LVI SBPO SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
DE 04 ATÉ 07 NOVEMBRO 2024
FORTALEZA | CEARÁ

ANÁLISE ESTATÍSTICA PARA PREVISÃO DE PARADAS NÃO PROGRAMADAS EM UMA CÉLULA DE FUNDIÇÃO

Tháyna Alcântara Viera Lúcio
thaynaengmec@gmail.com
Universidade Federal de Itajubá
Avenida BPS, 1303, Itajubá – MG
José Arnaldo Barra Montevechi
montevechi@gmail.com
Universidade Federal de Itajubá
Avenida BPS, 1303, Itajubá – MG
Rafael de Carvalho Miranda
rafael.miranda@unifei.edu.br
Universidade Federal de Itajubá
Avenida BPS, 1303, Itajubá – MG
Jonathan Serafim Lúcio
serafim.epr@gmail.com
Universidade Federal de Itajubá
Avenida BPS, 1303, Itajubá – MG
Guilherme Ferreira Lopes
gf.lopes@yahoo.com.br
Universidade Federal de Itajubá
Avenida BPS, 1303, Itajubá – MG

RESUMO

Este estudo explora a combinação de Regressão Linear Múltipla (RLM) e modelagem ARIMA para prever o Tempo de Parada Total em uma célula de fundição de uma indústria automotiva. A análise estatística dos dados de manutenção corretiva foi realizada, considerando variáveis como Manutenção Elétrica, Mecânica e Ferramental. O método *Stepwise* foi empregado para selecionar variáveis significativas na modelagem RLM, com a análise de resíduos seguida pelo ajuste do modelo ARIMA. O modelo RLM ajustado explicou 95,38% da variabilidade no Tempo de Parada Total, enquanto o modelo combinado RLM + ARIMA apresentou um Erro Absoluto Médio Percentual (MAPE) de 10% na previsão dos tempos de parada. Esses achados sugerem que o modelo integrado pode fornecer previsões precisas, auxiliando na alocação eficiente de recursos de manutenção e na identificação de padrões subjacentes não capturados pela regressão linear.

PALAVRAS-CHAVE. Regressão linear múltipla, ARIMA, MAPE, Modelo Combinado, Previsão
Tópicos: POI – PO na Indústria

ABSTRACT

This study explores the combination of Multiple Linear Regression (MLR) and ARIMA modeling to predict Total Downtime in a casting cell of an automotive industry. Statistical analysis of corrective maintenance data was conducted, considering variables such as Electrical, Mechanical, and Tooling Maintenance. The Stepwise method was employed to select significant variables in the MLR modeling, with residual analysis followed by the adjustment of the ARIMA model. The adjusted MLR model explained 95.38% of the variability in Total Downtime, while the combined MLR + ARIMA model presented a Mean Absolute Percentage Error (MAPE) of 10% in downtime prediction. These findings suggest that the integrated model can provide accurate predictions, aiding in the efficient allocation of maintenance resources and identifying underlying patterns not captured by linear regression.

KEYWORDS. Multiple Linear Regression, ARIMA, MAPE, Combined Model, Forecasting
Paper topics: POI – PO in Industry

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

1. Introdução

A análise estatística é fundamental para a compreensão dos desafios industriais e empresariais, proporcionando insights valiosos que embasam decisões informadas. No contexto da Manutenção Industrial, a análise de dados de dispositivos e equipamentos é indispensável para identificar os fatores que afetam a eficiência e a produtividade [Silva 2016].

A previsão de dados, por sua vez, aplica modelos estatísticos para antecipar o comportamento futuro dos equipamentos com base em dados históricos. Técnicas como análise de séries temporais, regressão e modelos preditivos avançados, como redes neurais artificiais, são empregadas para estimar a vida útil dos componentes, determinar o momento ideal para manutenções preventivas e prevenir paradas não planejadas [Işikdag 2023].

Nesta pesquisa, foram aplicados métodos estatísticos avançados como a Regressão Linear Múltipla (RLM) e o modelo Autoregressive Integrated Moving Average (ARIMA). A RLM é amplamente utilizada como ferramenta preditiva em diversas frentes de estudo, pois permite a análise de múltiplos preditores simultaneamente [Abdullah et al. 2019; Yorifuji et al. 2019]. Já o modelo ARIMA é comumente empregado na previsão de séries temporais [Işikdag et al. 2023].

O objetivo deste estudo é analisar os dados relativos ao tempo total de paradas não programadas em uma célula de fundição. Para isso, será utilizada a regressão linear múltipla para desenvolver uma equação preditiva que represente adequadamente os dados e seja capaz de prever estados futuros. O modelo ARIMA será aplicado para aprimorar a previsão, utilizando os resíduos da modelagem por RLM como base. Assim, será construído um modelo combinado com o intuito de obter previsões mais precisas e robustas.

