UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA METALÚRGICA E DE MATERIAIS
CURSO DE ENGENHARIA METALÚRGICA

LUCAS RODRIGUES COELHO

ANÁLISE EXPLORATÓRIA E CORRELAÇÃO DAS VARIÁVEIS DE LINGOTAMENTO CONTÍNUO DE UMA SIDERÚRGICA

FORTALEZA
2019

LUCAS RODRIGUES COELHO

ANÁLISE EXPLORATÓRIA E CORRELAÇÃO DAS VARIÁVEIS DE LINGOTAMENTO CONTÍNUO DE UMA SIDERÚRGICA

Trabalho de conclusão de curso apresentado ao Curso de Engenharia Metalúrgica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Metalúrgica. Área de concentração: Siderurgia.

Orientador: Prof. Dr. Hamilton Ferreira Gomes de Abreu.

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C617a Coelho, Lucas Rodrigues.
Análise exploratória e correlação das variáveis de lingotamento contínuo de uma siderúrgica / Lucas Rodrigues Coelho. – 2019.
43 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Metalúrgica, Fortaleza, 2019.
Orientação: Prof. Dr. Hamilton Ferreira Gomes de Abreu.

1. Lingotamento contínuo. 2. Análise exploratória. 3. Correlação de dados. I. Título.

CDD 669

LUCAS RODRIGUES COELHO

ANÁLISE EXPLORATÓRIA E CORRELAÇÃO DAS VARIÁVEIS DE LINGOTAMENTO CONTÍNUO DE UMA SIDERÚRGICA

Trabalho de conclusão de curso apresentado ao Curso de Engenharia Metalúrgica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Metalúrgica. Área de concentração: Siderurgia.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

_____
Prof. Dr. Hamilton Ferreira Gomes de Abreu (Orientador)
Universidade Federal do Ceará (UFC)

_____
Prof. Dr. Jorge Luiz Cardoso
Universidade Federal do Ceará (UFC)

_____
Prof. Rômulo Alves Soares
Universidade Federal do Ceará (UFC)

À minha namorada, aos meus amigos e aos meus colegas de profissão.

AGRADECIMENTOS

À minha namorada, Juliana Teófilo, pela paciência e apoio em todos os momentos difíceis e pela parceria em todos os momentos felizes.

Ao Prof. Dr. Hamilton Ferreira Gomes de Abreu, pela orientação e amizade prestada ao longo do processo de feitura deste trabalho.

Aos professores participantes da banca examinadora Prof. Dr. Jorge Luiz Cardoso e Prof. Rômulo Alves Soares pelo tempo, pelas valiosas colaborações

A todos os amigos e colegas que conheci ao longo da graduação pois foram, cada um, uma peça para construção do que sou hoje.

Aos amigos do LACAM, João Vitor, Amilton Cardoso, Arthur Araújo, Igor Anjos, Thiago Cesar, Dyego Almeida, Beatriz Pinho, Soraia Castro, Letícia Muniz, Mirela Oliveira, Caio David, Matheus Vieira, Ítalo Maciel, Whescley de Abreu, Pablo Leão e Pedro Paulo, pelo companheirismo, conhecimento compartilhado e felicidades proporcionadas.

“O que pode um matemático não acadêmico fazer para mudar o mundo?” (SCHUTT, Rachel)

RESUMO

Seguindo a tendência da indústria de coletar e basear decisões em dados e fatos, o presente trabalho utilizou ferramentas da análise de dados, subdivisão da ciência de dados, para tratar um conjunto de observações retirado do banco de dados de uma siderúrgica e realizar análise exploratória. O objetivo desse trabalho foi o de realizar análise exploratória das informações coletadas no processo, detectar tendências e analisar a correlação entre variável e variável, variável e defeito, além de formular hipóteses com base nos resultados. A plotagem de histogramas, diagramas de caixas e gráficos de dispersão foi utilizada para analisar os dados através de metodologia estatística, e por fim foi feita a correlação através de gráficos de dispersão. Foram encontradas tendências em relação a velocidade de lingotamento e teor de hidrogênio, além de correlações fracas que indicam alguma dependência no relacionamento das variáveis de tempo de permanência de panela e teor de hidrogênio, e teor de hidrogênio com a presença de trincas longitudinais nas placas produzidas na siderúrgica. Por fim o trabalho sugere trabalhos futuros com base nas hipóteses formuladas.

Palavras-chave: Lingotamento contínuo. Análise exploratória. Correlação de dados.

ABSTRACT

Following the industry's trend of collecting and guiding the decisions based on data, this essay utilizes data analysis tools, a subdivision of data science, to treat and analyze a group of data, taken from a steelmaking plant's databank. The objective of this work was to execute an exploratory data analysis on the information collected on the process of continuous casting, detect trends, and check the correlation between variables, and between variables and quality defects. Histogram plots, boxplots, and scatterplots were used to analyze the data using a statistical methodology, and as a last step, a correlation between variables, and variables and quality defects was done. Trends in relation to continuous casting speed and hydrogen content in the steel were found, besides weak correlations that indicates some dependence on the relationship between time spent by the ladle on the continuous casting tower and hydrogen content, and hydrogen content with the presence of longitudinal cracks on the slabs produced in the steelmaking plant. Finally, hypotheses and future works proposals were presented.

Keywords: Continuous Casting. Exploratory Data Analysis. Data Correlation.

LISTA DE GRÁFICOS

SUMÁRIO

1 INTRODUÇÃO

Segundo estudo da McKinsey & Company (2018), a produção de dados na indústria tem crescimento constante nos últimos anos devido ao barateamento da armazenagem e aumento do poder de processamento dos computadores atuais. Esse movimento é suportado pela indústria 4.0 que conecta máquinas e sensores à internet e grandes bancos de dados (SILVA, 2017).

Nesse meio, a análise de dados cresce no mesmo passo, desenvolvendo maneiras de estudar os dados, simplifica-los e gerar embasamento para a tomada de decisão na indústria.

Na indústria siderúrgica essa realidade também está presente. Sensores de temperatura e teor de hidrogênio, registro de tempo e de velocidade de processo, tudo isso são informações valiosas e coletadas a cada momento da produção, 24 horas por dia.

Por causa da grande quantidade de dados em formato de tabela produzidos nos processos, é importante conhecer metodologias de análise de dados como a análise exploratória de dados. Com ela é possível detectar informações que na forma tabular, ou seja, números organizados em colunas e linhas, não seria possível ou seria extremamente demorado e trabalhoso, algo incompatível com o que se espera na indústria 4.0.

Parte importante da análise exploratória é a visualização dos dados para que seja possível entender como se comportam. Por exemplo, um histograma aliado à média de um conjunto de dados plotados em um gráfico geram uma visualização que transmite rapidamente se uma variável tem seus valores desviando muito ou pouco da média. O que foi previamente citado pode ser resumido em um só número: o desvio padrão. Entretanto, posto da primeira forma, alguém com pouca familiaridade com estatística saberia apontar o que foi dito com base no gráfico, mas dificilmente saberia apontar olhando apenas para o desvio padrão.

O presente trabalho realiza uma análise exploratória segundo metodologia estatística de algumas variáveis cujas observações pertencem ao processo de lingotamento contínuo de uma siderúrgica. Foram utilizadas ferramentas da análise de dados e da estatística para descrever sucintamente as variações do processo, detectar padrões em relação a defeitos de qualidade e formular hipóteses e tendências.

1.1 Objetivos

1. Realizar análise exploratória do conjunto de dados a fim de entender as variações do processo através da plotagem de gráficos.
2. Buscar correlações entre variáveis e defeitos de qualidade, e entre variáveis.
3. Formulação de hipóteses sobre os dados produzidos no processo de lingotamento contínuo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Estatística

Para entender melhor o comportamento dos dados que são analisados, conceitos estatísticos são utilizados para os resumir. Conceitos como média, moda, mediana, correlação entre variáveis, desvio padrão e distribuições de frequência são amplamente utilizados e combinados a fim de demonstrar tendências que de outra maneira seriam impossíveis de detectar apenas visualizando os dados. Esse trabalho trata de análise de variáveis em ambiente unidimensional ou bidimensional, isso é, análises das observações de uma ou duas varáveis de um conjunto de dados, respectivamente (HÄRDLE e SIMAR, 2003).

2.1.1 Média amostral

Representada na estatística por $\bar{x}$, é definida como a razão do somatório das observações, representadas por x_i, com i variando de 1 a n, pela quantidade de observações, n, no conjunto (TERRELL, 1999). Ou simplesmente por:

$$\bar{x} = \frac{\sum_1^n x_i}{n}$$

2.1.2 Mediana

Se um conjunto de dados for ordenado do menor para o maior, a mediana é o valor central dessa sequência ordenada de dados. Por representar o valor central, é natural que a forma de cálculo seja diferente para quantidades ímpares, onde só há um valor no centro da sequência, e para quantidades pares, onde existem dois (HÄRDLE e SIMAR, 2003). Para quantidades pares, a mediana é calculada por:

$$M = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

,

e quando for ímpar é calculada simplesmente por:

$$M = x_{\frac{n}{2}+1}$$

2.1.3 Quartis inferior e superior

Os quartis são semelhantes à mediana, mas para os valores que estão a 1 quarto e a 3 quartos do início da sequência ordenada do conjunto de dados. A partir da posição do valor da mediana, denominada aqui por z, calculasse: $posição\ do\ quartil\ = {(z+1)}/{2}$. O quartil inferior é o número na posição calculada anteriormente, tomando como referência o início da sequência, e o quartil superior, o final da sequência. No caso de a posição ser uma fração, tomasse os números inteiros imediatamente maior e menor, os valores nas respectivas posições da sequência, e calculasse a média desses números, definindo assim os quartis inferior e superior.

2.1.4 Variáveis contínuas e variáveis discretas

Variáveis contínuas são aquelas cujos valores não apresentam saltos entre si, ou seja, variam continuamente ao longo de todo um intervalo, podendo assumir infinitos valores entre dois pontos distintos (O'NEIL e SCHUTT, 2013). São variáveis como peso, altura, velocidade ou distância.

Já as variáveis discretas são aquelas que assumem valores específicos e limitados. Podem tomar apenas valores definidos e nenhum outro entre esses valores (O'NEIL e SCHUTT, 2013). Exemplos de variáveis discretas são: O número de pessoas em uma sala ou tipos aços.

