor, que serão menos precisos. Logo, a interpretação desses parâmetros poderá ser enviesada.

Para melhorar a qualidade do modelo e resolver os problemas relacionados com a multicolinearidade, optou-se por implementar a metodologia PCA.

4.4 *Principal component analysis*

Num processo contínuo, a alteração do valor de uma das variáveis *input* pode levar a uma mudança na forma como outras variáveis do processo se relacionam. Logo, em um projeto de melhoria de um processo contínuo, a implementação de *Principal Component Analysis* (PCA) é vantajosa, na medida em que permite detetar desvios nas relações entre as variáveis do processo (Hild et al., 2000).

Assim, recorrendo ao Minitab, identificaram-se os principais componentes dos regressores incluídos no modelo.

A análise da Figura G.4 permitiu identificar que as distâncias e pressões medidas nos vários pontos da prensa eram as variáveis com maior impacto na previsão do valor da resistência à tração. Na Figura G.8 é possível perceber que a multicolinearidade se foca nas distâncias e pressões da prensa. Por esse motivo, decidiu-se realizar uma análise PCA a essas variáveis independentes.

O número de componentes a considerar é uma decisão crucial. Caso seja demasiado reduzido, poderá haver informação essencial que não está a ser tida em conta na análise. Se se incluírem demasiados componentes, estar-se-á a introduzir ruído na análise (Josse and Husson, 2012).

O critério para determinar o número de componentes a considerar foi a percentagem de variação representada pelos mesmos. De acordo com a literatura, os componentes devem ter capacidade de explicar, pelo menos, 80% da variação dos dados.

O primeiro passo foi analisar as distâncias da prensa e identificar os componentes principais. A distância da prensa é medida em 34 pontos, obtendo-se 34 variáveis. A análise PCA permite transformá-las em 10 variáveis que explicam 81,4% da variação registada.

Repetiu-se este processo para as pressões da prensa. A pressão é medida em 53 pontos da prensa, resultando em 53 variáveis distintas. Recorrendo aos 10 principais componentes, é possível explicar 81,9% da variação registada.

Substituíram-se as distâncias e as pressões da prensa por 10 componentes principais, respetivamente. De seguida, elaborou-se um modelo através de uma regressão linear múltipla.

O modelo obtido tem um coeficiente de determinação de 38,38%, um R² ajustado de 34,55% e um R² previsto de 28,44%, pior do que no modelo anterior. No entanto, as estatísticas de erro apresentam valores melhores: MAPE de 7,40% e RMSPE de 9,58%.

Os resultados obtidos e a validação dos pressupostos encontram-se no Apêndice H. O modelo elaborado respeita todos os pressupostos referentes à regressão linear múltipla, pelo que é possível concluir que a implementação de PCA foi vantajosa e permitiu melhorar a qualidade do modelo.

Capítulo 5

Resultados

5.1 Estabilização do processo produtivo

Após o desenvolvimento do modelo preditivo, revelou-se pertinente criar normas para as variáveis da linha. Isto é, estabelecer valores fixos para algumas e limitar o intervalo de valores de outras. Esta medida vai permitir reduzir a variabilidade do processo e, consequentemente, a variabilidade dos resultados obtidos.

Para determinar os valores ótimos, recorreu-se ao *software RStudio* para elaborar uma árvore de regressão. O resultado obtido encontra-se na Figura 5.1.

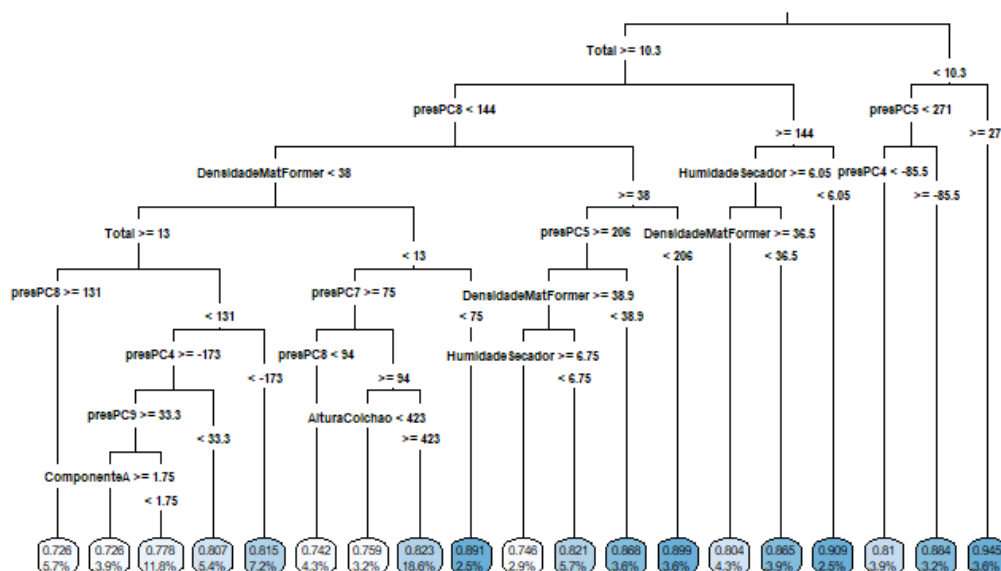


Figura 5.1: Árvore de regressão para determinar o valor da resistência à tração.

O resultado foi uma árvore com 19 nós terminais. Das 11 variáveis independentes introduzidas, apenas uma (quantidade de resina utilizada) não foi incluída nos critérios de decisão da árvore.

Tomou-se a decisão de utilizar uma árvore de regressão para determinar a melhor gama de valores para as variáveis da linha estatisticamente significativas devido a três razões principais.

Em primeiro lugar, pelo facto de se possuir um elevado volume de dados, o que permite tirar conclusões fidedignas desta análise, já que se trata de uma amostra significativa.

Por outro lado, a alternativa de fazer desenho de experiências (*Design Of Experiences* - DOE) com tantas variáveis seria muito complexo, dispendioso e demorado, além de ser inviável.

Por último, devido ao contexto durante a elaboração da dissertação, que levou à interrupção temporária da atividade da empresa, não foi possível realizar testes na linha/presenciais. Assim, a árvore de regressão permitiu limitar a gama de valores dos parâmetros significativos.

Um dos objetivos do projeto foi reduzir a quantidade de matéria prima utilizada (correspondente à variável "Total"), uma vez que foi identificado um sobreconsumo que implicava uma diminuição na margem operacional da empresa A.

Através da análise da Figura 5.1, é possível identificar que não se deve utilizar menos de 10,3 kg/m², já que se obteria valores de resistência à tração superiores à especificação do cliente. Por outro lado, uma vez que se pretende reduzir o consumo de matéria prima, aconselha-se a que não se ultrapassem os 13 kg por m² produzido.

Assim, definiu-se como gama de valores para a quantidade de MP utilizada entre 10,3 e 13 kg/m².

5.2 Impacto financeiro da implementação do modelo

Para que se justifique alterar um processo, é necessário que os benefícios do novo processo ultrapassem os custos inerentes à mudança (Kumar and Sosnoski, 2009).

Segundo Montgomery (2009), o impacto de um projeto deve ser avaliado de acordo com o benefício financeiro que vai trazer à empresa, permitindo medir de forma objetiva o sucesso do projeto. Atualmente, os custos adicionais de produção dividem-se em três vertentes: sobreconsumo de matéria prima, requalificação e desqualificação.

Na empresa A, em 2019, tendo em conta o segmento de produtos que foi alvo de estudo na presente dissertação, a taxa de sobreconsumo de matéria prima (MP) foi 6,47%, a taxa de requalificação foi de 0,31% e a taxa de desqualificação foi de 0,18%. Na Tabela 5.1 está representado o impacto anual de cada tipo de custo por ponto percentual.

Tabela 5.1: Custos adicionais de produção.

Sobreconsumo de MP (k€/%)	Requalificação (k€/%)	Desqualificação (k€/%)
45,74	28,15	58,64

Os custos de requalificação e desqualificação (custos de rejeição) são referentes à matéria prima incorporada até ao momento da rejeição e aos custos de produção até ao momento de rejeição.

Na Equação 5.1 calcularam-se os custos adicionais de produção referentes a 2019. Este valor vai ser a referência para determinar qual o benefício que o projeto desenvolvido vai proporcionar.

$$Custos_{2019} = 45,47 * 6,47\% + 28,15 * 0,31\% + 58,64 * 0,18\% = 315,22 \text{ k€} \quad (5.1)$$

Analisando os resíduos (diferença entre o valor real e o valor previsto) do modelo obtido após PCA, verificou-se que tinham um valor esperado de 0,02 e um desvio padrão de 0,07. Ou seja, se se ajustar a linha produtiva para um valor de resistência à tração de 0,68Pa, o valor previsto da previsão vai ser 0,70Pa com um desvio padrão de 0,07.

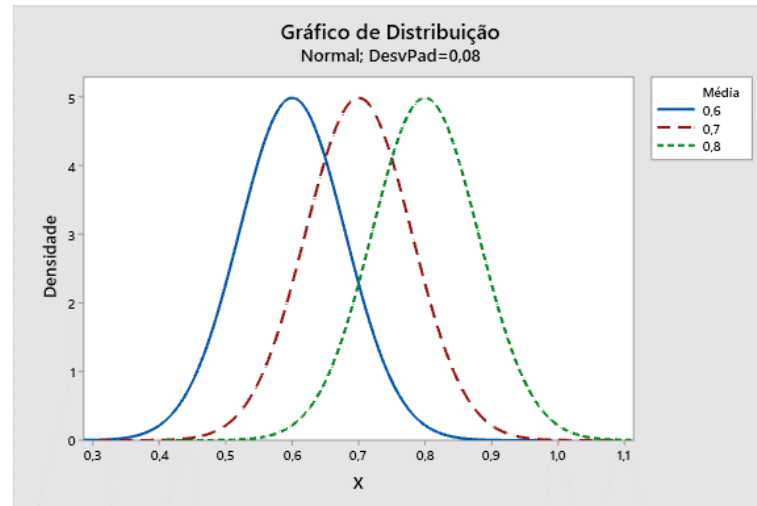


Figura 5.2: Variação da curva da distribuição normal de acordo com o valor esperado.