2. Referencial Teórico

2.1 Regressão Linear

Segundo Rabajante [2020], modelos matemáticos podem ser uma ferramenta valiosa de "primeiros socorros" diante de problemas complexos. A regressão linear é uma técnica estatística fundamental utilizada para entender a relação entre uma variável dependente (Y) e uma ou mais variáveis independentes (X). Seu objetivo é modelar essa relação linear por meio de uma equação matemática, permitindo prever ou estimar valores da variável dependente com base nos valores das variáveis independentes, fornecendo assim informações para a correlação de diferentes contextos (Furletti et al., 2022). Existem dois tipos principais de regressão linear:

- Regressão Linear Simples (RLS): É a forma mais básica de regressão linear, onde há apenas uma variável independente influenciando a variável dependente.
- Regressão Linear Múltipla (RLM): Envolve mais de uma variável independente que influencia a variável dependente.

Entre os principais parâmetros da regressão linear, destacam-se os coeficientes de regressão, que representam o efeito das variáveis independentes sobre a variável dependente. O coeficiente de determinação (R^2) indica a proporção da variabilidade da variável dependente explicada pelas variáveis independentes [Yorifuji et al., 2019]. Além disso, os resíduos, que são as diferenças entre os valores reais e os previstos pelo modelo, ajudam a identificar a parcela dos dados que o modelo não consegue explicar. A análise dos resíduos é essencial para avaliar a precisão do modelo [Oliveira et al., 2020].

2.1.1 Regressão Linear Múltipla

Os modelos de RLM são adequados para a análise de ambientes nos quais diversas variáveis afetam um único resultado [Lin et al., 2020]. A RLM é uma técnica de modelagem que explica o impacto de variáveis independentes sobre uma variável dependente. O modelo de regressão linear múltipla pode ser expresso pela seguinte Equação (1):

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_nX_n + \varepsilon \quad (1)$$

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

Na equação, o Y representa a variável dependente, a corresponde a constante, X_1, X_2,... X_n são as variáveis independentes, b_1, b_2,... b_n são os coeficientes de regressão estimados pelo método dos mínimos quadrados e ε é o erro associado.

O parâmetro estatístico R é amplamente utilizado para avaliar a precisão e a qualidade do ajuste da linha de regressão, variando de 0 a 1, sendo que valores mais elevados indicam uma relação linear mais forte entre as variáveis observadas e previstas [McLinden et al., 2016; Pourmohammad et al., 2020].

De acordo com Oliveira et al. [2020], alguns cuidados devem ser observados na aplicação da RLM, relacionados aos dados utilizados na construção do modelo. Esses cuidados incluem:

- Independência das Observações: Cada observação deve ser independente. A autocorrelação entre observações pode comprometer a precisão dos coeficientes estimados.
- Seleção de Variáveis no Modelo: Diversos métodos podem ser usados para identificar as variáveis independentes mais relevantes.
- Homocedasticidade: A variabilidade dos resíduos deve ser constante em todos os níveis das variáveis independentes.
- Independência dos Resíduos: Os resíduos não devem estar correlacionados entre si.
- Ausência de Outliers: Valores atípicos podem distorcer a regressão e devem ser tratados adequadamente.
- Número Suficiente de Observações: A amostra deve ser maior do que o número de variáveis independentes para garantir resultados confiáveis.

Para a seleção das variáveis no modelo desta pesquisa, foi utilizado o método *Stepwise*, que inclui variáveis progressivamente, realizando uma análise a cada nova inclusão para garantir sua relevância [Chimuco, 2015].

Atender a todos os pressupostos é essencial para garantir a validade e a confiabilidade dos resultados da RLM.

2.2 Autoregressive Integrated Moving Average (ARIMA)

O modelo ARIMA, também conhecido como modelo de Box-Jenkins, foi desenvolvido por George Box e Gwilym Jenkins em 1970, visando descrever séries temporais por meio de um modelo matemático. A sua forma mais geral para modelos não sazonais é representada por (p, d, q), onde p representa a parte autorregressiva (AR), d o grau de diferenciação (I), e q a ordem da média móvel (MA). Este modelo é aplicável tanto a séries estacionárias quanto a não estacionárias. O processo de construção de um modelo ARIMA envolve as etapas de especificação, identificação, estimação e verificação [Moretti e Toloi, 2018].

A metodologia Box-Jenkins envolve três fases principais: Identificação, Estimação e Teste, e Aplicação [Makridakis et al., 1998]. A Figura 1 resume essas etapas.