2.1.5 Desvio padrão amostral

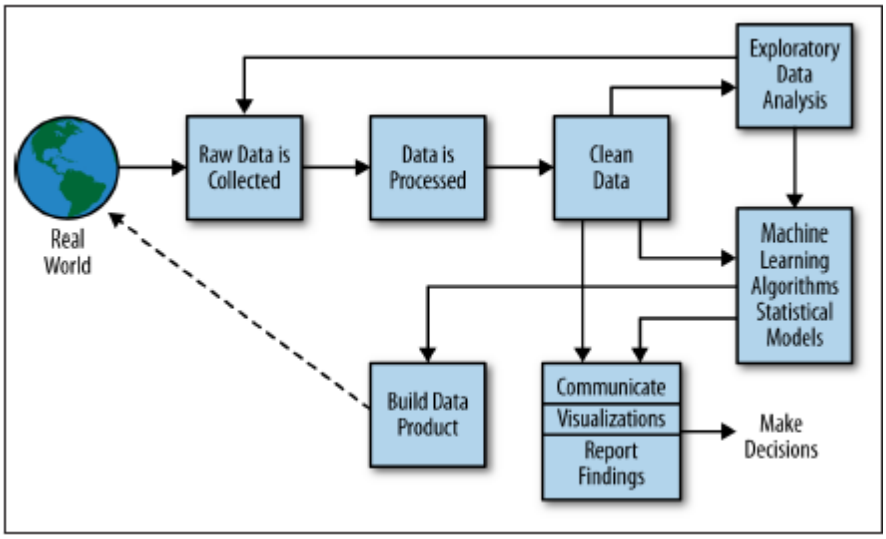
Verificar o quão dispersos os dados estão em relação à sua média é de grande valia para a análise. Com esse objetivo, calcula-se o desvio padrão dos dados. Observações próximas da média, indicam que estão distribuídas dentro de um intervalo menor de valores, ou seja, houve pouca dispersão. Terrel define o desvio padrão amostral (s) como a raiz quadrada da variância amostral (s^2) calculada por (TERRELL, 1999):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

2.2 Análise de dados

A análise de dados é uma subdivisão da ciência de dados que se concentra na análise subsequente ao tratamento e limpeza de dados brutos (figura 1), com o objetivo de descrever e explorar como o grupo de dados analisado se comporta, e que relações podem ser observadas naquela população. Focos neste trabalho, dois tipos de análise são: A análise exploratória de dados (EDA – do inglês *Exploratory Data Analysis*) (O'NEIL e SCHUTT, 2013). Esta se enquadra, como o nome sugere, na exploração do conjunto de dados, geralmente se utilizando da plotagem de gráficos, chamadas de visualizações, outra subdivisão da ciência de dados, para facilitar a compreensão de como as variáveis do conjunto de dados se correlacionam e para desenvolver ideias de maneira intuitiva (WANG, ZHANG, *et al.*, 2018).

Figura 1. Processo de análise de dados segundo a ciência de dados



Fonte: Rachel Schutt and Cathy O'Neil (2014, p. 41)

2.2.1 Limpeza dos dados

Problemas durante a coleta de grandes quantidades de dados, que são coletados durante um grande intervalo de tempo, podem ocorrer (O'NEIL e SCHUTT, 2013). Equipamentos podem falhar, problemas eletrônicos podem gerar dados surreais ou mesmo não gerar dado algum durante um número de observações. Esses problemas se traduzem no conjunto de dados na forma de observações não disponíveis, valores extremos como zeros ou 9999 e erros de digitação em variáveis de texto.

Para as variáveis discretas, funções que retornam os valores únicos ou duplicados contidos no conjunto de dados são excelentes para encontrar valores que deveriam ser idênticos, mas que foram preenchidos de maneira levemente diferente, ou valores que estão duplicados, no caso de variáveis onde todas as observações deveriam ser únicas. Para as variáveis contínuas, uma função que retorne como os dados estão distribuídos estatisticamente é uma ótima maneira de detectar valores zero onde não deveria existir zero, ou valores no limite da escala, que podem indicar um valor não condizente com o coletado, sendo então mais prudente o descarte desses. Ambas as variáveis podem ser também testadas com funções que indiquem observações não disponíveis, que revelam falha na coleta (WICKHAM, GROLEMUND e GARRETT, 2017).

2.3 Análises gráficas

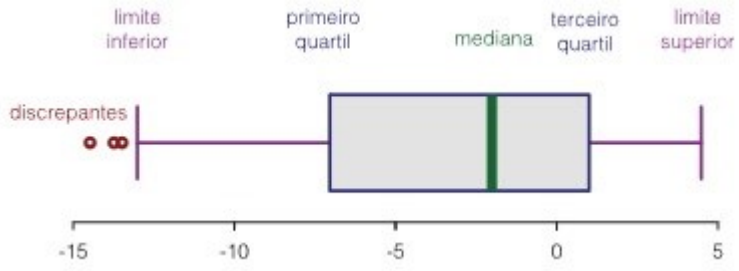
São as análises em forma de plotagem de gráfico. Essas auxiliam a visualização de informações que na forma bruta, seriam dificilmente analisados. Com a análise gráfica, os dados são tratados em forma de gráficos, resumindo o comportamento dos dados a uma forma mais amigável ao ser humano, facilitando o entendimento e a detecção de tendências ou correlações (WICKHAM, GROLEMUND e GARRETT, 2017).

Neste trabalho foram utilizados diferentes tipos de análises gráficas.

2.3.1 Diagrama de caixa

Diagrama de caixa é uma técnica de análise gráfica muito útil para identificar visualmente a mediana, os primeiro e terceiro quartis, os limites inferior e superior e, quando presentes, as discrepâncias (HÄRDLE e SIMAR, 2003), como mostra a figura 2. A visualização de um conjunto de dados unidimensional, ou como apresentado nesse trabalho, a análise de todas as observações, isolando-se uma das variáveis do conjunto de dados multidimensional, utilizando o diagrama de caixa é uma maneira prática de obter informações da distribuição desses dados e, o mais importante, da presença de dados discrepantes.

Figura 2. Representação de um diagrama de caixa



Fonte: Google imagens.

A construção do diagrama de caixa é realizada através de um sumário com cinco valores: Mediana, primeiro e terceiro quartis, e os extremos. Os limites inferior e superior são calculados a partir dos quartis inferior e superior. Sendo F_L o quartil inferior e F_U o quartil superior, o limite inferir é calculado por $F_L - 1{,}5(F_U - F_L)$ e o limite superior é calculado por $F_U + 1{,}5(F_U - F_L)$.

2.3.2 Histogramas

Um histograma é um gráfico de barras, sem espaço entre cada uma, que representam um intervalo, fechado na esquerda, de tamanho h (HÄRDLE e SIMAR, 2003). Cada intervalo, ou cesta, B_j pode ser representado por:

$$B_j(x_0, h) = [x_0 + (j-1)h, x_0 + jh)$$

A constante h define a largura da cesta, ou seja, o tamanho do intervalo de valores que está contido dentro de cada intervalo B_j, e o gráfico é construído como uma sequência de n cestas B_j, contendo do valor mínimo ao máximo dos dados.

A utilização de histogramas para estimar a densidade dos dados serve como uma maneira de se ter uma noção da distribuição dos dados. Um h muito pequeno gera um gráfico que não expressa de maneira satisfatória a distribuição dos dados, com muitos picos irrelevantes. Por outro lado, um h grande demais gerará um gráfico com grandes blocos, escondendo informação e deixando o gráfico pouco estruturado (HÄRDLE e SIMAR, 2003).

2.3.3 Gráfico de dispersão

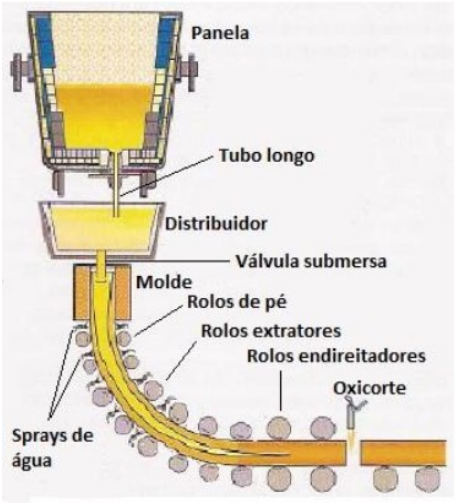
Gráficos de dispersão são gráficos cujas coordenadas são governadas por variáveis contínuas. É um dos gráficos mais úteis na análise estatística (TERRELL, 1999).

São utilizados para entender a natureza da relação entre duas ou três variáveis distintas entre si para detecção de possíveis agrupamentos de dados e tendências em relação a essas variáveis (HÄRDLE e SIMAR, 2003). No presente trabalho, foi utilizado na forma bidimensional para comparar duas variáveis, nos eixos x e y.

2.4 Lingotamento contínuo

O processo de lingotamento contínuo é o processo onde metal líquido é solidificado de maneira contínua, ou seja, virtualmente sem interrupções do processo. Isso ocorre graças à maneira como é feita essa solidificação (THOMAS, 2001). Metal líquido carregado em panelas de aço com interior recoberto por refratários são carregadas em uma torre de lingotamento, sobre um distribuidor e um molde. O metal pode ser despejado através de tubo (chamadas de tubo longo), equipamentos refratários que isolam o metal do ar, ou de maneira livre dentro do distribuidor. Esse por sua vez distribui o metal para os veios de lingotamento que também podem ou não utilizar válvulas (aqui chamadas de válvula submersa) para um ou mais moldes, que tem o trabalho de criar um gradiente de temperatura entre a superfície do líquido que toca o molde e o centro líquido do metal. O metal semisolidificado é então movimento através de rolos de lingotamento, que suportam a fina casca de metal sólido, dão forma e resfriam a peça. Ao final, o metal é cortado através de oxicorte para separar o produto final na forma de placas tarugos ou perfis (COELHO, 2013). Um esquemático do processo pode ser visto na figura 3.

Figura 3. Esquemático do processo de lingotamento contínuo.

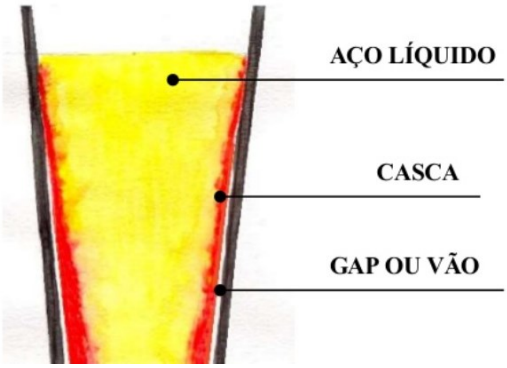


Fonte: Estudo comparativo entre fluxantes aplicados no lingotamento contínuo do aço SAE 1046 MOD (2014, p. 8)

A fabricação é balanceada para que seja mantida uma casca de metal na interface Molde-Metal líquido de forma que o metal líquido no centro seja mantido dentro do encapsulamento de metal sólido ao longo do processo fabril, como mostra a figura 4, sendo completamente solidificado ao final. Essa é a maneira mais eficiente de produzir metal sólido em grandes quantidades (THOMAS, 2001). Os metais fabricados com esse procedimento têm formas retangulares, quadradas ou em forma de "osso de cachorro", que são produtos semiacabados que passaram por outros processos de fabricação a fim de se formar o produto final. O lingotamento contínuo pode ser usado na fabricação de aço, alumínio, cobre, níquel e outros metais (THOMAS, 2001).

Atualmente, 85% da produção de aço mundial é feita através do lingotamento contínuo (COELHO, 2013). O método se mostra eficiente devido à sua capacidade de produção elevada a custo produtivo baixo, embora o custo inicial do processo seja mais alto do que o custo dos lingotamentos convencionais (THOMAS, 2001).

Figura 4. Solidificação no molde



Fonte: Curso básico de lingotamento contínuo (2016, p. 37).

2.4.1 Placas

As placas são um dos produtos semiacabados produzidos pelas siderúrgicas (figura 5). É o produto final do processo de lingotamento contínuo de placas. Suas dimensões são governadas pelas dimensões do molde e da máquina de lingotamento contínuo, podendo alcançar os mais variados comprimentos, espessuras e larguras (IRVING, 1993).

Figura 5. Placa de aço após o oxicorte

Fonte: Próprio autor.

2.4.2 Defeitos de qualidade

Os defeitos de qualidade são todos os casos onde há falha no processo resultando em um produto cujas especificações estão fora do que foi definido pelo cliente, seja este interno ou externo (CAMPOS, 1992).

Na indústria siderúrgica estudada, os defeitos de qualidade no lingotamento são os defeitos apresentados na placa de aço ou durante o processo, que gerassem um desacordo com as exigências do cliente externo.