Pretende-se determinar para que valor objetivo ajustar a linha, de modo a minimizar os custos adicionais de produção e a ter maior benefício. A Figura 5.2 permite perceber a variação dos resultados obtidos face ao valor objetivo definido.

Após a elaboração do modelo, verificou-se a normalidade dos erros (Apêndice H). Através da Equação 5.2 pode-se concluir que, como os erros seguem uma distribuição normal e o valor previsto é constante ¹, os valores reais também vão estar normalmente distribuídos.

$$ValorReal = ValorPrevisto + Erro \quad (5.2)$$

Se o valor da resistência à tração for inferior a 0,70Pa, as placas vão ser rejeitadas (requalificadas ou desqualificadas). Admitindo que os valores seguem uma distribuição normal, é possível calcular a probabilidade de as placas serem rejeitadas. Tendo em conta o volume de vendas correspondente a cada espessura em 2019, é possível dividir a taxa de rejeição em taxa de desqualificação (T_D) e taxa de requalificação (T_R).

Para a produção das placas que constituem a amostra estudada neste projeto, utilizaram-se, em média, 12,83 kg de MP por m² produzido. Através da árvore de decisão, definiu-se que os valores correspondentes à variável Total deviam variar entre 10,3 e 13 kg/m².

¹ Considera-se que o valor previsto vai ser igual ao valor esperado da previsão

Uma vez que a taxa de sobreconsumo está diretamente relacionada com a quantidade de matéria prima utilizada, é possível calcular qual a taxa de sobreconsumo correspondente à gama de valores indicada (Figura 5.3).



Figura 5.3: Taxas de sobreconsumo associadas à gama de valores definida.

Assim, desenvolveu-se uma função (Equação 5.3) que calcula os custos adicionais de produção para vários cenários. Sendo que se pretende eleger o cenário que minimiza o valor da função.

$$Custos = 45,47 * T_S + 28,15 * T_R + 58,64 * T_D \quad (5.3)$$

No caso de se utilizar o limite máximo de MP definido, os custos de sobreconsumo vão ser de 300k€ por ano. Se, em média, se utilizar 10,3kg por m² produzido, estima-se que os custos de sobreconsumo corresponderão a 237k€ anuais.

Valor Objetivo	Valor Esperado	Desvio Padrão	%Desqualificação	%Requalificação	Desqualificação (k€/ Ano)	Requalificação (k€/ Ano)	Custo Rejeição K€/ano
0,7	0,72	0,07	14,99%	19,54%	878,93 €	550,19 €	1 429,12 €
0,71	0,73	0,07	12,73%	16,60%	746,68 €	467,40 €	1 214,08 €
0,72	0,74	0,07	10,65%	13,89%	624,48 €	390,91 €	1 015,38 €
0,73	0,75	0,07	8,76%	11,43%	513,92 €	321,70 €	835,62 €
0,74	0,76	0,07	7,09%	9,25%	416,00 €	260,41 €	676,40 €
0,75	0,77	0,07	5,65%	7,36%	331,08 €	207,25 €	538,33 €
0,76	0,78	0,07	4,42%	5,76%	258,99 €	162,12 €	421,11 €
0,77	0,79	0,07	3,39%	4,43%	199,06 €	124,60 €	323,66 €
0,78	0,80	0,07	2,56%	3,34%	150,28 €	94,07 €	244,36 €
0,79	0,81	0,07	1,90%	2,48%	111,42 €	69,75 €	181,17 €
0,8	0,82	0,07	1,38%	1,80%	81,10 €	50,77 €	131,87 €
0,81	0,83	0,07	0,99%	1,29%	57,95 €	36,28 €	94,23 €
0,82	0,84	0,07	0,69%	0,90%	40,64 €	25,44 €	66,07 €
0,83	0,85	0,07	0,48%	0,62%	27,96 €	17,50 €	45,46 €
0,84	0,86	0,07	0,32%	0,42%	18,87 €	11,81 €	30,69 €
0,85	0,87	0,07	0,21%	0,28%	12,50 €	7,82 €	20,32 €
0,86	0,88	0,07	0,14%	0,18%	8,12 €	5,08 €	13,20 €
0,87	0,89	0,07	0,09%	0,11%	5,17 €	3,24 €	8,40 €
0,88	0,90	0,07	0,06%	0,07%	3,23 €	2,02 €	5,25 €
0,89	0,91	0,07	0,03%	0,04%	1,98 €	1,24 €	3,21 €
0,9	0,92	0,07	0,02%	0,03%	1,19 €	0,74 €	1,93 €
0,91	0,93	0,07	0,01%	0,02%	0,70 €	0,44 €	1,14 €
0,92	0,94	0,07	0,01%	0,01%	0,40 €	0,25 €	0,66 €
0,93	0,95	0,07	0,00%	0,01%	0,23 €	0,14 €	0,37 €
0,94	0,96	0,07	0,00%	0,00%	0,13 €	0,08 €	0,21 €
0,95	0,97	0,07	0,00%	0,00%	0,07 €	0,04 €	0,11 €
0,96	0,98	0,07	0,00%	0,00%	0,04 €	0,02 €	0,06 €
0,97	0,99	0,07	0,00%	0,00%	0,02 €	0,01 €	0,03 €
0,98	1,00	0,07	0,00%	0,00%	0,01 €	0,01 €	0,02 €
0,99	1,01	0,07	0,00%	0,00%	0,00 €	0,00 €	0,01 €
1	1,02	0,07	0,00%	0,00%	0,00 €	0,00 €	0,00 €

Figura 5.4: Custos de rejeição de vários cenários.

Tendo como *baseline* o valor gasto em custos adicionais de produção no ano anterior (315k€), calculou-se qual o valor mínimo de resistência à tração que se deve ter como objetivo para gastar menos do que em 2019.

Quanto mais alto for o valor previsto pelo modelo, menores serão as taxas de desqualificação e requalificação. Na Figura 5.4 identificaram-se os valores mínimos, tendo em conta os gastos de sobreconsumo.

Se, em média, se utilizar 13 kg por m² produzido, para ter benefícios, ter-se-á de definir os parâmetros da linha a fim de obter, no mínimo, uma resistência à tração de 0,86Pa (linha a verde). Caso se utilize, em média, 10,3 kg por m² produzido, deve-se parametrizar a linha a fim de obter, no mínimo, uma resistência à tração de 0,82Pa (linha a amarelo).

Os custos de requalificação e desqualificação obtidos são superiores ao que serão quando se implementar o modelo. Devido à limitação da gama de valores aceitável para algumas variáveis e a fixação do valor de outras, a variabilidade do processo e dos resultados vai reduzir.

Assim, o modelo vai ter uma maior capacidade para prever o valor da resistência à tração. Logo, os erros vão ter um menor desvio padrão, pelo que as taxas de requalificação e desqualificação também serão inferiores, refletindo-se em custos mais baixos.

No entanto, face à impossibilidade de prever o valor do desvio padrão, utilizou-se o desvio padrão do modelo. O cenário obtido é mais pessimista do que será a realidade, pelo que a obtenção de benefícios neste cenário significa que as mais valias provenientes do projeto vão ser ainda mais vantajosas para a empresa A.

Capítulo 6

Conclusão

Pelo facto de, em indústria de processo, a intervenção dos operadores corresponder a funções de suporte, melhorar o processo produtivo tem um impacto direto nos resultados obtidos pela organização.

6.1 Plano de ações

Devido ao contexto em que foi desenvolvido o projeto, não foi possível realizar várias análises fundamentais para o seu sucesso. No entanto, face à perspectiva de regressar à empresa, é conveniente traçar um plano de ações a levar a cabo aquando desse regresso.

Numa primeira fase, será fundamental realizar testes no laboratório. Recolher amostras para fazer o teste ANOVA para validar que a qualidade da placa se mantém longitudinalmente e realizar os testes R&R para garantir que o sistema de medição utilizado é capaz. Se alguma das hipóteses for rejeitada, será necessário tomar medidas para a corrigir. Se ambas as hipóteses forem confirmadas, deverá seguir-se para uma fase de avaliação do modelo desenvolvido. Este deverá ser implementado e confrontar os valores previstos com os valores medidos em laboratório. Se se detetar que o modelo tem desvios significativos face aos valores reais, deve-se recolher mais amostras para poder melhorar a qualidade de previsão do modelo.

Para reduzir a variabilidade do processo, será vantajoso aplicar desenho de experiências nas variáveis mais significativas para otimizar esses parâmetros.

6.2 Perspetivas de trabalho futuro

6.2.1 *Improve*

A aplicação de desenho de experiências nas variáveis identificadas como significativas iria permitir estabilizar o modelo e determinar gamas de valores ótimas para esses parâmetros.

Relativamente às variáveis com menor impacto na qualidade do produto final, propõe-se fixar o seu valor, de modo a diminuir a variabilidade do processo e dos seus resultados.

Para tal, a criação de normas e *standards* relativamente aos valores das variáveis de produção seria vantajosa. Através de gestão visual, é possível criar um sistema que alerte os operadores para desvios face aos valores pretendidos e que lhes permita detetar e fazer os ajustes necessários mais rapidamente.

Os modelos foram obtidos através de regressões lineares múltiplas. No entanto, esta pode não ser a melhor abordagem se as relações entre as variáveis não forem lineares ou se houver interações complexas entre as variáveis.

Estes dois problemas são frequentes em processos contínuos. Através de ferramentas de *Machine Learning* é possível melhorar a qualidade dos modelos.

6.2.2 Control

Até ao momento do estudo, são registados os valores dos parâmetros da prensa em tempo real. A integração do modelo nesse sistema permitiria, face aos valores reais das variáveis da linha, prever a qualidade do produto final. Caso se identificasse um desvio relativamente ao valor objetivo, seria lançado um alerta para que os operadores pudessem ajustar os parâmetros da linha, assegurando qualidade à primeira.

A implementação de um *andon* corresponde a um sinal luminoso que se aciona para assinalar uma situação indesejada. Neste caso, seria uma luz que se ativaria quando o modelo previsse que, face às condições da linha naquele momento, as placas produzidas não iriam cumprir os requisitos de qualidade acordados.

Para garantir a consistência dos resultados obtidos e que as melhorias implementadas se mantêm e são suficientes, deve-se implementar um sistema de SPC - *Statistical Process Control*. Verificar a eficiência do processo mensalmente permitirá identificar eventuais problemas e resolvê-los mais rapidamente.

Através da monitorização das cartas de controlo será possível identificar causas assinaláveis de variação e proceder à sua eliminação, melhorando a *performance* do processo (MacGregor and Kourti, 1995). Trata-se de um método preventivo e não reativo, evitando a propagação dos efeitos dos problemas que possam surgir.

6.3 Conclusões

O presente caso de estudo, que se insere no setor dos derivados de madeira, tinha como principal objetivo reduzir a taxa de desqualificação devido ao valor da resistência à tração dos produtos finais.

Através de regressões lineares múltiplas entre as variáveis envolvidas no processo foi possível identificar quais as variáveis *input* e de controlo estatisticamente significativas para a qualidade do produto final (*output*), com um nível de confiança de 95%.

Concluiu-se que, tal como identificado pelos operadores, a quantidade de matéria prima utilizada tem um efeito significativo no valor da resistência à tração. Recorrendo a uma árvore de decisão determinou-se uma gama de valores para esta variável. O objetivo foi assegurar a qualidade

do produto final e, em simultâneo, reduzir os custos adicionais de produção, nomeadamente os associados ao sobreconsumo de matéria prima.

O modelo desenvolvido permite, face às condições da linha de produção, prever o valor final da resistência à tração e, caso necessário, identificar os ajustes a fazer aos parâmetros produtivos. Assim, confere maior capacidade de controlo aos operadores, reduzindo a imprevisibilidade do processo. No entanto, não foi possível testar o modelo, não sendo possível verificar se os objetivos traçados inicialmente foram atingidos.

Com a introdução desta ferramenta no sistema de controlo da empresa será possível diminuir as taxas de rejeição (requalificação e desqualificação) e controlar o sobreconsumo de matéria prima. A obtenção de bons resultados na gama de produtos que foi alvo de melhoria vai motivar a replicação da metodologia adotada para outras famílias de produtos da empresa A e para outras sucursais do grupo a que a organização pertence.

Bibliografia

- Aba, Eli K.; Hayden, M. A. (2013). Heart of the matter. *Fortune*, 168(8):24.
- Adnan, N., Ahmad, M. H., and Adnan, R. (2006). A Comparative Study On Some Methods For Handling Multicollinearity Problems The Problem of Multicollinearity Methods for Handling Multicollinearity. 22(2):109–119.
- Aguayo, R. (2010). *Dr Deming: The american who taught the japanese about quality*.
- AIMMP (2020). Associação das Indústrias de Madeira e Mobiliário de Portugal.
- Alfons, A., Croux, C., and Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Annals of Applied Statistics*, 7(1):226–248.
- Arnheiter, E. D. and Maleyeff, J. (2005). The integration of Lean Management and Six Sigma. 17(1):5–18.
- Borges, F. H. and Dalcol, P. R. T. (2002). Indústrias de Processo: Comparações e Caracterizações.
- Chen, J. R., Lin, Y. H., and Leu, Y. G. (2018). Predictive model based on decision tree combined multiple regressions. *ICNC-FSKD 2017 - 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, (1):1855–1858.
- Chen, K.-s., Wang, K.-j., and Chang, T.-c. (2017). A novel approach to deriving the lower confidence limit of indices C_{pu} , C_{pl} , and C_{pk} in assessing process capability. 7543.
- Choi, W., Paulson, S. E., Casmassi, J., and Winer, A. M. (2013). Evaluating meteorological comparability in air quality studies: Classification and regression trees for primary pollutants in California’s South Coast Air Basin. *Atmospheric Environment*, 64(x):150–159.
- Clapp, N. E., Young, T. M., and Guess, F. M. (2008). Predictive modeling the internal bond of medium density fiberboard using a modified principal component analysis. *Forest Products Journal*, 58(4):49–55.
- Deming, W. E. (2000). *The New Economics for Industry, Government, Education*.
- Dobbin, K. K. and Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC Medical Genomics*, 4(1):31.
- Excel, S. f. (2013). Destructive Gage R&R Analysis.
- Fransoo, J. C. and Rutten, W. G. M. M. (1993). A Typology of Production Control Situations in Process Industries. pages 47–57.

- George, M. L., Rowlands, D., Price, M., and Maxey, J. (2005). *The Lean Six Sigma Pocket Toolbook*. McGraw-Hill.
- Gins, G., Vanlaer, J., and Impe, J. V. (2013). The Impact of Process Variability on Statistical Process Monitoring. *2013 Conference on Control and Fault-Tolerant Systems (SysTol)*, pages 306–311.
- Giorgi, E. (2018). On the goodness-of-fit of generalized linear geostatistical models. *Spatial Statistics*, 28:79–83.
- Green, F. B. (2006). Six-sigma and the revival of TQM. *Total Quality Management and Business Excellence*, 17(10):1281–1286.
- Gygi, C., DeCarlo, N., and Williams, B. (2005). *Six Sigma for Dummies*.
- Hayes, T., Usami, S., Jacobucci, R., and Mcardle, J. J. (2016). Using Classification and Regression Trees (CART) and Random Forests to Analyze Attrition : Results From Two Simulations. 30(4):911–929.
- Hild, C., Sanders, D., and Cooper, T. (2000). Six Sigma on continuous processes: How and why it differs. *Quality Engineering*, 13(1):1–9.
- Imai, M. (1986). *Kaizen, The Key to Japan's Competitive Success*.
- Imai, M. (2012). *Gemba Kaizen - A Commonsense Approach to Continuous Improvement Strategy*.
- Johnston, A. B., Maguire, L., and McGinnity, M. (2008). Disentangling causal relationships of a manufacturing process using genetic algorithms and Six Sigma techniques. *International Journal of Production Research*, 46.
- Jones, D. M., Watton, J., and Brown, K. J. (2005). Comparison of hot rolled steel mechanical property prediction models using linear multiple regression, non-linear multiple regression and non-linear artificial neural networks. *Ironmaking and Steelmaking*, 32(5):435–442.
- Josse, J. and Husson, F. (2012). Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistics and Data Analysis*, 56(6):1869–1879.
- Kadlec, P., Gabrys, B., and Strandt, S. (2009). Data-driven Soft Sensors in the Process Industry. 33:795–814.
- Kaizen Institute (2017). Total Quality Management.
- Kumar, S. and Sosnoski, M. (2009). Using DMAIC Six Sigma to systematically improve shopfloor production quality and costs. *International Journal of Productivity and Performance Management*, 58(3):254–273.
- Lawrence, R. L. and Wright, A. (2001). Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis. *Photogrammetric Engineering & Remote Sensing*, 67(10):1137–1142.
- MacGregor, J. F. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3):403–414.

- Mgbere, O., Singh, M., and Arafat, R. (2015). Measurement system analysis (MSA) of empirically derived composite measure of preventive care counseling practices of HIV medical care providers. *International Journal of Medical Science and Public Health*, 4(12):1640.
- Moksony, F. (2014). Small is beautiful, Interpretation of R² in social research. *Szociologiai Szemle*, Special is(January 1999):130–138.
- Montgomery, D. (2009). *Introduction To Statistical Quality Control*, volume 10. John Wiley & Sons <https://www.spcforexcel.com/knowledge/measurement-systems-analysis/destructive-gage-rr-analysis> <https://www.spcforexcel.com/knowledge/measurement-systems-analysis/destructive-gage-rr-analysis> <https://www.spcforexcel.com/knowledge/measurement-systems-analysis/destructive-gage-rr-analysis>
- Montgomery, D., Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. 5th edition.
- Montgomery, D. and Runger, G. (2013). *Applied Statistics and Probability for Engineers*. 6 edition.
- Mundfrom, D. J., Smith, M. D., and Kay, L. W. (2018). The effect of Multicollinearity on Prediction in Regression Models. (January):23–28.
- Naghibi, S. A., Pourghasemi, H. R., and Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental Monitoring and Assessment*, 188(1):1–27.
- Neter, J., Wasserman, W., and Kutner, M. H. (1983). *Applied Linear Regression Models*.
- Portugal, D. d. E. d. B. d. (2016). Análise das empresas dos setores da madeira, da cortiça e do papel. Technical report.
- Pyzdek, T. (2003). *The Six Sigma Handbook: The Complete Guide for Greenbelts, Blackbelts, and Managers at All Levels*, volume 44. McGraw-Hill, 2nd edition.
- Roggo, Y., Pauli, V., Jelsch, M., Pellegatti, L., Elbaz, F., Ensslin, S., Kleinebudde, P., and Krumme, M. (2020). Continuous manufacturing process monitoring of pharmaceutical solid dosage form: A case study. *Journal of Pharmaceutical and Biomedical Analysis*, 179:112971.
- Saigal, S. and Mehrotra, D. (2012). Performance Comparison of Time Series Data Using Predictive Data Mining Techniques. *Advances in Information Mining*, 4(1):57–66.
- Salmerón, R., García, J., García, C. B., and Martín, M. M. (2017). A note about the corrected VIF. *Statistical Papers*, 58(3):929–945.
- Saunders, L. J., Russell, R. A., and Crabb, D. P. (2012). The Coefficient of Determination : What Determines a Useful R² Statistic? 53(11):6830–6832.
- Shahriari, K. and Ding, F. (2011). Process variability and inherent efficiency enhancement in industrial processes: Two case studies in pulp and paper industry. *2011 IEEE International Conference on Control Applications (CCA)*, pages 1262–1269.
- Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*.
- Smętkowska, M. and Mrugalska, B. (2018). Using Six Sigma DMAIC to improve the quality of the production process : a case study. *Procedia - Social and Behavioral Sciences*, 238:590–596.