2.4.2.1 Rebarba de corte

Ao cortar o veio de aço solidificado por oxicorte para formar placas de tamanho menor, é formada uma rebarba, resto de metal liquefeito e resolidificado durante o corte, que se mostra como uma protuberância alongada e fina em toda a largura oposta à área de contato com o maçarico de corte. Na siderúrgica cujo processo foi estudado, Rebarba de corte se dá ao nome do defeito que consiste na não retirada da rebarba pelo rebarbador e cujo defeito só foi detectado após o resfriamento completo da placa.

2.4.2.2 Rebarba quente

Rebarba quente, na siderúrgica estudada, se dá à rebarba cuja detecção foi feita ainda com a placa quente, logo após o processo de rebarbação pelo rebarbador.

2.4.2.3 Trinca longitudinal

Na siderúrgica estudada, trinca longitudinal é o nome dado para o defeito que se mostra como trinca, no sentido longitudinal da placa, ao longo de qualquer uma das faces da placa e após o seu resfriamento.

2.4.3 Rebarbador

É o equipamento utilizado para retirar a rebarba formada durante o processo de oxicorte. Consiste de várias séries de martelos metálicos conectados a um eixo giratório, que está ligado a um motor elétrico, como mostra a figura 6. Ao passar a placa sobre o rebarbador, esse gira até uma velocidade determinada e então se projeta pra cima, atingindo os martelos na extremidade da placa onde se forma a rebarba. Esses martelos realizam uma raspagem de cada borda inferior das placas produzidas para que seja feito o nivelamento dessas bordas.

Figura 6. Eixo de um rebarbador com martelos

Fonte: Numtec Alpine Metal Tech.

3 DESENVOLVIMENTO

3.1 Composição dos dados

O conjunto de dados, coletado automaticamente através do sistema de controle de produção da siderúrgica, contém 66190 observações e 11 variáveis, entre elas, variáveis contínuas e variáveis discretas. Seis das variáveis do conjunto são discretas. São elas: Código da placa, Código da corrida, Número do Veio, Espessura da Placa, Classe do aço e Defeito na placa. As outras cinco são variáveis contínuas, Velocidade Média de Lingotamento, Delta de velocidade, Teor de hidrogênio, Temperatura do aço e Tempo de permanência da panela na torre. Alguns desses dados não apresentam potencial de correlação, como por exemplo a identificação da placa e o número do veio, já que ambas são determinadas pelo processo e uma correlação entre essas não faz sentido. Já a correlação entre veio e defeito, por exemplo, pode revelar uma tendência interessante e são o centro deste trabalho.

3.2 Limpeza e Manipulação dos dados

Dados geralmente contém erros e precisam ser limpos. Uma análise de cada variável foi realizada para identificar possíveis erros que inviabilizem a utilização, como dados faltando. Nisso o software de análise de dados RStudio, que utiliza a linguagem de programação R, auxilia bastante no processo por possuir funções que resumem os dados dando uma visão rápida da distribuição desses dados, com máximo, mínimo, primeiro e terceiro quartos, média e moda. Os pacotes usados durante o desenvolvimento desse trabalho foram: *ggplot2*, *cowplot*, *stringr* e *colorspace*. Entrar em detalhes sobre quais funções foram utilizadas para analisar os dados foge do escopo deste trabalho, portanto elas serão omitidas, sendo apresentados apenas os resultados das operações realizadas e qual medida foi tomada para contornar as falhas dos dados.

Analisando a estrutura do conjunto de dados, foi possível ver o tipo de cada variável, designado automaticamente pelo software no momento do carregamento dos dados. Essa designação por vezes é errada, por exemplo: Dados numéricos podem ser carregados como caracteres e vice-versa. Dessa forma, é necessária a checagem desses tipos de dados e, eventualmente, corrigir os dados para que a análise possa ocorrer sem erros. A tabela 1 mostra como o software designou automaticamente e como foi feita a correção.

Tabela 1. Relação das variáveis com seus tipos de dados

Variável	**Tipo automaticamente designado**	**Tipo corrigido**
Código da placa	Caracteres	Caracteres
Código da corrida	Caracteres	Caracteres
Número do Veio	Caracteres	Categórico
Espessura da Placa	Numérico	Categórico
Velocidade Média de Lingotamento	Numérico	Numérico
Classe do aço	Caracteres	Categórico
Delta de velocidade (real – programada)	Numérico	Numérico
Teor de hidrogênio	Numérico	Numérico
Temperatura do aço	Numérico	Numérico
Defeito na placa	Caracteres	Categórico
Tempo de permanência da panela na torre	Data	Numérico

Fonte: Elaborado pelo autor.

A mudança das variáveis que tiveram seu tipo corrigido para categórico é justificada pelo fato de tais variáveis serem definidas pelo processo de lingotamento da siderúrgica, como por exemplo, a espessura da placa. Apesar de ser representada por um número, essa espessura não é uma variável contínua, e sim uma variável discreta, com valores fixos predefinidos. Isso quer dizer que os valores podem ser postos em categorias, ou cestas, sem prejuízo à análise.

Após estruturar corretamente os dados, foi realizada a análise de cada variável à procura de possíveis erros nos dados devido à frequência com que problemas de coleta podem surgir na forma de dados inesperados ou faltando. Para encontrar tais erros, os dados foram divididos em variáveis discretas e contínuas e aplicadas funções diferentes na linguagem R.

Na variável Número do Veio, foram observadas 3 categorias. Entretanto, fisicamente só existem 2 veios na siderúrgica cujo processo foi estudado. Uma análise mais aprofundada resultou na detecção do erro: 1 (uma) observação estava registrada como *Veio2* no lugar de *Veio 2*. Foi então substituído esse valor pelo valor esperado.

Dentre as observações da variável Velocidade Média de Lingotamento, 67 delas registraram velocidade de 0 m/min. Entretanto, essa velocidade é absurda no lingotamento contínuo, ou caracteriza um problema durante o processo que faz com que esses valores devam ser desconsiderados durante a sua análise por não representarem a realidade do processo. Por conseguinte, as mesmas 67 observações da variável Delta de Velocidade, também devem ser desconsideradas pelo fato dessa variável ter seu valor dependente da anterior.

Outra variável que precisou de limpeza foi a variável Teor de Hidrogênio. Os valores observados estavam em um intervalo de 0 ppm a 9,99 ppm. O extremo menor desse intervalo indica erro de coleta pois o processo na siderúrgica estudada não prevê aços com níveis tão baixos de hidrogênio. Já o extremo superior exibe o limite de escala do sensor de hidrogênio, sendo impossível precisar o valor real de hidrogênio. Por isso, ambos foram desconsiderados durante suas análises.

Ao analisar o intervalo de valores contidos nas observações da variável Temperatura do aço, foi encontrado valores mínimos de 1470ºC. Quando comparadas às temperaturas Liquidus, os valores mínimos eram de 1515ºC, indicando que a menor temperatura dos aços produzidos deveria ser acima de 1515ºC. Partindo dessa premissa, foi feita a análise de todas as variáveis relacionadas a essas observações e foi constatado o seguinte: Todas pertencem a uma mesma corrida, os valores de hidrogênio dessa corrida são zero ppm e a velocidade de lingotamento é inferior à velocidade programada – normalmente relacionada a aços com superaquecimento acima do recomendado, não o contrário, como é o caso de um aço com temperatura de 1470ºC. Foi então decidido que os dados não faziam sentido e, por isso, as observações dessa corrida foram retiradas do conjunto de dados para evitar erros nas análises futuras.

As variáveis que não foram citadas no parágrafo anterior estavam com integridade satisfatória das observações e não precisaram passar por manipulação ou limpeza dos dados nesse momento inicial.

Ao final da limpeza, o conjunto de dados continha 66174 observações. A diferença de 16 observações se deu pelo motivo de uma corrida, citada acima como contendo erro em diversas variáveis, com 16 placas, ter sido retirada dos dados. Para uma análise multivariada, as observações que apresentaram erro deveriam ser totalmente excluídas. Entretanto, o presente trabalho utiliza análises até o máximo de duas dimensões (bivariada), dessa forma as outras variáveis que apresentaram problemas em relação aos dados não tiveram essas observações retiradas por completo pois essa ação também removeria as observações de outras variáveis que não apresentaram qualquer erro. Ao longo das análises, quando necessário, os dados que contém erro foram temporariamente removidos.

3.3 Metodologia de análise

Härdle et al. citam uma metodologia de comparação de dados utilizando 3 ferramentas principais: Diagramas de caixa, histogramas e gráficos de dispersão. O objetivo

da utilização dessas análises é o de encontrar dados que estejam mais dispersos que outros, descobrir a existência de *outliers* e a existência de subgrupos dentro dos dados (HÄRDLE e SIMAR, 2003).

Nesse trabalho, foi utilizada a metodologia citada anteriormente a fim de entender melhor como se comporta o processo de lingotamento apenas olhando para as variáveis estudadas. Diagramas de caixa foram utilizados para encontrar a existência de outliers nas observações de variáveis contínuas, em seguida foram plotados histogramas dessas variáveis (no caso de variáveis discretas, foi realizada a contagem de cada categoria de valor), com cestas definidas utilizando o *Silverman's rule of thumb* (SILVERMAN, 1986), gerando uma visualização de como se distribuíram as observações dentro do intervalo de cada variável já excluídas de seus outliers para ser possível detectar padrões ou picos relevantes que indiquem uma tendência. Em seguida foram plotados histogramas das mesmas variáveis, isolando-se as observações que apresentaram defeito para ser possível notar regiões onde esses defeitos possivelmente se acumularam e, por fim, foi utilizado um conjunto de dados modificado, composto pela soma do número de placas com defeito em cada corrida, para que fosse possível gerar um gráfico de dispersão que teve como intuito observar se houve alguma tendência relacionando a presença de defeitos às variáveis estudadas.

Algumas variáveis não passaram por todas as análises pelo motivo de, em determinados casos, não haver sentido prático do ponto de vista do processo de lingotamento uma correlação entre essas variáveis.

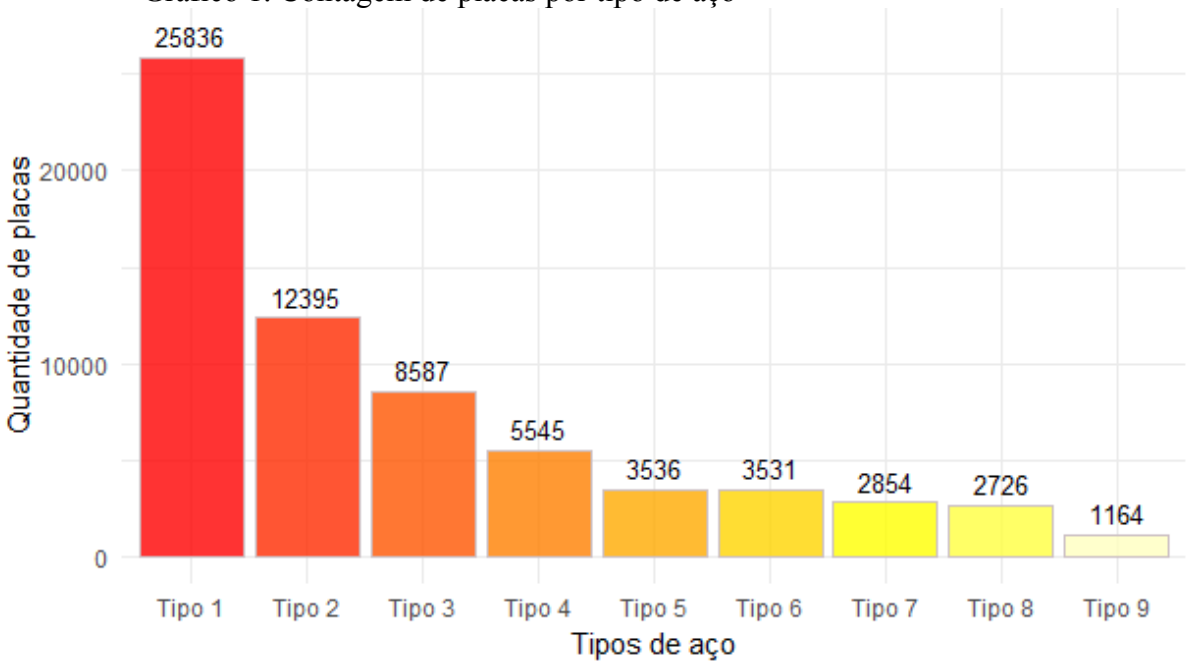
3.4 Resultados e discussão

3.4.1 Análises exploratórias

Foi iniciada a análise exploratória a partir do tipo de aço. Foi analisado como se comportou a produção da siderúrgica estudada em relação aos tipos de aço produzidos.

O gráfico 1 mostra que o aço de tipo 1 foi largamente mais produzido que os outros, sendo a produção do segundo tipo mais produzido apenas 47,9% da produção do primeiro. É importante ter esse conhecimento pois defeitos que se apresentarem em grande número no tipo 1, não necessariamente expressam uma tendência do tipo 1 a apresentar tal defeito, apenas indica que a quantidade de defeitos foi alta pois a produção desse tipo também foi.

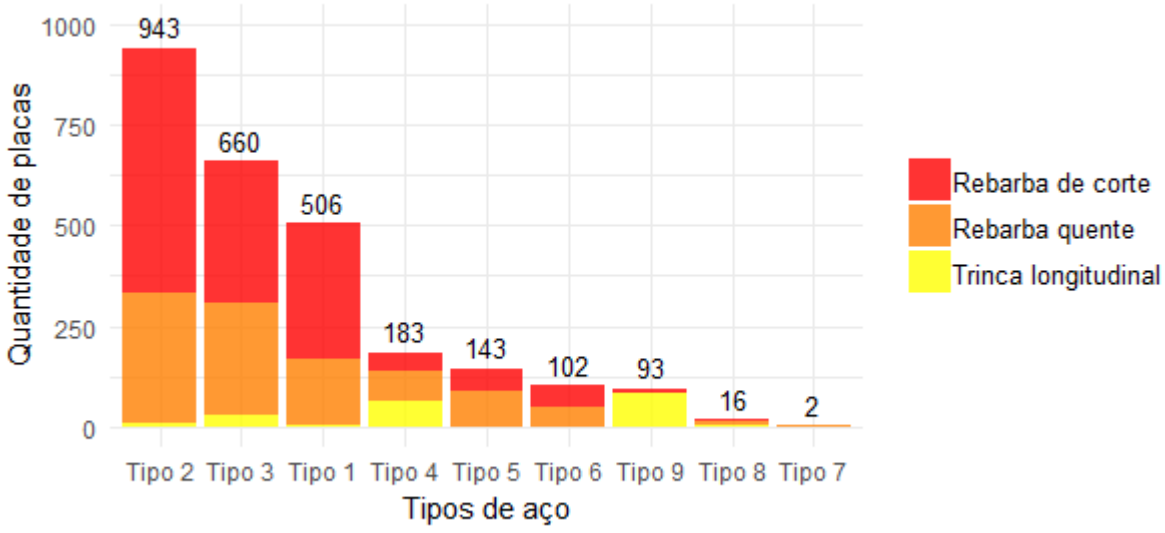
Gráfico 1. Contagem de placas por tipo de aço



Fonte: Elaborado pelo autor.

O gráfico 2 mostra que o defeito de rebarba de corte ocorreu em maior proporção nos aços do tipo 1, 2, 3 e 6. O defeito de rebarba quente aparenta igual distribuição em todos os aços. O defeito de trinca longitudinal se apresentou em sua maioria nos aços tipo 9 e tipo 4.

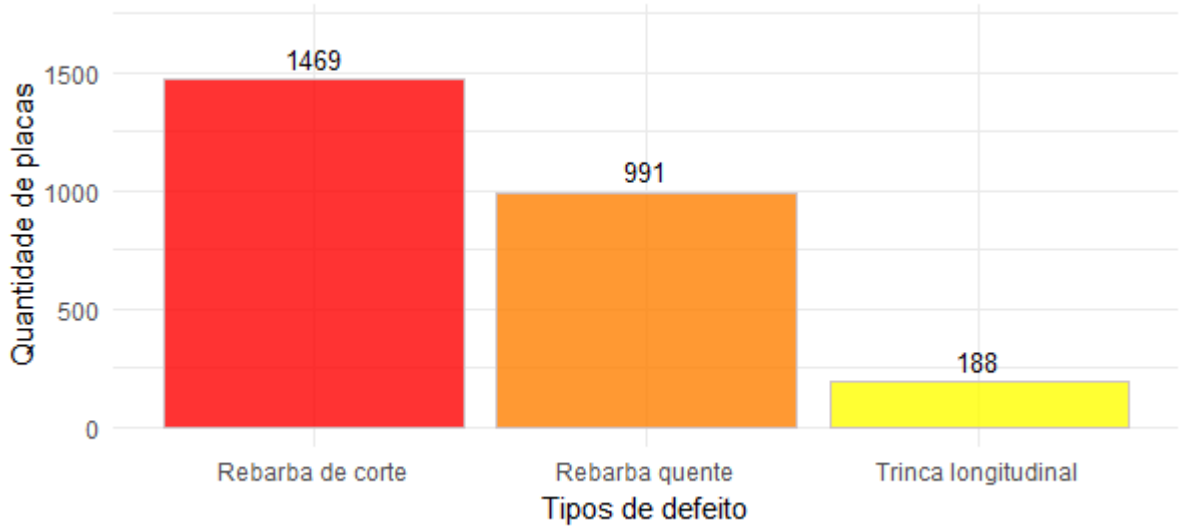
Gráfico 2. Contagem dos defeitos em relação ao tipo de aço



Fonte: Elaborado pelo autor.

Em seguida foi feita a contagem dos defeitos por tipo. É possível observar no gráfico 3 que há uma maior presença de rebarba de corte, sendo a segunda maior as rebarbas quentes e por último, com quantidade bem inferior às anteriores, as placas com trinca longitudinal. Em um universo de 66174 placas, os defeitos de rebarba de corte foram os mais frequentes, totalizando 2.2% das placas produzidas no período, seguidos dos defeitos de rebarba quente, com 1.5% do total e por último os defeitos de trinca longitudinal, com 0.28%.

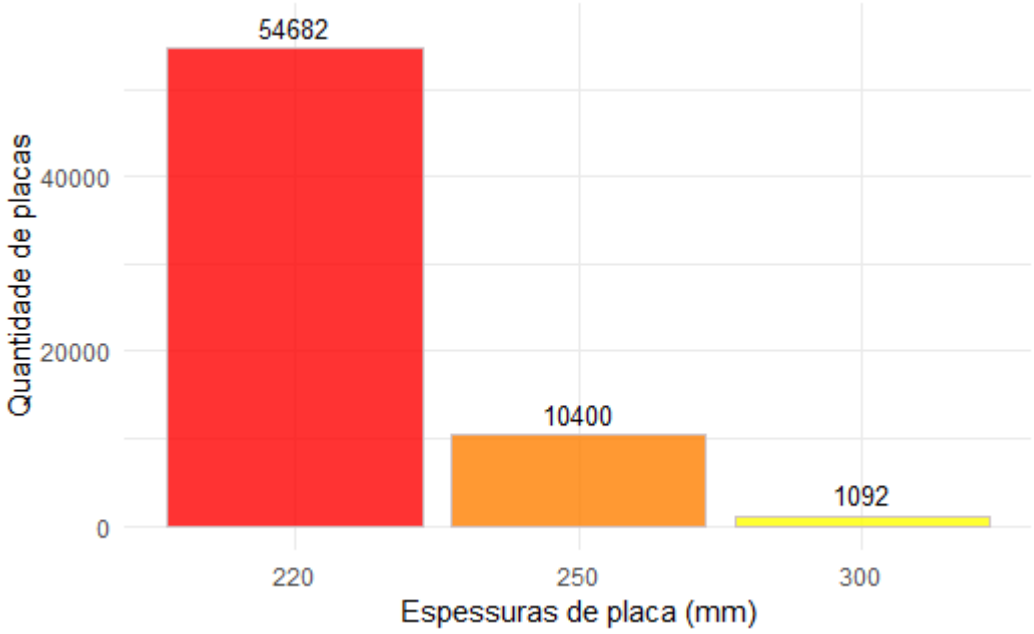
Gráfico 3. Contagem dos defeitos ocorridos no período



Fonte: Elaborado pelo autor.

Em seguida foi realizada a contagem de placas por espessura. No gráfico 4 é possível perceber que a produção de placas de 220 mm foi bem maior do que as de placas com 250 mm e 300 mm, e que a mesma premissa em relação à quantidades de placas de tipo 1 deve ser observada quando analisando dados que relacionem uma variável com a quantidade de placas de determinada espessura.

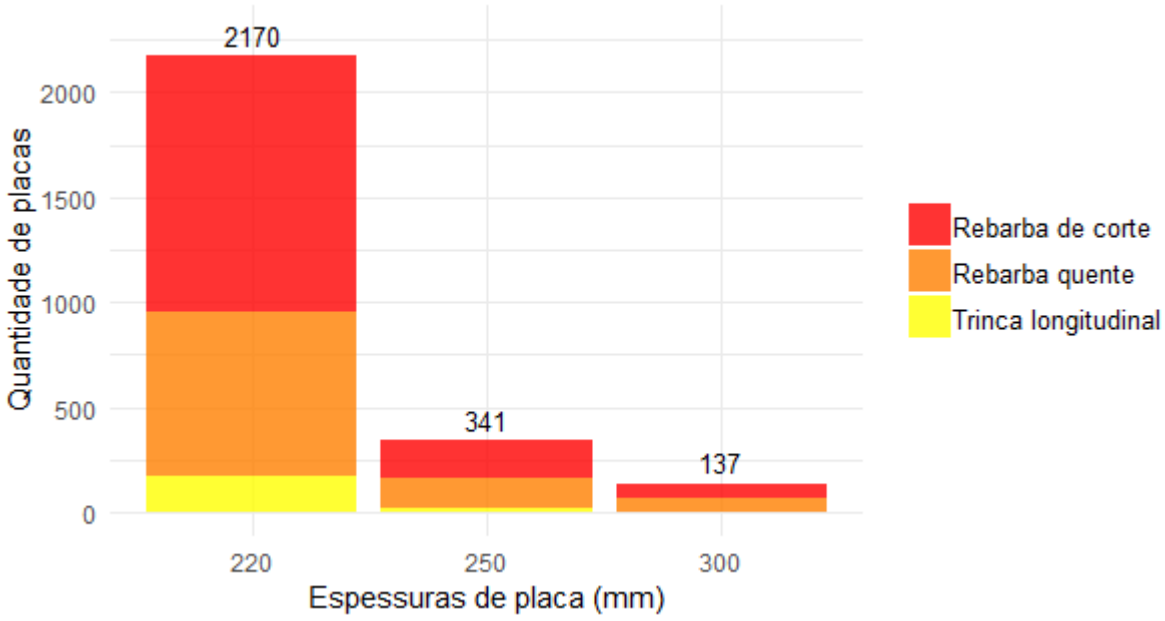
Gráfico 4. Contagem de placas por espessura de placa



Fonte: Elaborado pelo autor.