- Suryanarayana, T. and Mistry, P. (2016). Methodology. pages 45–49.
- Telmo, C., Lousada, J., and Moreira, N. (2010). Proximate analysis, backwards stepwise regression between gross calorific value, ultimate and chemical analysis of wood. *Bioresource Technology*, 101(11):3808–3815.
- Teunter, R. H., Babai, M. Z., and Syntetos, A. A. (2010). ABC Classification: Service levels and inventory costs. *Production and Operations Management*, 19(3):343–352.
- Tjahjono, B., Ball, P., Vitanov, V. I., Scorzafave, C., Nogueira, J., Calleja, J., Minguet, M., Narasimha, L., Rivas, A., Srivastava, A., Srivastava, S., and Yadav, A. (2010). Six sigma: A literature review. *International Journal of Lean Six Sigma*, 1(3):216–233.
- Toledo, J. C. d., Ferro, J. R., and Truzzi, O. M. S. (1986). Indústrias de processo contínuo : novos rumos para a organização do trabalho. 26(1):103–105.
- Wuest, T., Weimer, D., Irgens, C., and Thoben, K. D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production and Manufacturing Research*, 4(1):23–45.
- Yang, Y. (2019). *Application Research of Process Capability Analysis in Manufacturing Quality Control*. Springer International Publishing.
- Zehir, C., Ertosun, G., Zehir, S., and Müceldilli, B. (2012). Total Quality Management Practices' Effects on Quality Performance and Innovative Performance. *Procedia - Social and Behavioral Sciences*, 41:273–280.

Anexo A

Simulação de dados para realizar teste ANOVA

As localizações 1 e 3 correspondem às extremidades da placa e a localização 2 corresponde ao centro (longitudinalmente) da placa.

Tabela A.1: Dados utilizados na simulação do teste ANOVA.

Localização	Densidade da Placa (kg/m ³)	Resistência à Tração (Pa)
1	700,0	0,73
1	683,2	0,68
1	695,8	0,77
1	707,0	0,74
1	692,3	0,72
1	705,6	0,78
1	702,8	0,73
1	712,6	0,71
1	697,2	0,73
1	720,3	0,77
1	681,8	0,72
1	705,6	0,75
1	684,6	0,73
1	695,1	0,69
1	689,5	0,75
1	708,4	0,72
1	710,5	0,77
1	690,2	0,77
1	710,5	0,70
1	705,6	0,78
1	680,4	0,72
1	702,8	0,71
1	680,4	0,73
1	718,9	0,77
1	716,1	0,73
2	720,3	0,70
2	706,3	0,67

Continua na página seguinte

Tabela A.1 – continua da página anterior

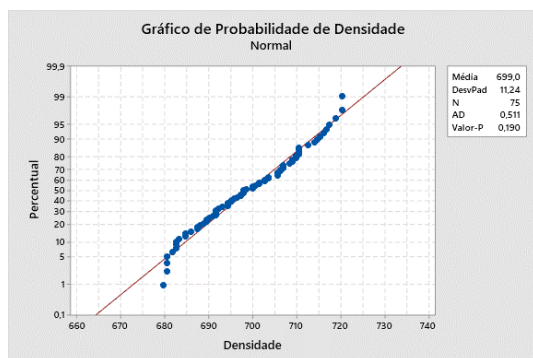
Localização	Densidade da Placa (kg/m³)	Resistência à Tração (Pa)
2	709,8	0,77
2	700,7	0,76
2	696,5	0,70
2	697,9	0,68
2	691,6	0,78
2	701,4	0,73
2	694,4	0,71
2	697,9	0,75
2	682,5	0,77
2	691,6	0,68
2	689,5	0,73
2	707,0	0,72
2	716,8	0,68
2	694,4	0,76
2	714,0	0,70
2	694,4	0,72
2	703,5	0,78
2	684,6	0,69
2	691,6	0,74
2	701,4	0,78
2	709,8	0,68
2	682,5	0,73
2	697,2	0,76
3	710,5	0,72
3	706,3	0,77
3	695,1	0,75
3	709,1	0,72
3	687,4	0,68
3	715,4	0,73
3	691,6	0,74
3	686,0	0,73
3	714,7	0,74
3	710,5	0,70
3	682,5	0,77
3	687,4	0,77
3	693,0	0,76
3	709,1	0,74
3	700,0	0,71
3	688,8	0,74
3	697,9	0,74
3	707,0	0,71
3	703,5	0,78
3	717,5	0,74
3	680,4	0,72
3	690,9	0,67
Continua na página seguinte		

Tabela A.1 – continua da página anterior

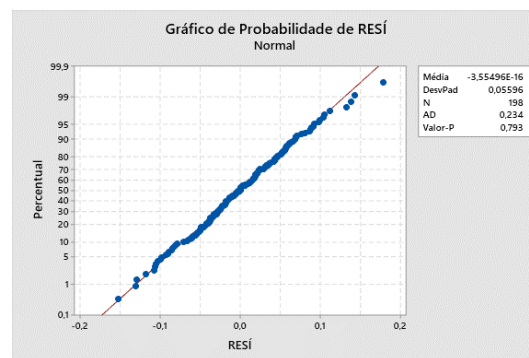
Localização	Densidade da Placa (kg/m³)	Resistência à Tração (Pa)
3	698,6	0,70
3	688,1	0,72
3	679,7	0,74

Anexo B

Gráficos para validar os pressupostos referentes aos erros do teste ANOVA

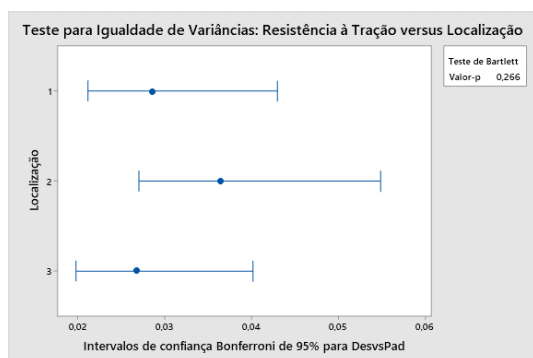


(a) Densidade da placa

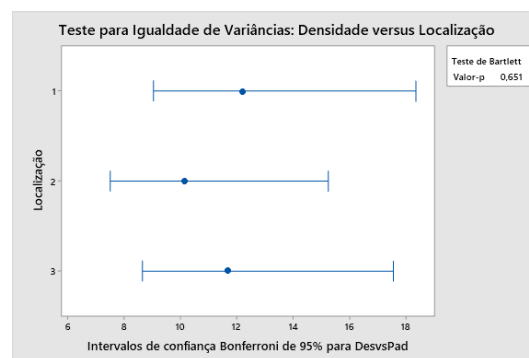


(b) Resistência à tração

Figura B.1: Teste à normalidade.



(a) Densidade da placa



(b) Resistência à tração

Figura B.2: Teste à homogeneidade da variância.

Anexo C

Resultados do modelo da resistência à tração

C.1 Resultados do modelo

Equação de Regressão

Resistência à Tração = 0,190 + 0,001713 Densidade Núcleo - 0,01695 FatorX - Resina
+ 0,03118 Dist_Esq_Frame11-13 + 0,003364 Dist_Esq_Frame19-20
- 0,01358 Dist_Esq_Frame24-25 - 0,02674 Dist_Dir_Frame9-10
- 0,01117 Pres_Esq_Frame19-20 + 0,00727 Pres_Esq_Frame28-29
- 0,00491 Pres_Esq_Frame30-32 + 0,000783 Pres_Cen_Frame 2
- 0,001859 Pres_Cen_Frame9-10 - 0,001563 Pres_Cen_Frame30-32
+ 0,00306 Pres_Cen_Frame33 + 0,00327 Pres_Dir_Frame17-18
+ 0,00985 Pres_Dir_Frame19-20 + 0,002040 Pres_Dir_Frame26-27
- 0,00740 Pres_Dir_Frame28-29 + 0,00620 Pres_Dir_Frame30-32
- 0,00461 Pres_Dir_Frame33 + 0,00476 Densidade MatFormer
- 0,03406 Humidade Secador

Figura C.1: Equação da regressão linear múltipla - Resistência à tração.

Sumário do Modelo

S	R2	R2(aj)	PRESQ	R2(pred)	AICc	BIC
0,0592091	53,17%	47,59%	0,801152	39,20%	-528,44	-459,16

Figura C.2: Sumário do modelo - Resistência à tração.