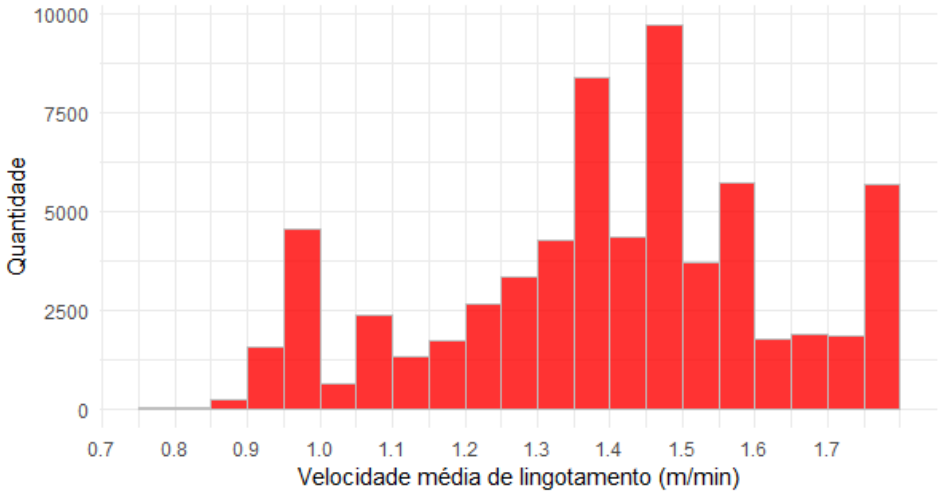
No gráfico 5 nota-se uma fração praticamente idêntica de defeitos para todas as espessuras, tendo as placas de 220 mm uma quantidade de rebarba de corte ligeiramente maior que as demais.

Gráfico 5. Contagem dos defeitos em relação à espessura de placa



Fonte: Elaborado pelo autor.

Gráfico 6. Histograma da velocidade média de lingotamento

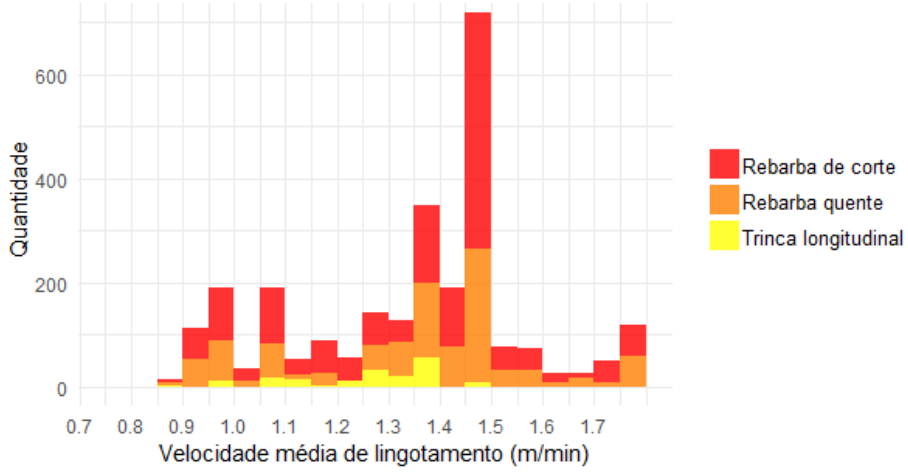


Fonte: Elaborado pelo autor.

Utilizando distribuição de frequência em um histograma, foi analisado no gráfico 6 como se distribui a velocidade média das placas produzidas no período. É possível perceber que a maior produção se concentra em torno de 1,45 m/min, com quantidades relevantes de placas sendo produzidos em torno de 1,80 m/min e outras quantidades relevantes em torno de 0,85 m/min. Essa variação se dá de acordo com a velocidade do fluxo de aço entre molde e distribuidor. A média amostral foi 1,41 m/min e o desvio padrão amostral, 0,23 m/min.

Quando plotada com os defeitos, vê-se um grande pico de rebarba de corte na região de 1.45 m/min e a maioria das trincas longitudinais concentradas entre 1.20 e 1.40 m/min, como mostra o gráfico 7.

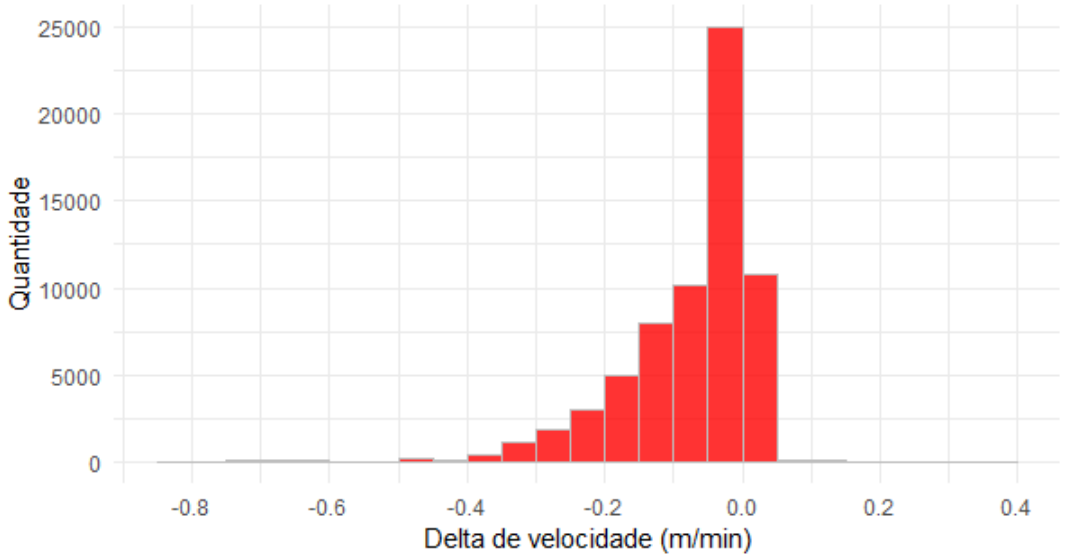
Gráfico 7. Histograma da velocidade média em relação aos defeitos



Fonte: Elaborado pelo autor.

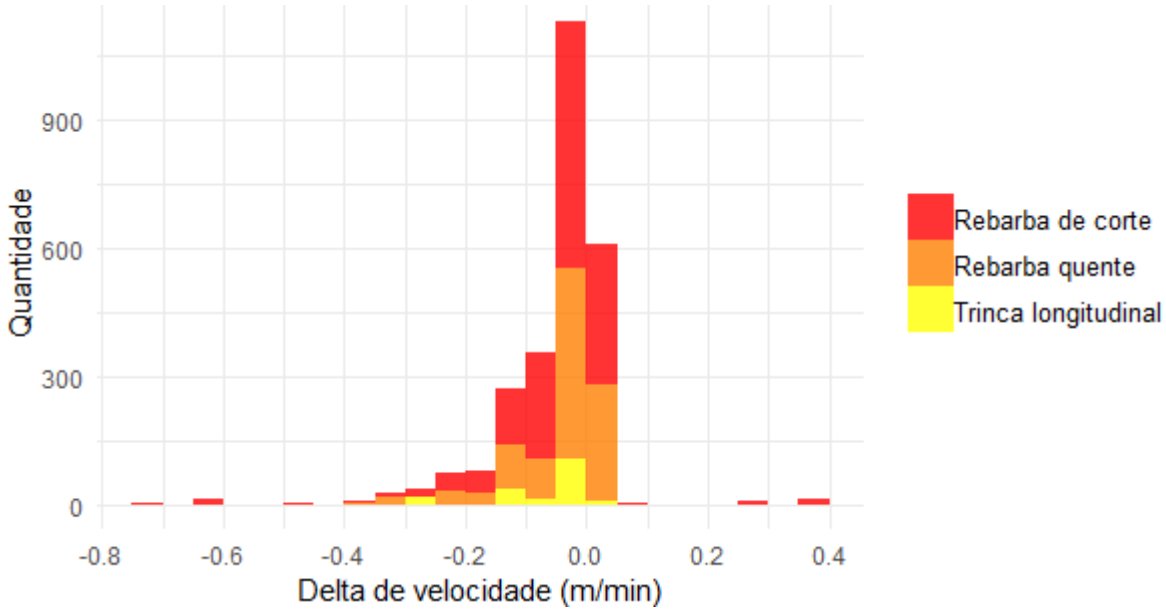
O gráfico 8 mostra como o delta de velocidade média de lingotamento está distribuído no processo. Percebe-se que o processo, na maioria dos casos, trabalha com uma velocidade média entre -0,10 m/min e 0,00 m/min da velocidade ideal de processo, com quantidades relevantes de placas sendo produzidas com velocidades de até 0,09 m/min acima e entre 0,11 m/min e 0,20 m/min abaixo. A média amostral foi -0,08 m/min e o desvio padrão amostral, 0,10 m/min.

Gráfico 8. Histograma do delta de velocidade média de lingotamento



Fonte: Elaborado pelo autor.

Gráfico 9. Histograma dos defeitos em relação ao delta de velocidade

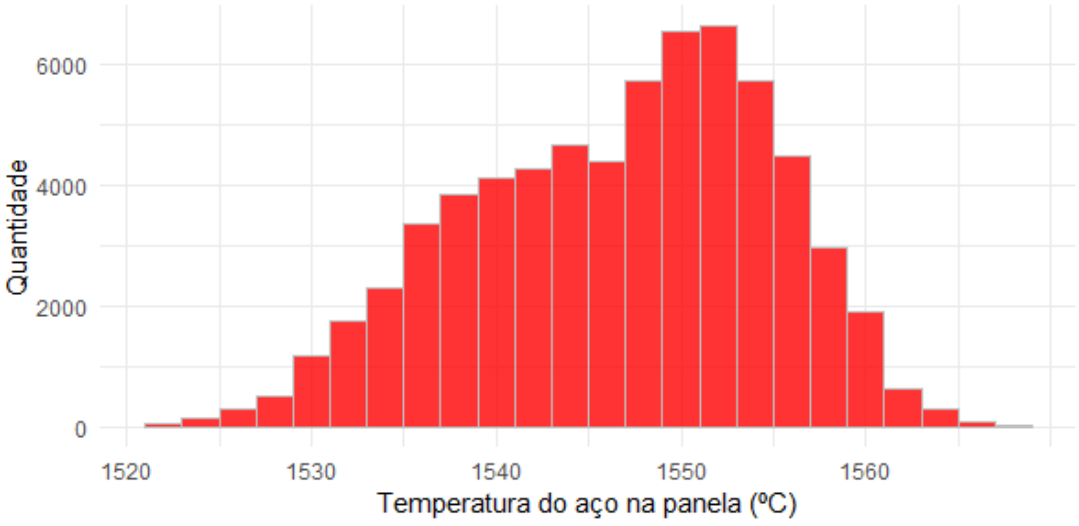


Fonte: Elaborado pelo autor.

Analisando o gráfico 9 da ocorrência de defeitos na distribuição de delta de velocidade é possível notar uma maior presença de defeitos nas velocidades mais próximas da velocidade ideal de processo, com uma distribuição menor ao longo dos outros deltas de velocidade.

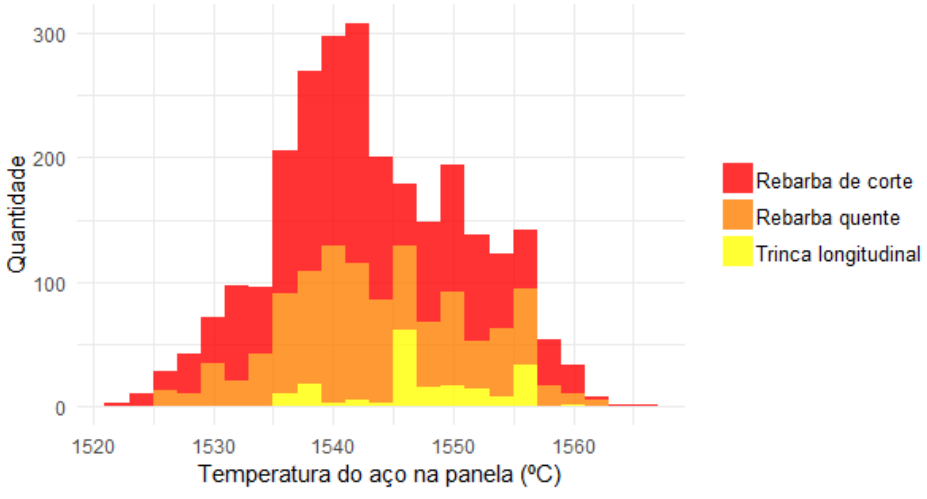
A seguir foi analisada a temperatura do aço que chega ao lingotamento, no gráfico 10. Os resultados mostram uma grande quantidade de placas com temperatura 1552ºC e uma gradual variação tanto para mais quanto para menos. A média amostral foi 1546,14ºC e o desvio padrão amostral, 8,18ºC.

Gráfico 10. Histograma das temperaturas de aço na panela



Fonte: Elaborado pelo autor.

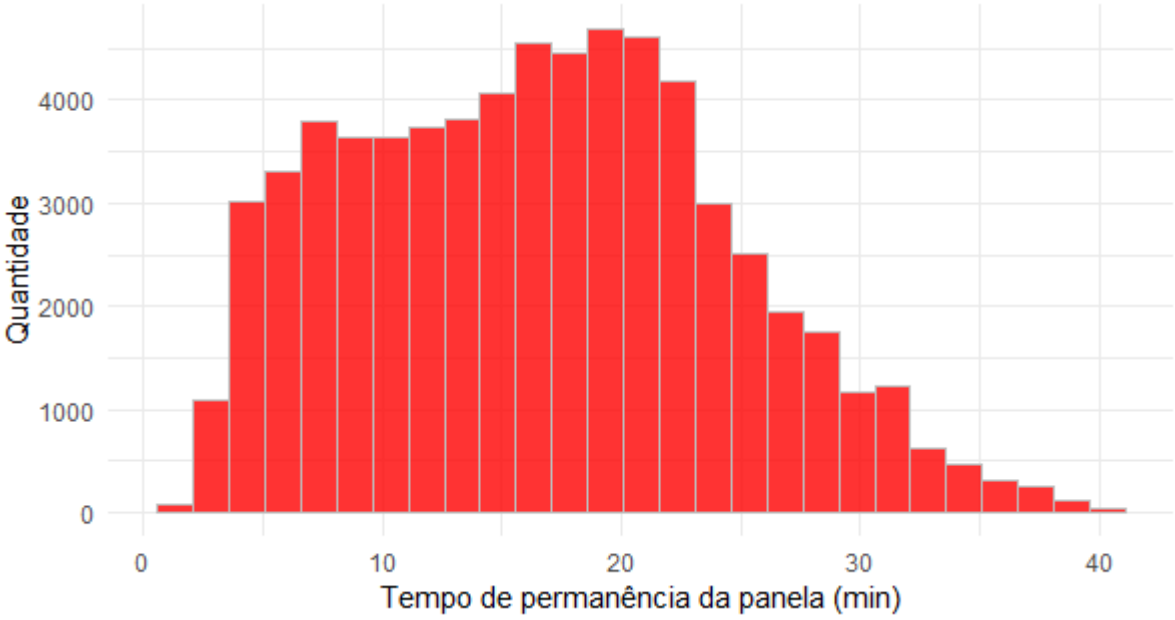
Gráfico 11. Histograma dos defeitos em relação às temperaturas de aço na panela



Fonte: Elaborado pelo autor.

Já quando confrontados os defeitos nessa mesma variável, nota-se uma maior concentração destes com picos mais próximos de 1540ºC para todos os defeitos, como mostrado no gráfico 11.

Gráfico 12. Histograma do tempo de permanência da panela



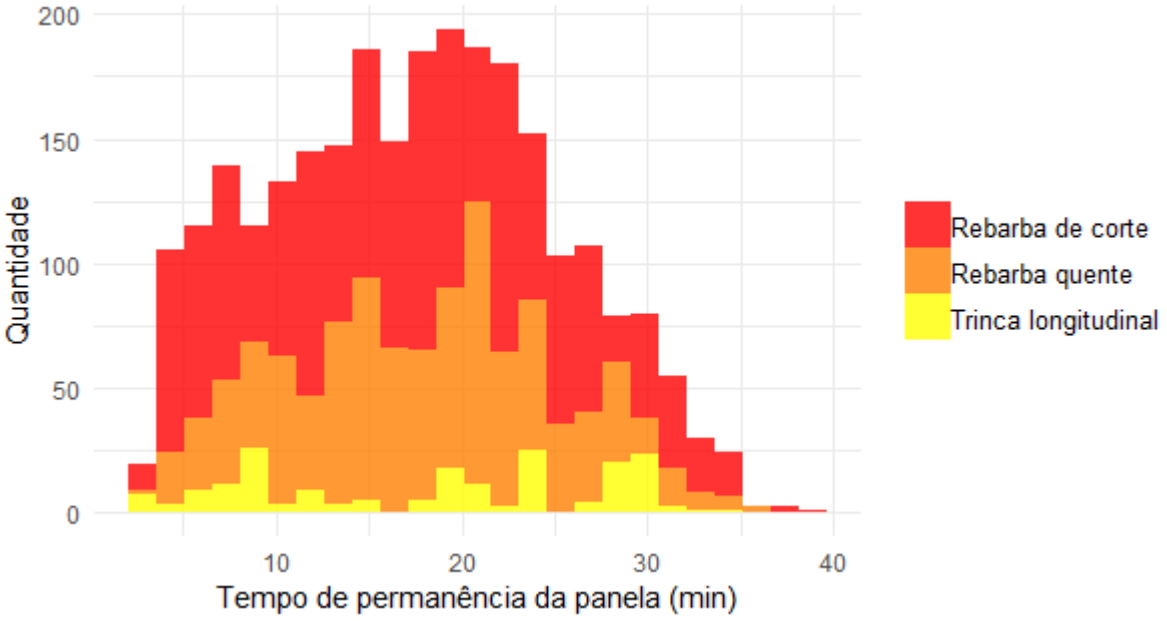
Fonte: Elaborado pelo autor.

No gráfico 12 foi analisado o tempo de permanência da panela na torre antes da abertura da panela. É possível notar que o processo utiliza, no maior número de casos, tempos de permanência de aproximadamente 19 minutos, com uma distribuição para mais e para menos, demonstrando um possível ajuste durante as corridas e tempo do processo. A média amostral foi 16,6 minutos e o desvio padrão amostral, 8,0 minutos.

Os defeitos quando distribuídos no gráfico 13 apresentaram um resultado semelhante, sem grandes observações a serem feitas.

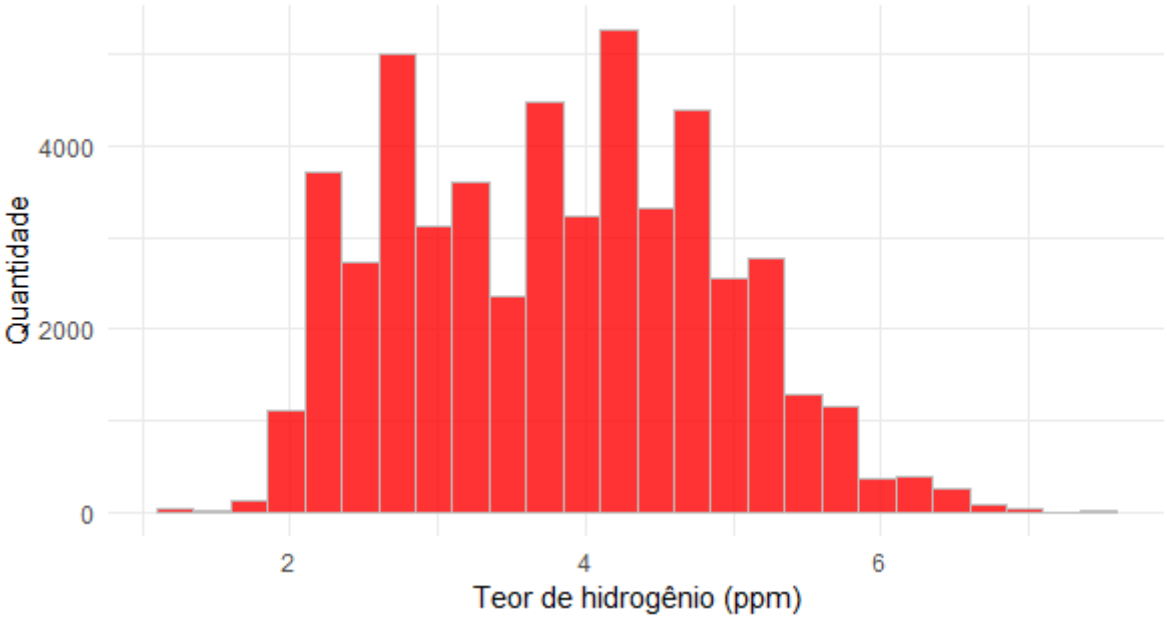
A distribuição do teor de hidrogênio (gráfico 14) durante o lingotamento apresentou picos variados, mostrando um baixo controle fino do teor de hidrogênio no processo, corroborado pelo desvio padrão amostral calculado. A média amostral foi 3,78 ppm e o desvio padrão amostral, 1,06 ppm.

Gráfico 13. Histograma dos defeitos em relação ao tempo de permanência



Fonte: Elaborado pelo autor.

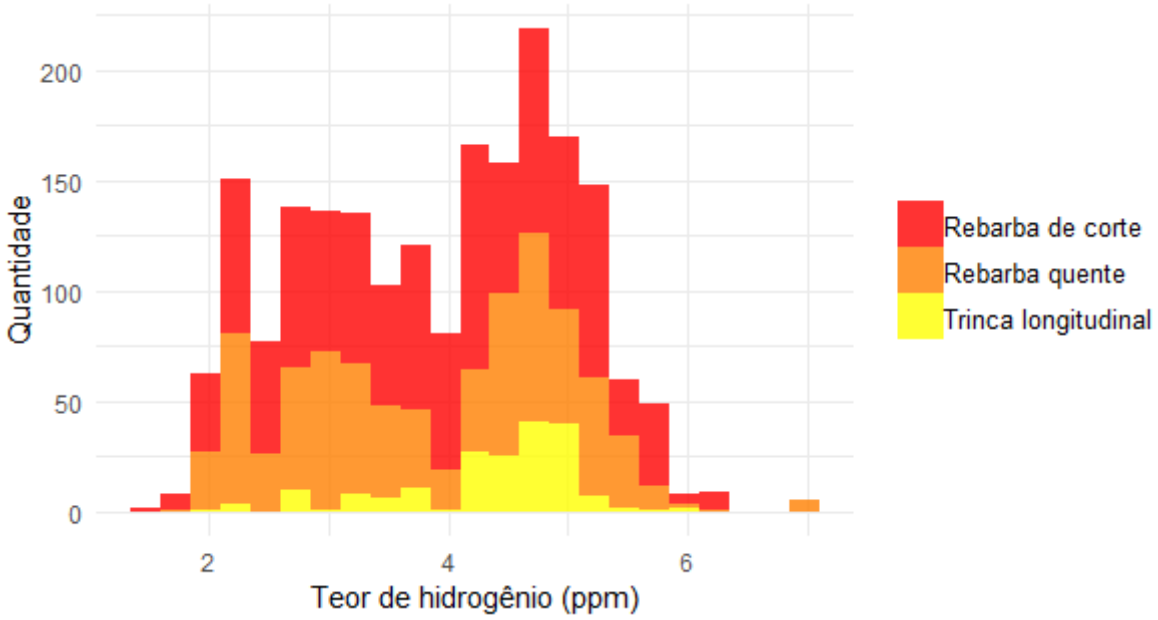
Gráfico 14. Histograma do teor de hidrogênio



Fonte: Elaborado pelo autor.