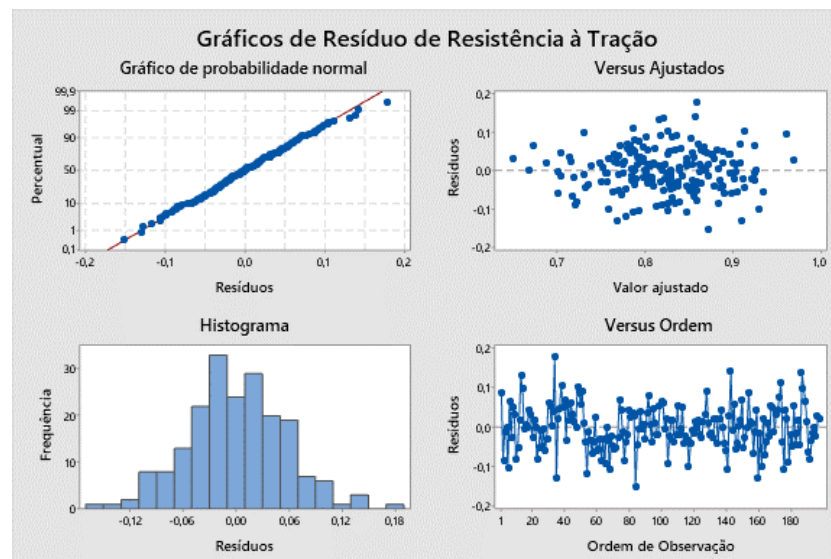


Figura C.3: Gráficos de resíduos - Resistência à tração.

Análise de Variância

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	21	0,70067	53,17%	0,70067	0,033365	9,52	0,000
Densidade Núcleo	1	0,20580	15,62%	0,11974	0,119735	34,15	0,000
FatorX - Resina	1	0,00087	0,07%	0,03423	0,034231	9,76	0,002
Dist_Esq_Frame11-13	1	0,01838	1,39%	0,06216	0,062159	17,73	0,000
Dist_Esq_Frame19-20	1	0,00173	0,13%	0,04412	0,044123	12,59	0,000
Dist_Esq_Frame24-25	1	0,19277	14,63%	0,05620	0,056202	16,03	0,000
Dist_Dir_Frame9-10	1	0,00428	0,32%	0,04395	0,043951	12,54	0,001
Pres_Esq_Frame19-20	1	0,00004	0,00%	0,04626	0,046262	13,20	0,000
Pres_Esq_Frame28-29	1	0,00009	0,01%	0,04412	0,044122	12,59	0,000
Pres_Esq_Frame30-32	1	0,02219	1,68%	0,01714	0,017139	4,89	0,028
Pres_Cen_Frame 2	1	0,01091	0,83%	0,03085	0,030847	8,80	0,003
Pres_Cen_Frame9-10	1	0,02205	1,67%	0,05430	0,054300	15,49	0,000
Pres_Cen_Frame30-32	1	0,00339	0,26%	0,02573	0,025726	7,34	0,007
Pres_Cen_Frame33	1	0,00429	0,33%	0,02780	0,027803	7,93	0,005
Pres_Dir_Frame17-18	1	0,03313	2,51%	0,02845	0,028452	8,12	0,005
Pres_Dir_Frame19-20	1	0,01441	1,09%	0,03936	0,039358	11,23	0,001
Pres_Dir_Frame26-27	1	0,00714	0,54%	0,01776	0,017758	5,07	0,026
Pres_Dir_Frame28-29	1	0,01917	1,46%	0,03509	0,035093	10,01	0,002
Pres_Dir_Frame30-32	1	0,00375	0,28%	0,02215	0,022154	6,32	0,013
Pres_Dir_Frame33	1	0,01814	1,38%	0,04357	0,043566	12,43	0,001
Densidade MatFormer	1	0,00718	0,54%	0,02091	0,020911	5,96	0,016
Humidade Secador	1	0,11096	8,42%	0,11096	0,110956	31,65	0,000
Erro	176	0,61701	46,83%	0,61701	0,003506		
Total	197	1,31768	100,00%				

Figura C.4: Análise de variância - Resistência à tração.

Anexo D

Resultados do modelo da densidade do núcleo

D.1 Resultados do modelo

Equação de Regressão

Densidade Núcleo = $11685 - 27,8 \text{ Altura do Colchão} + 1,0192 \text{ Densidade Placa}$
 $+ 0,0996 \text{ Velocidade_Linha} - 0,725 \text{ Temperatura1} + 0,741 \text{ Temperatura5}$

Figura D.1: Equação da regressão linear múltipla - Densidade do núcleo.

Sumário do Modelo

S	R2	R2(aj)	PRESQ	R2(pred)	AICc	BIC
12,4898	39,71%	38,11%	31315,0	35,96%	1546,61	1568,92

Figura D.2: Sumário do modelo - Densidade do núcleo.

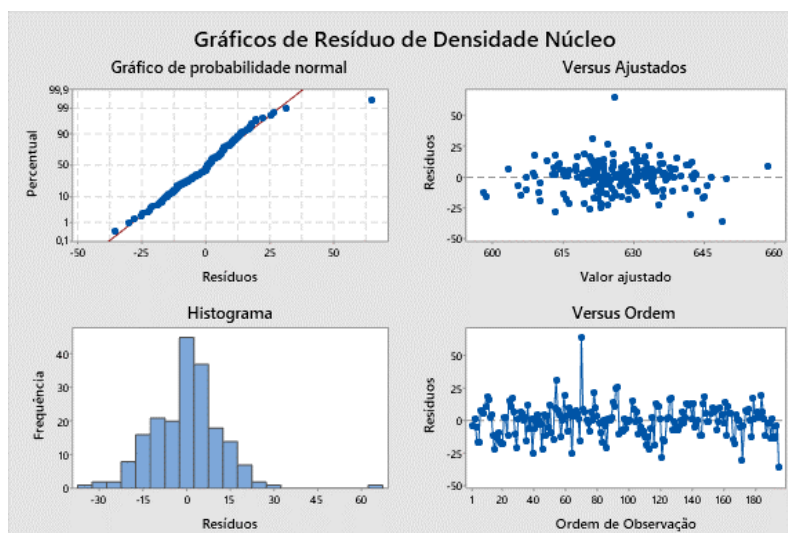


Figura D.3: Gráficos de resíduos - Densidade do núcleo.

Análise de Variância

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	5	19416,6	39,71%	19416,6	3883,3	24,89	0,000
Altura do Colchão	1	434,4	0,89%	1141,3	1141,3	7,32	0,007
Densidade Placa	1	17352,6	35,49%	16943,9	16943,9	108,62	0,000
Velocidade_Linha	1	437,5	0,89%	1314,2	1314,2	8,42	0,004
Temperatura1	1	131,2	0,27%	639,3	639,3	4,10	0,044
Temperatura5	1	1060,8	2,17%	1060,8	1060,8	6,80	0,010
Erro	189	29482,9	60,29%	29482,9	156,0		
Total	194	48899,5	100,00%				

Figura D.4: Análise de variância - Densidade do núcleo.

Anexo E

Resultados do modelo da densidade da placa

E.1 Resultados do modelo

Equação de Regressão

Densidade Placa = $486,2 - 2,034 \text{ Esp Saída} + 2,423 \text{ Total (kg)} - 3,57 \text{ Humidade Fibras}$
 $+ 0,499 \text{ Pressão Operação} + 0,561 \text{ Pressão Controle}$

Figura E.1: Equação da regressão linear múltipla - Densidade da placa.

Sumário do Modelo

S	R2	R2(aj)	PRESQ	R2(pred)	AICc	BIC
8,58633	13,48%	11,23%	15369,7	6,06%	1421,86	1444,29

Figura E.2: Sumário do modelo - Densidade da placa.

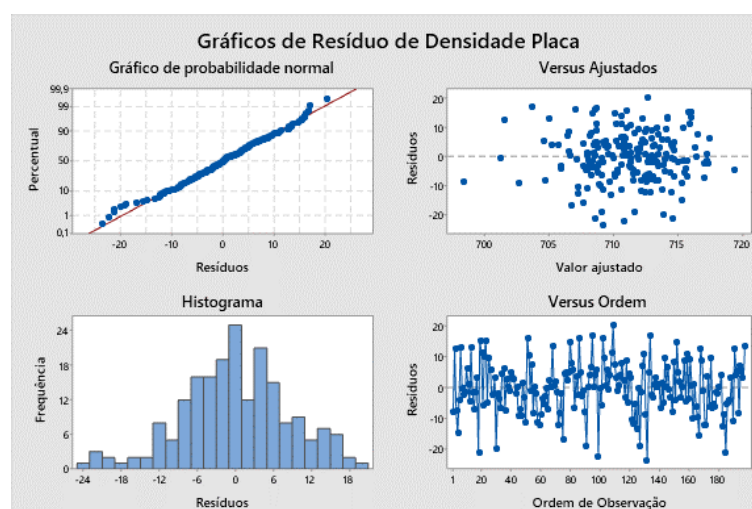


Figura E.3: Gráficos de resíduos - Densidade da placa.

Análise de Variância

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	5	2205,5	13,48%	2205,5	441,09	5,98	0,000
Esp Saída	1	11,2	0,07%	793,8	793,83	10,77	0,001
Total (kg)	1	34,7	0,21%	634,7	634,66	8,61	0,004
Humidade Fibra	1	1354,7	8,28%	541,9	541,90	7,35	0,007
Pressão Operação	1	481,5	2,94%	331,9	331,93	4,50	0,035
Pressão Controlo	1	323,3	1,98%	323,3	323,31	4,39	0,038
Erro	192	14155,2	86,52%	14155,2	73,73		
Total	197	16360,7	100,00%				

Figura E.4: Análise de variância - Densidade da placa.