Quando distribuído em relação aos defeitos, apresentou dois picos relevantes em torno de 5 ppm para todos os defeitos e em torno de 3 ppm para defeitos de rebarba de corte e rebarba quente (gráfico 15).

Gráfico 15. Histograma dos defeitos em relação às temperaturas de panela



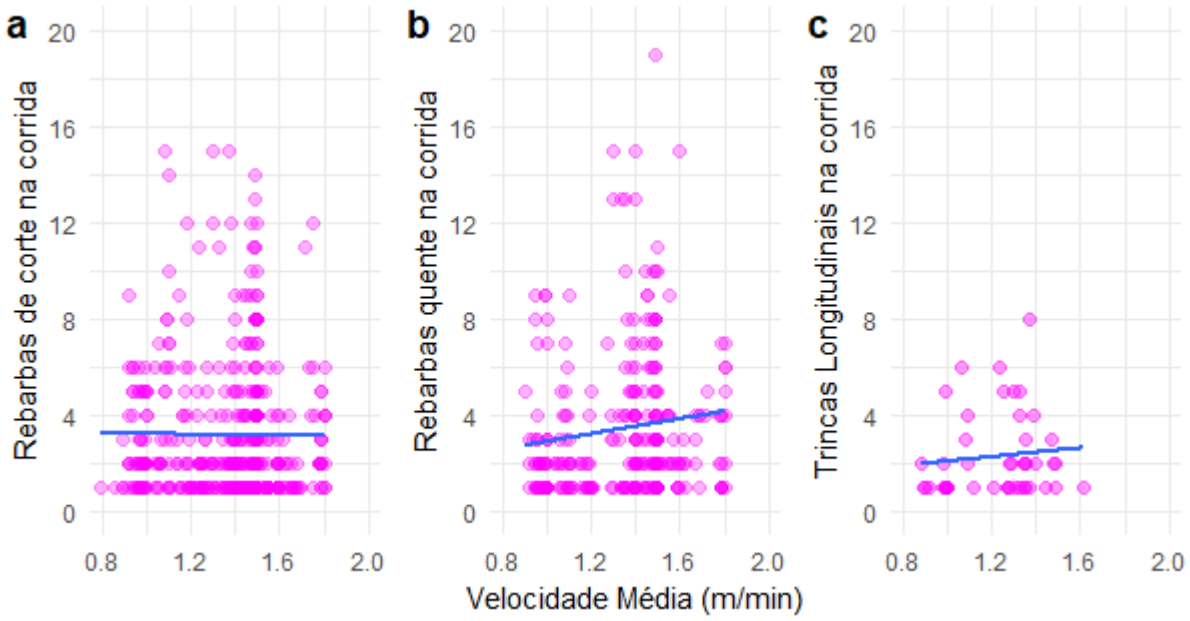
Fonte: Elaborado pelo autor.

3.4.2 Análises de correlação

Utilizando o método de correlação já discutido, foram analisadas as variáveis Velocidade média de lingotamento, Delta de velocidade, Teor de hidrogênio, Temperatura do aço e Tempo de permanência do aço, contra as variáveis modificadas de rebarba de corte, rebarba quente e trinca longitudinal. O objetivo dessa análise foi de levantar possíveis correlações entre os defeitos e as outras variáveis contínuas e aleatórias do conjunto de dados.

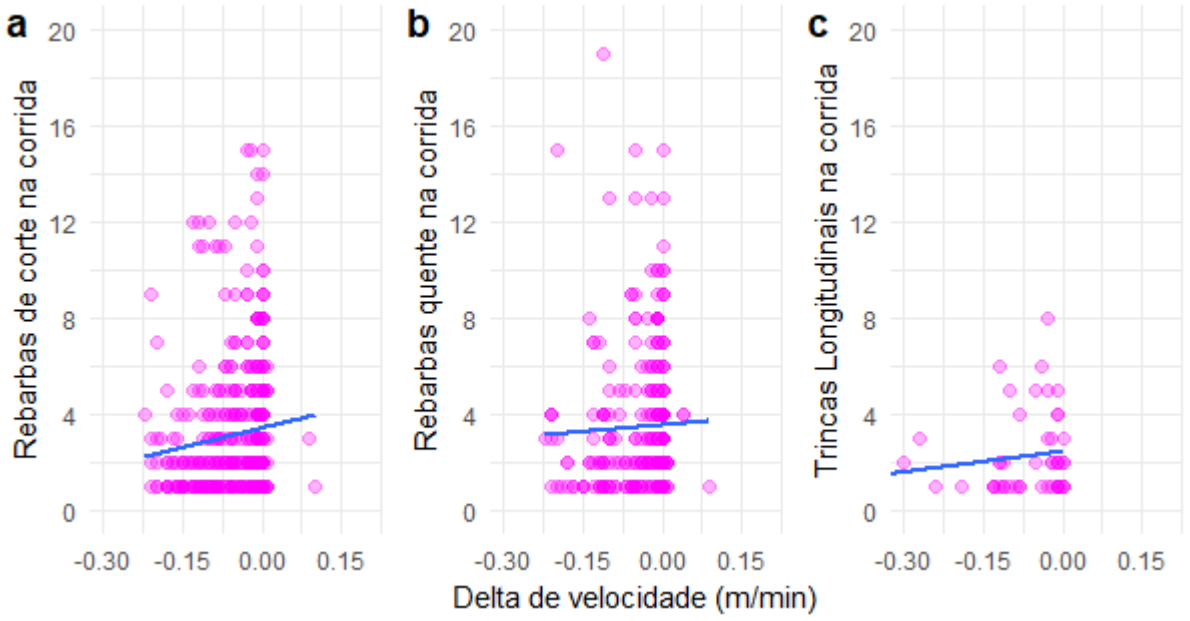
Foram plotados 16 gráficos de dispersão a fim de ser possível verificar visualmente uma correlação. As plotagens foram: 1 (uma) para cada variável contínua, formando um conjunto e 1 conjunto para cada defeito, e mais 1 gráfico com a dispersão de hidrogênio vs. o tempo de permanência da panela na torre de lingotamento. Os resultados são apresentados na página seguinte e em seguida discutidos.

Gráfico 16. Gráficos de dispersão de velocidade média vs. (a) rebarba de corte, (b) rebarba quente e (c) trincas longitudinais.



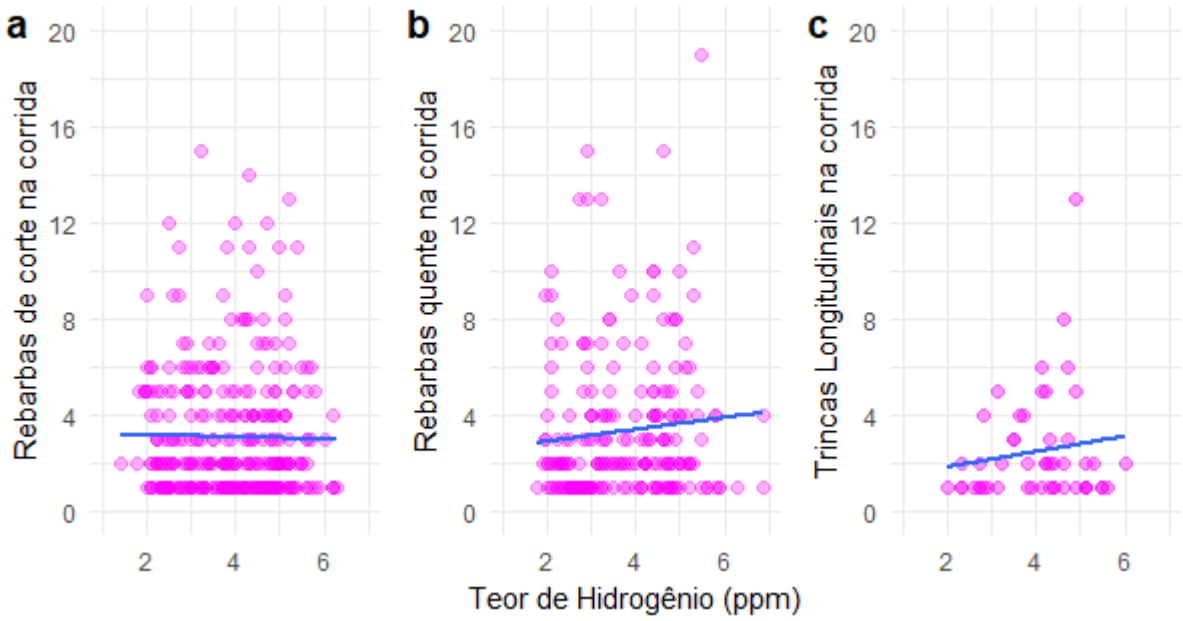
Fonte: Elaborado pelo autor.

Gráfico 17. Gráficos de dispersão de delta de velocidade vs. (a) rebarba de corte, (b) rebarba quente e (c) trincas longitudinais.



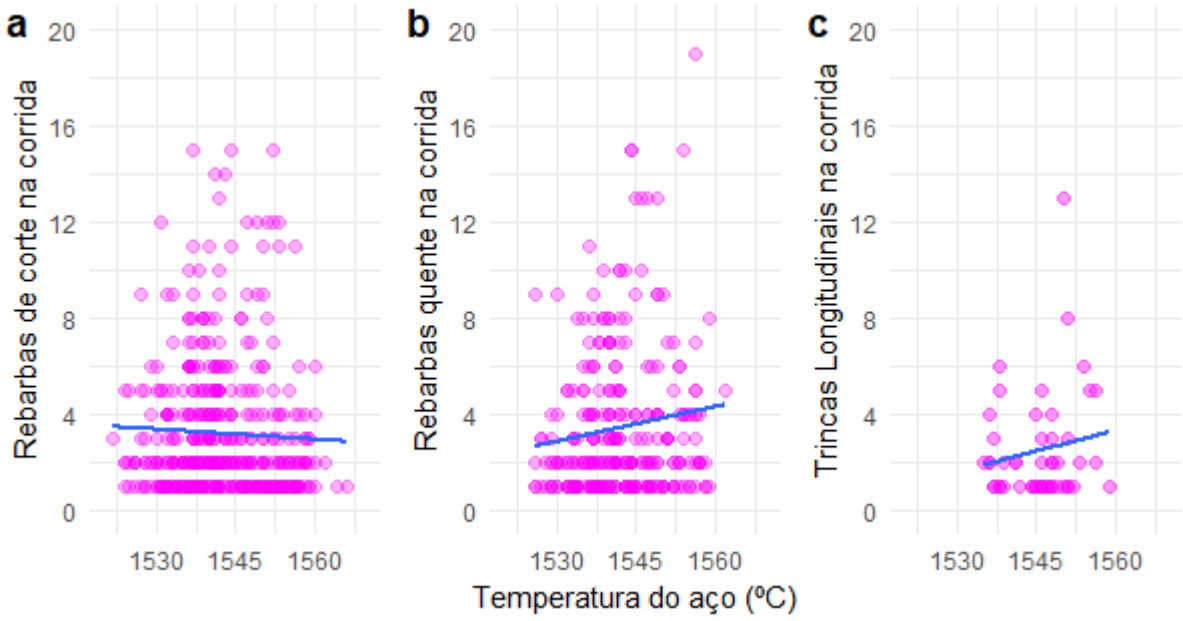
Fonte: Elaborado pelo autor.