Anexo F

Diagrama de relações

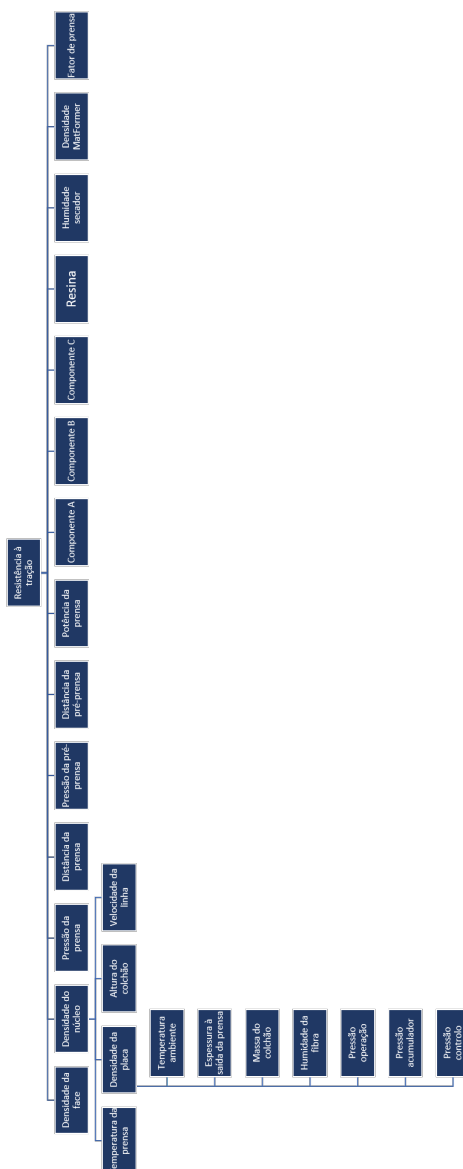


Figura F.1: Diagrama de relações com os três níveis.

Anexo G

Resultados e validação dos pressupostos do modelo de 1 Nível

G.1 Resultados do modelo

Equação de Regressão

Resistência à Tração = 1,041 - 0,05673 Total (kg) - 0,1696 FatorX - Componente A
+ 0,03020 Dist_Esq_Frame11-13 - 0,00979 Dist_Esq_Frame24-25
- 0,0854 Dist_Dir_Frame6 + 0,1602 Dist_Dir_Frame8
- 0,0596 Dist_Dir_Frame9-10 - 0,00304 Dist_Dir_Frame17-18
- 0,00289 Dist_Dir_Frame28-29 - 0,002618 Pres_Esq_Frame9-10
+ 0,00479 Pres_Esq_Frame28-29 + 0,002364 Pres_Cen_Frame7
- 0,002062 Pres_Cen_Frame8 + 0,00213 Pres_Cen_Frame33
+ 0,001385 Pres_Dir_Frame2 - 0,001013 Pres_Dir_Frame5
- 0,001307 Pres_Dir_Frame6 + 0,001455 Pres_Dir_Frame8
+ 0,002019 Pres_Dir_Frame9-10 + 0,001957 Pres_Dir_Frame26-27
- 0,00584 Pres_Dir_Frame28-29 - 0,00368 Pres_Dir_Frame33
+ 0,00767 Densidade MatFormer

Figura G.1: Equação da regressão linear múltipla - Resistência à tração (1 nível).

Sumário do Modelo

S	R2	R2(aj)	PRESQ	R2(pred)	AICc	BIC
0,0589510	55,79%	50,02%	0,793966	42,61%	-532,94	-457,95

Figura G.2: Sumário do modelo - Resistência à tração (1 nível).

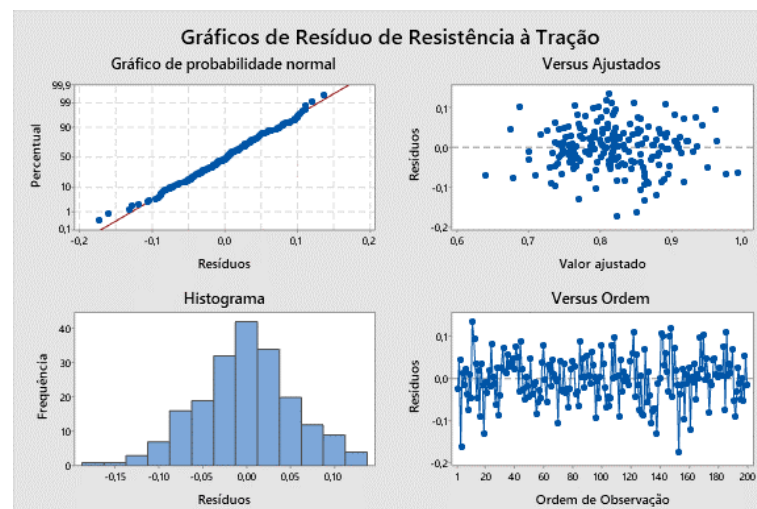


Figura G.3: Gráficos de resíduos - Resistência à tração (1 nível).

Análise de Variância

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	23	0,77191	55,79%	0,77191	0,033561	9,66	0,000
Total (kg)	1	0,09917	7,17%	0,15562	0,155616	44,78	0,000
FatorX - Componente A	1	0,03885	2,81%	0,05157	0,051568	14,84	0,000
Dist_Esq_Frame11-13	1	0,04490	3,25%	0,04898	0,048978	14,09	0,000
Dist_Esq_Frame24-25	1	0,06977	5,04%	0,04726	0,047262	13,60	0,000
Dist_Dir_Frame6	1	0,00045	0,03%	0,01839	0,018388	5,29	0,023
Dist_Dir_Frame8	1	0,04276	3,09%	0,04198	0,041978	12,08	0,001
Dist_Dir_Frame9-10	1	0,00897	0,65%	0,01471	0,014714	4,23	0,041
Dist_Dir_Frame17-18	1	0,01734	1,25%	0,02059	0,020593	5,93	0,016
Dist_Dir_Frame28-29	1	0,02099	1,52%	0,01691	0,016910	4,87	0,029
Pres_Esq_Frame9-10	1	0,00391	0,28%	0,03393	0,033929	9,76	0,002
Pres_Esq_Frame28-29	1	0,00141	0,10%	0,02698	0,026979	7,76	0,006
Pres_Cen_Frame7	1	0,00018	0,01%	0,05963	0,059630	17,16	0,000
Pres_Cen_Frame8	1	0,04205	3,04%	0,03613	0,036128	10,40	0,002
Pres_Cen_Frame33	1	0,01803	1,30%	0,01370	0,013700	3,94	0,049
Pres_Dir_Frame2	1	0,12290	8,88%	0,08086	0,080859	23,27	0,000
Pres_Dir_Frame5	1	0,06777	4,90%	0,05070	0,050702	14,59	0,000
Pres_Dir_Frame6	1	0,01144	0,83%	0,01829	0,018295	5,26	0,023
Pres_Dir_Frame8	1	0,02990	2,16%	0,03040	0,030400	8,75	0,004
Pres_Dir_Frame9-10	1	0,01062	0,77%	0,01839	0,018392	5,29	0,023
Pres_Dir_Frame26-27	1	0,00609	0,44%	0,01802	0,018016	5,18	0,024
Pres_Dir_Frame28-29	1	0,03397	2,45%	0,03125	0,031247	8,99	0,003
Pres_Dir_Frame33	1	0,02370	1,71%	0,02761	0,027611	7,95	0,005
Densidade MatFormer	1	0,05676	4,10%	0,05676	0,056759	16,33	0,000
Erro	176	0,61164	44,21%	0,61164	0,003475		
Total	199	1,38355	100,00%				

Figura G.4: Análise de variância - Resistência à tração (1 nível).

G.2 Validação de pressupostos

G.2.1 Normalidade multivariada

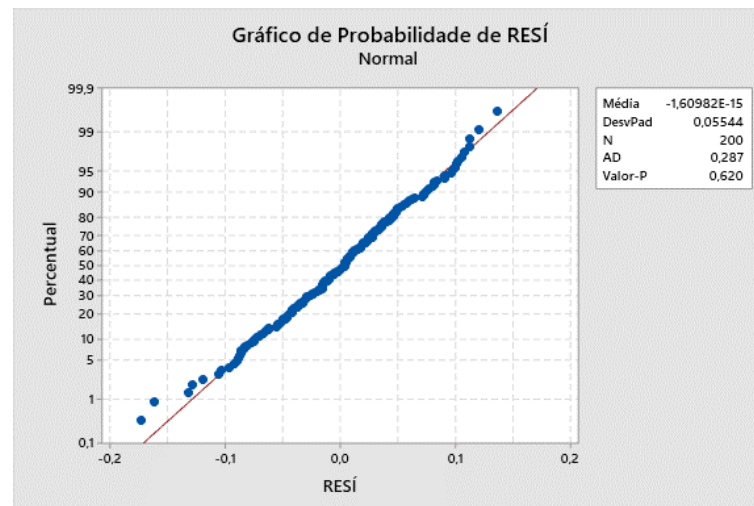


Figura G.5: Gráfico normal de resíduos.

Após analisar o gráfico da Figura G.5, pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que os pontos se aproximam da linha diagonal de referência.

Esta afirmação é suportada pelo teste de *Anderson-Darling*, em que se obteve um valor de 0,287 e um *p-value* de 0,620. Ou seja, é possível concluir que os resíduos estão distribuídos normalmente.

G.2.2 Influência multivariada dos outliers

Para averiguar se há *outliers* nos dados utilizados, recorre-se ao valor da distância de Cook, registrado na penúltima coluna da Figura G.6.