Gráfico 18. Gráficos de dispersão de teor de hidrogênio vs. (a) rebarba de corte, (b) rebarba quente e (c) trincas longitudinais.



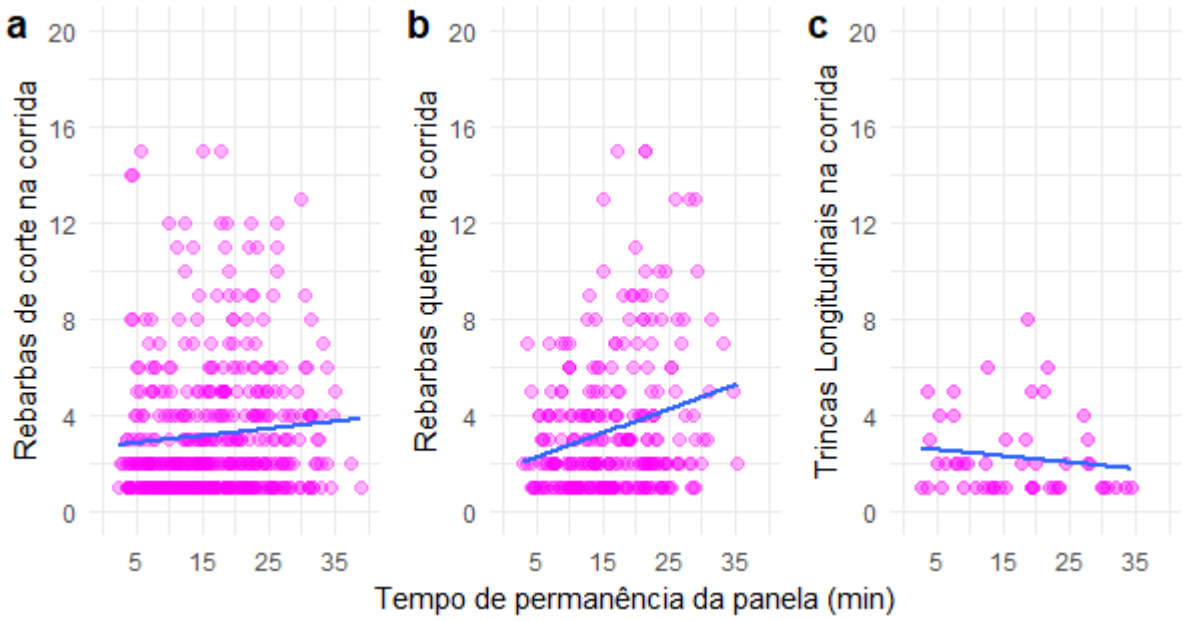
Fonte: Elaborado pelo autor.

Gráfico 19. Gráficos de dispersão de temperatura do aço vs. (a) rebarba de corte, (b) rebarba quente e (c) trincas longitudinais.



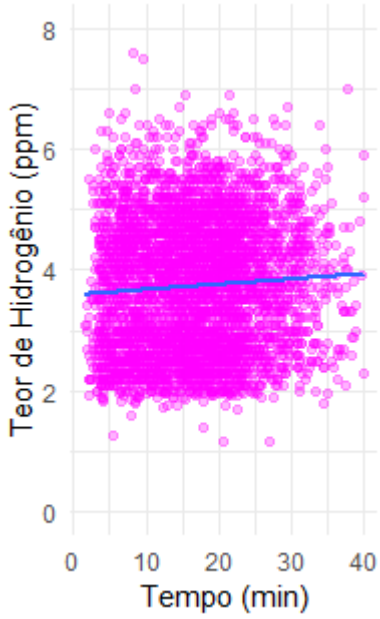
Fonte: Elaborado pelo autor.

Gráfico 20. Gráficos de dispersão de tempo de permanência da panela vs. (a) rebarba de corte, (b) rebarba quente e (c) trincas longitudinais.



Fonte: Elaborado pelo autor.

Gráfico 21. Gráfico de dispersão de teor de hidrogênio vs. tempo de permanência da panela



Fonte: Elaborado pelo autor.

Observando as dispersões com o auxílio de uma linha de regressão, que mostra a tendência desses pontos levando em consideração cada par (x,y), foi possível notar que houve tendências nos dados, algumas mais fortes como a presença de rebarba quente em panelas que

tiveram mais tempo de permanência (gráfico 20b), e outras tendências irrelevantes como a influência do teor de hidrogênio (gráfico 18a) e da velocidade média (gráfico 16a) em relação à rebarba de corte. É possível notar também uma tendência positiva no teor de hidrogênio em relação ao tempo de permanência da panela na torre de lingotamento (gráfico 21). Outras tendências foram negativas como a presença de trinca em panelas com mais tempo de permanência na torre (gráfico 20c) e a rebarba de corte em relação à temperatura do aço na panela (gráfico 19a).

Por fim, foi gerada a correlação com base no grupo de dados modificado que é apresentada na tabela 2.

Tabela 2. Valores de correlação das observações do grupo de dados modificado

Defeito/Variável **Variável**	**Rebarba de corte**	**Rebarba quente**	**Trinca longitudinal**	**Teor de hidrogênio**
Velocidade Média de Lingotamento	-0,001	0,023	-0,103	-
Delta de velocidade (real – programada)	0,055	0,028	0,033	-
Teor de hidrogênio	-0,019	-0,025	0,023	1
Temperatura do aço	-0,043	-0,017	-0,038	-
Tempo de permanência da panela na torre	0,083	0,017	-0,075	0,063

Fonte: Elaborado pelo autor.

4 CONCLUSÃO

Através da análise exploratória do processo de lingotamento contínuo da siderúrgica estudada constata-se que há um baixo ajuste fino do processo em termos de teores de hidrogênio no aço, evidenciado pela distribuição inconstante e formada por diversos picos ao longo do intervalo de valores na distribuição de frequência do teor de hidrogênio das placas produzidas no período.

Os aços do tipo 1, 2, 3 e 6 apresentaram uma maior ocorrência de rebarba de corte se comparados aos outros tipos de aço estudados e que os aços de tipo 4 e 9, podem ter uma tendência à trinca, entretanto, nesse último caso, é necessário uma quantidade amostral maior já que apenas uma porcentagem pequena dos aços apresentou tal defeito.

Os defeitos de rebarba de corte foram os que representaram a maior quantidade dentre os defeitos de placa da siderúrgica, consistindo 2,2% do total de placas produzidas no período, seguido do defeito de rebarba quente, representando 1,5% do total de placas.

O valor de correlação medido para as variáveis apresentou-se baixo, evidenciando uma correlação fraca entre as variáveis e seus defeitos. Tal fato pode ter relação com a grande variedade de situações impostas ao processo cujos valores não são computados na compilação dos dados. É possível que uma análise com mais variáveis e em um ambiente multidimensional trouxesse melhores resultados do ponto de vista de correlação.

Além disso, observando o conjunto de análises, pode-se formular as seguintes hipóteses:

1. Os defeitos de rebarba de corte, rebarba quente e trinca longitudinal ocorrem em maior quantidade quando submetidos a velocidades de lingotamento mais próximas da velocidade definida como ideal para o processo da siderúrgica, como corrobora os gráficos de dispersão, os histogramas de delta de velocidade e o valor de correlação calculado entre delta de velocidade e os defeitos estudados.
2. O teor de hidrogênio aumenta com o tempo de permanência da panela na torre de lingotamento, como evidenciado pelo gráfico de dispersão entre as duas variáveis e a correlação positiva entre elas.
3. O teor de hidrogênio tem relação com a presença de trincas longitudinais. Essa hipótese leva em consideração que houve correlação positiva entre o defeito e o teor de hidrogênio, que o histograma de trincas longitudinais relacionados ao teor de hidrogênio tem uma distribuição enviesada em direção aos valores de 5 ppm, enviesamento que não ocorre com os outros defeitos no histograma.

4.1 Sugestões de trabalhos futuros

1. Confirmação da relevância das hipóteses através de metodologias estatísticas mais robustas.
2. Realização das mesmas análises em um conjunto de dados contendo a composição química real (ou calculada) dos elementos da liga que compõe o aço estudado, levantando correlações e tendências da composição química e dos defeitos.
3. Realização da curva do teor de hidrogênio de uma panela de aço na torre de lingotamento.
4. Análise da variação do teor de hidrogênio em termos de desvio padrão e subsequente classificação do processo dentro do padrão sigma.
5. Teste da significância das correlações apresentadas.
6. Análise de regressão com defeitos designados como variáveis dependentes e as demais variáveis como independentes afim de serem levantadas novas inferências, como por exemplo, identificar a forma como cada variável afeta a quantidade de defeitos e quais variáveis afetam de modo significante.

REFERÊNCIAS

CAMPOS, V. F. **TQC:** controle da qualidade total: (no estilo japonês). Belo Horizonte: Fundação Christiano Ottoni, 1992.

CARDOSO DA ROCHA, V. **Estudo Comparativo entre Fluxantes Aplicados no Lingotamento Contínuo do Aço SAE 1046 MOD**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2014.

COELHO, G. C. **Solidificação:** Lingotamento Contínuo. Lorena: Universidade de São Paulo, 2013.

COMPANY, M. &. **Analytics comes of age**. [S.l.]: [s.n.], 2018.

GATIGNON, H. **Statistical Analysis of Management Data**. 1ª. ed. New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers, 2003.

HÄRDLE, W.; SIMAR, L. **Applied Multivariate Statistical Analysis**. 1ª. ed. [S.l.]: Springer, 2003.

IRVING, W. H. **Continuous casting of steel**. 1ª. ed. London: The Institute of Materials, 1993.
O'NEIL, C.; SCHUTT, R. **Doing Data Science:** Straight Talk from the Frontline. [S.l.]: O'Reilly Media Inc., 2013.

SELENE XIA, B.; GONG, P. Review of business intelligence through data analysis. **Benchmarking: An International Journal**, Vol. 21, n. 2, 2014. 300-311. Disponivel em: <https://doi.org/10.1108/BIJ-08-2012-0050>.

SILVA, D. G. **Indústria 4.0:** Conceito, Tendências e Desafios. Ponta Grossa: Universidade Tecnológica Federal do Paraná, 2017.

SILVERMAN, B. W. Density Estimation for Statistics and Data Analysis. **Monographs on Statistics and Applied Probability**, London, 1986.

TERRELL, G. R. **Mathematical Statistics - A Unified Introduction**. 1ª. ed. New York: Springer, 1999.

THOMAS, B. G. Continuous Casting. In: NA **Encyclopedia of Materials:** Science and Technology. [S.l.]: Elsevier Science Ltd., 2001. p. 1595-1599.

WANG, J. et al. Industrial Big Data Analytics: Challenges, Methodologies, and Applications, 2018.

WICKHAM, H.; GROLEMUND; GARRETT. **R for Data Science:** Import, Tidy, Transform, Visualize, and Model Data. 1ª. ed. Sebastopol: O'Reilly, 2017.