Os valores registrados são todos inferiores a 1. Pelo que se pode concluir que este pressuposto foi assegurado.

Ajustados e Diagnósticos para Observações Atípicas

Resistência			EP do				
Obs.	à Tração	Ajuste	Ajustado	IC de 95%	Resíd	Resíd Pad	Resíd Del
3	0,6900	0,8517	0,0174	(0,8174; 0,8861)	-0,1617	-2,87	-2,93
11	0,9500	0,8126	0,0177	(0,7776; 0,8475)	0,1374	2,44	2,48
19	0,6400	0,7694	0,0163	(0,7372; 0,8016)	-0,1294	-2,28	-2,31
122	0,9300	0,8172	0,0184	(0,7808; 0,8535)	0,1128	2,01	2,03
138	0,7300	0,8619	0,0274	(0,8078; 0,9161)	-0,1319	-2,53	-2,57
141	1,0000	0,8914	0,0234	(0,8452; 0,9375)	0,1086	2,01	2,02
148	0,9300	0,8094	0,0206	(0,7688; 0,8501)	0,1206	2,18	2,21
153	0,6500	0,8239	0,0171	(0,7903; 0,8576)	-0,1739	-3,08	-3,16
161	0,7600	0,8797	0,0257	(0,8290; 0,9303)	-0,1197	-2,25	-2,28
170	0,8700	0,7644	0,0276	(0,7098; 0,8189)	0,1056	2,03	2,05
186	0,9600	0,8473	0,0179	(0,8120; 0,8826)	0,1127	2,01	2,02
D de							
Obs.	HI	Cook	DAJUSTADOS				
3	0,087167	0,03	-0,90628	R			
11	0,090088	0,02	0,78026	R			
19	0,076700	0,02	-0,66635	R			
122	0,097495	0,02	0,66810	R			
138	0,216315	0,07	-1,34922	R			
141	0,157367	0,03	0,87502	R			
148	0,122096	0,03	0,82310	R			
153	0,083664	0,04	-0,95486	R			
161	0,189464	0,05	-1,10294	R			
170	0,219611	0,05	1,08571	R			
186	0,092080	0,02	0,64433	R			

R Resíduo grande

Figura G.6: Tabela de observações atípicas.

G.2.3 Independência estatística

Estatística de Durbin-Watson

Estatística de Durbin-Watson = 1,81745

Figura G.7: Estatística de *Durbin-Watson*.

A estatística de *Durbin-Watson* permite confirmar se os erros estão correlacionados entre si. Uma vez que o valor é inferior a 2, verifica-se a presença de correlação entre os erros.

No entanto, como é próximo de 2 (valor obtido quando não há correlação), pode considerar-se que a correlação não é significativa e que a independência estatística se encontra assegurada.

G.2.4 Multicolinearidade

Coefficientes

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P
Constante	1,041	0,109	(0,825; 1,257)	9,51	0,000
Total (kg)	-0,05673	0,00848	(-0,07347; -0,04000)	-6,69	0,000
FatorX - Componente A	-0,1696	0,0440	(-0,2565; -0,0827)	-3,85	0,000
Dist_Esq_Frame11-13	0,03020	0,00805	(0,01433; 0,04608)	3,75	0,000
Dist_Esq_Frame24-25	-0,00979	0,00266	(-0,01503; -0,00455)	-3,69	0,000
Dist_Dir_Frame6	-0,0854	0,0371	(-0,1587; -0,0121)	-2,30	0,023
Dist_Dir_Frame8	0,1602	0,0461	(0,0692; 0,2511)	3,48	0,001
Dist_Dir_Frame9-10	-0,0596	0,0290	(-0,1167; -0,0024)	-2,06	0,041
Dist_Dir_Frame17-18	-0,00304	0,00125	(-0,00551; -0,00058)	-2,43	0,016
Dist_Dir_Frame28-29	-0,00289	0,00131	(-0,00547; -0,00030)	-2,21	0,029
Pres_Esq_Frame9-10	-0,002618	0,000838	(-0,004272; -0,000965)	-3,12	0,002
Pres_Esq_Frame28-29	0,00479	0,00172	(0,00140; 0,00818)	2,79	0,006
Pres_Cen_Frame7	0,002364	0,000571	(0,001238; 0,003490)	4,14	0,000
Pres_Cen_Frame8	-0,002062	0,000639	(-0,003323; -0,000800)	-3,22	0,002
Pres_Cen_Frame33	0,00213	0,00107	(0,00001; 0,00425)	1,99	0,049
Pres_Dir_Frame2	0,001385	0,000287	(0,000818; 0,001951)	4,82	0,000
Pres_Dir_Frame5	-0,001013	0,000265	(-0,001536; -0,000490)	-3,82	0,000
Pres_Dir_Frame6	-0,001307	0,000569	(-0,002431; -0,000183)	-2,29	0,023
Pres_Dir_Frame8	0,001455	0,000492	(0,000484; 0,002426)	2,96	0,004
Pres_Dir_Frame9-10	0,002019	0,000878	(0,000287; 0,003751)	2,30	0,023
Pres_Dir_Frame26-27	0,001957	0,000860	(0,000261; 0,003654)	2,28	0,024
Pres_Dir_Frame28-29	-0,00584	0,00195	(-0,00968; -0,00200)	-3,00	0,003
Pres_Dir_Frame33	-0,00368	0,00131	(-0,00626; -0,00110)	-2,82	0,005
Densidade MatFormer	0,00767	0,00190	(0,00392; 0,01141)	4,04	0,000
Termo	VIF				
Constante					
Total (kg)	14,28				
FatorX - Componente A	2,27				
Dist_Esq_Frame11-13	38,18				
Dist_Esq_Frame24-25	6,30				
Dist_Dir_Frame6	865,00				
Dist_Dir_Frame8	1201,31				
Dist_Dir_Frame9-10	484,32				
Dist_Dir_Frame17-18	2,32				
Dist_Dir_Frame28-29	1,76				
Pres_Esq_Frame9-10	7,81				
Pres_Esq_Frame28-29	12,94				
Pres_Cen_Frame7	7,49				
Pres_Cen_Frame8	7,27				
Pres_Cen_Frame33	5,30				
Pres_Dir_Frame2	2,75				
Pres_Dir_Frame5	4,36				
Pres_Dir_Frame6	5,91				
Pres_Dir_Frame8	8,25				
Pres_Dir_Frame9-10	9,34				
Pres_Dir_Frame26-27	1,50				
Pres_Dir_Frame28-29	15,42				
Pres_Dir_Frame33	6,51				
Densidade MatFormer	1,19				

Figura G.8: Coeficientes do modelo obtido.

Se houver multicolinearidade no modelo, significa que duas ou mais variáveis independentes fornecem informação redundante acerca da variável resposta.

A multicolinearidade traduz-se através do valor do VIF (*Variance Inflation Factor*). A presença de valores superiores a 10 revela a presença de multicolinearidade no modelo.

O modelo obtido é composto por vários regressores com um VIF superior a 10, pelo que a multicolinearidade pode ser considerada um problema.

G.2.5 Homocedasticidade

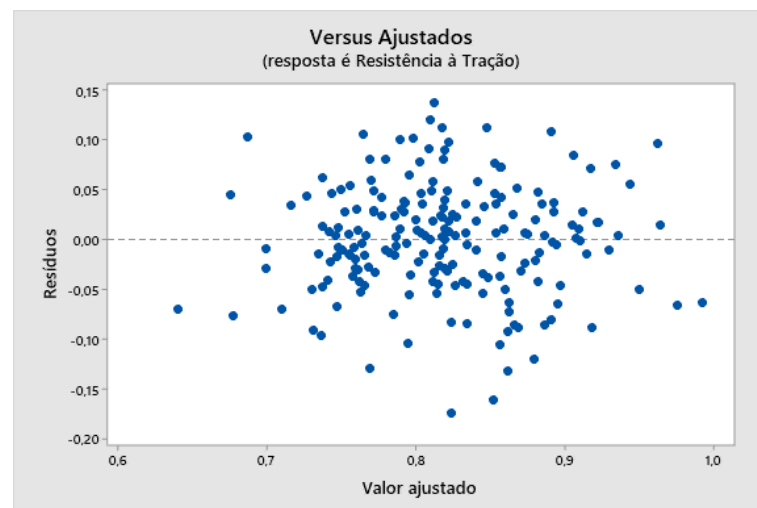


Figura G.9: Gráfico de resíduos versus ajustes.

O gráfico da Figura G.9, confirma que os resíduos têm variância constante. A dispersão dos dados evidencia aleatoriedade, não se verificando um padrão na sua distribuição. Logo, este pressuposto foi assegurado.

Anexo H

Resultados e validação dos pressupostos do modelo após PCA

H.1 Resultados do modelo

Equação de Regressão

Resistência à Tração = $-80,1 - 0,01651 \text{ Total (kg)} + 0,1926 \text{ Altura do Colchão}$
- $0,02008 \text{ FatorX - Resina} - 0,2365 \text{ FatorX - Componente A}$
- $0,000931 \text{ presPC4} - 0,000556 \text{ presPC5} - 0,001101 \text{ presPC7}$
+ $0,002001 \text{ presPC8} - 0,001137 \text{ presPC9} + 0,00856 \text{ Densidade MatFormer}$
- $0,02431 \text{ Humidade Secador}$

Figura H.1: Equação da regressão linear múltipla - Resistência à tração (após PCA).

Sumário do Modelo

S	R2	R2(aj)	PRESQ	R2(pred)	AICc	BIC
0,0658947	38,38%	34,55%	0,892633	28,44%	-476,00	-435,94

Figura H.2: Sumário do modelo - Resistência à tração (após PCA).

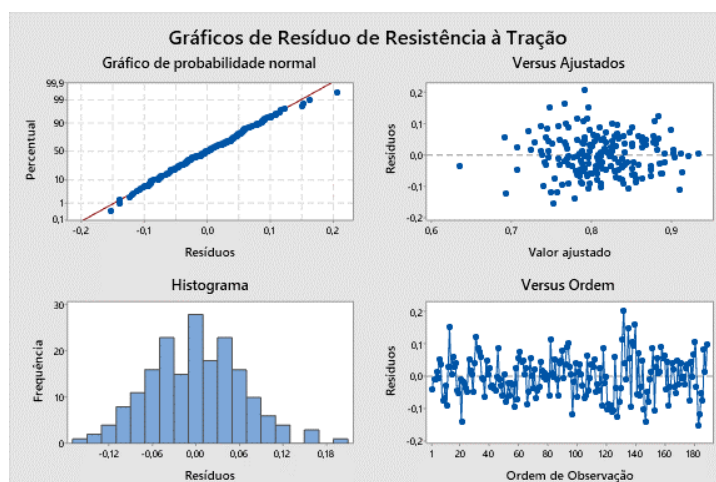


Figura H.3: Gráficos de resíduos - Resistência à tração (após PCA).

Análise de Variância

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	11	0,47878	38,38%	0,47878	0,043525	10,02	0,000
Total (kg)	1	0,11686	9,37%	0,13294	0,132945	30,62	0,000
Altura do Colchão	1	0,00841	0,67%	0,03094	0,030942	7,13	0,008
FatorX - Resina	1	0,01732	1,39%	0,03698	0,036980	8,52	0,004
FatorX - Componente A	1	0,03654	2,93%	0,10031	0,100306	23,10	0,000
presPC4	1	0,02912	2,33%	0,10085	0,100854	23,23	0,000
presPC5	1	0,00011	0,01%	0,04115	0,041150	9,48	0,002
presPC7	1	0,00051	0,04%	0,05697	0,056973	13,12	0,000
presPC8	1	0,15456	12,39%	0,18196	0,181959	41,91	0,000
presPC9	1	0,00618	0,50%	0,02308	0,023082	5,32	0,022
Densidade MatFormer	1	0,04541	3,64%	0,06778	0,067783	15,61	0,000
Humidade Secador	1	0,06376	5,11%	0,06376	0,063756	14,68	0,000
Erro	177	0,76855	61,62%	0,76855	0,004342		
Total	188	1,24733	100,00%				

Figura H.4: Análise de variância - Resistência à tração (após PCA).

H.2 Validação de pressupostos

H.2.1 Normalidade multivariada

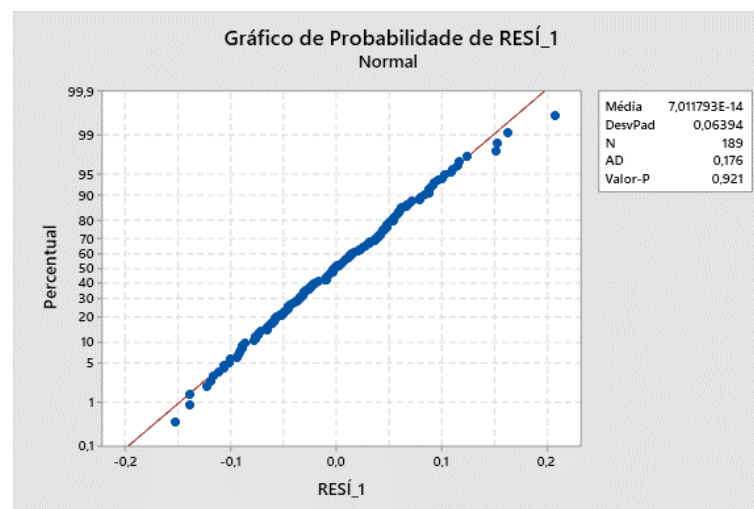


Figura H.5: Gráfico normal de resíduos.

O gráfico da Figura H.5 permite afirmar que os resíduos seguem uma distribuição normal, uma vez que os pontos se aproximam da linha diagonal de referência.

O teste de *Anderson-Darling* suporta esta afirmação através de um *p-value* de 0,921. Logo, é possível concluir que os resíduos estão distribuídos normalmente.

H.2.2 Influência multivariada dos outliers

Ajustados e Diagnósticos para Observações Atípicas

Obs.	Resistência à Tração	Ajuste	EP do Ajustado	IC de 95%	Resíd	Resíd Pad	Resíd Del
13	0,9500	0,7967	0,0096	(0,7777; 0,8156)	0,1533	2,35	2,38
21	0,6400	0,7800	0,0140	(0,7523; 0,8076)	-0,1400	-2,17	-2,20
92	0,7900	0,7519	0,0376	(0,6778; 0,8261)	0,0381	0,70	0,70
132	1,0000	0,7924	0,0193	(0,7542; 0,8305)	0,2076	3,30	3,39
135	0,9000	0,7475	0,0287	(0,6909; 0,8040)	0,1525	2,57	2,61
140	0,9300	0,7668	0,0209	(0,7254; 0,8081)	0,1632	2,61	2,66
147	0,6400	0,7795	0,0164	(0,7471; 0,8119)	-0,1395	-2,19	-2,21
183	0,6000	0,7532	0,0180	(0,7177; 0,7888)	-0,1532	-2,42	-2,45

Obs.	D de HI	Cook	DAJUSTADOS
13	0,021213	0,01	0,35083 R
21	0,045338	0,02	-0,47880 R
92	0,325289	0,02	0,48753 X
132	0,086127	0,09	1,04144 R
135	0,189070	0,13	1,26133 R
140	0,100946	0,06	0,89037 R
147	0,062093	0,03	-0,56846 R
183	0,074802	0,04	-0,69704 R

R Resíduo grande
X Atípicos X

Figura H.6: Tabela de observações atípicas.

Todos os regressores do modelo têm valores de distância de *Cook*, penúltima coluna da Figura H.6, inferiores a 1. Pelo que, se pode concluir que este pressuposto foi assegurado.

H.2.3 Independência estatística

Estatística de Durbin-Watson

Estatística de Durbin-Watson = 1,52600

Figura H.7: Estatística de *Durbin-Watson*.

O valor da estatística de *Durbin-Watson* está compreendido entre 1 e 2. Logo, pode-se concluir que apesar da sua existência, a correlação não é significativa e que a independência estatística se encontra assegurada.

H.2.4 Multicolinearidade

Coeficientes

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	-80,1	30,5	(-140,3; -19,8)	-2,62	0,009	
Total (kg)	-0,01651	0,00298	(-0,02240; -0,01062)	-5,53	0,000	1,27
Altura do Colchão	0,1926	0,0722	(0,0502; 0,3350)	2,67	0,008	1,84
FatorX - Resina	-0,02008	0,00688	(-0,03366; -0,00650)	-2,92	0,004	2,19
FatorX - Componente A	-0,2365	0,0492	(-0,3336; -0,1394)	-4,81	0,000	2,15
presPC4	-0,000931	0,000193	(-0,001312; -0,000550)	-4,82	0,000	2,92
presPC5	-0,000556	0,000181	(-0,000912; -0,000200)	-3,08	0,002	1,80
presPC7	-0,001101	0,000304	(-0,001701; -0,000501)	-3,62	0,000	2,21
presPC8	0,002001	0,000309	(0,001391; 0,002611)	6,47	0,000	2,26
presPC9	-0,001137	0,000493	(-0,002111; -0,000164)	-2,31	0,022	2,25
Densidade MatFormer	0,00856	0,00217	(0,00428; 0,01283)	3,95	0,000	1,18
Humidade Secador	-0,02431	0,00634	(-0,03683; -0,01179)	-3,83	0,000	2,11

Figura H.8: Coeficientes do modelo obtido.

A análise da Figura H.8 permite detetar que todas as variáveis independentes incluídas no modelo têm valores de VIF inferiores a 10. Logo, não há multicolinearidade significativa no modelo, pelo que este pressuposto foi garantido.

H.2.5 Homocedasticidade

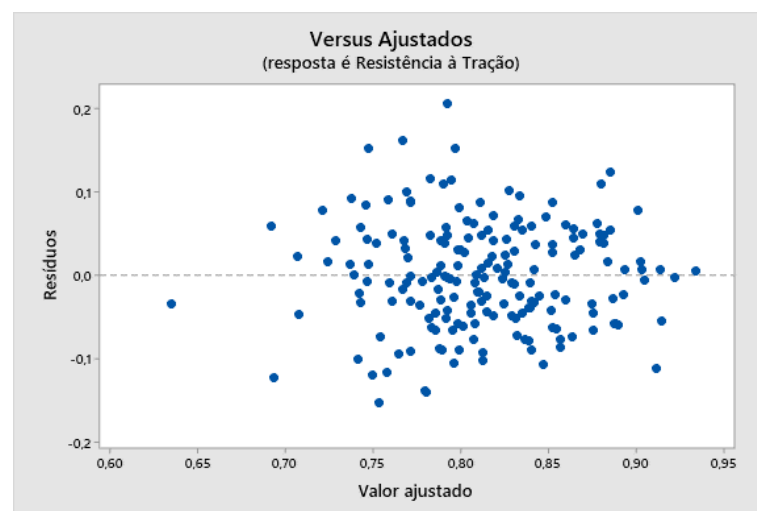


Figura H.9: Gráfico de resíduos versus ajustes.

A análise do gráfico da Figura H.9, revela a ausência de um padrão na distribuição dos dados. Uma vez que a dispersão dos dados evidencia aleatoriedade, pode-se considerar que a variância dos erros é similar.