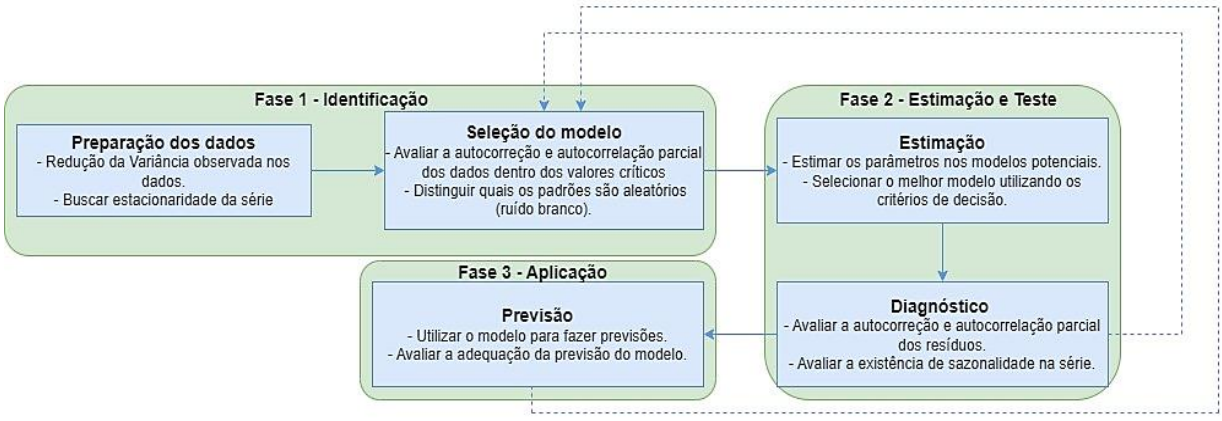


Figura 1: Esquema adaptado para utilização da metodologia de Box-Jenkins
Fonte: autores

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

- Fase de Identificação: A Função de Autocorrelação (ACF) é fundamental para a identificação de padrões em séries temporais, como sazonalidades e ciclos. Esta função desempenha um papel central ao avaliar a correlação entre os valores de uma série temporal em diferentes intervalos de tempo, permitindo a detecção de dependências ou relações temporais entre períodos sucessivos [Makridakis et al., 1998]. Em séries estacionárias, os erros exibem um comportamento conhecido como "ruído branco", caracterizado por variáveis aleatórias que não apresentam correlação entre si, o que significa que não há padrões sistemáticos, apenas flutuações aleatórias. Além da ACF, a Função de Autocorrelação Parcial (PACF) é igualmente relevante, pois mede a associação entre observações em diferentes períodos, excluindo a influência de observações intermediárias. Essa ferramenta é particularmente útil na determinação da ordem dos modelos autorregressivos (AR), uma vez que a PACF permite identificar a correlação direta entre um valor presente e seus valores passados, enquanto mantém constantes os demais fatores envolvidos [Makridakis et al., 1998]. Dessa forma, a PACF fornece uma visão mais precisa da dependência temporal dos dados, auxiliando na escolha do modelo de previsão adequado.
- Estimação e Teste: Nesse estágio, realiza-se a estimação dos coeficientes dos componentes autorregressivos (AR) e de médias móveis (MA), que compõem o modelo ARIMA, com base nos dados observados. Após a identificação do modelo, esses parâmetros são determinados para ajustar o modelo aos dados da série temporal. Durante o processo de estimação, alguns coeficientes podem ser considerados estatisticamente insignificantes, ou seja, apresentam um p-valor maior ou igual a 0,05. Nesses casos, os parâmetros podem ser excluídos do modelo, com o objetivo de melhorar o ajuste e a precisão das previsões. Além disso, uma avaliação detalhada dos resíduos deve ser conduzida, garantindo que não haja evidências de autocorrelação significativa ou de autocorrelação parcial entre os resíduos. A ausência de correlação nos resíduos indica que o modelo ajustado capturou adequadamente a estrutura da série temporal, pois os resíduos devem se comportar como "ruído branco", ou seja, sem padrões sistemáticos. Esse comportamento dos resíduos é um indicativo de que o modelo está bem ajustado e pronto para a etapa de previsão [Makridakis et al., 1998].
- Fase de Aplicação: Nessa etapa, o modelo ajustado é utilizado para realizar previsões com base nos dados da série temporal analisada [Makridakis et al., 1998]. Para avaliar a qualidade das previsões, diferentes métricas de desempenho podem ser empregadas, como o Erro Percentual Absoluto Médio (MAPE), que foi escolhido para este estudo. O MAPE é uma métrica comumente utilizada para medir a precisão do modelo, pois calcula a média dos erros percentuais absolutos entre os valores previstos e os observados, facilitando a interpretação dos resultados em termos percentuais. Embora outras métricas, como o Erro Quadrático Médio (RMSE) ou o Coeficiente de Determinação (R^2), possam ser relevantes na avaliação de previsões, o MAPE foi selecionado neste artigo por sua simplicidade e clareza na apresentação dos erros, permitindo uma comparação direta entre os modelos.

2.3 Manutenção Corretiva

Baldez et al. [2022] destacam que a manutenção corretiva é empregada em situações extremas de falha do equipamento ou máquina, devendo ser evitada, pois compromete sua vida útil e pode resultar em perdas significativas na produção. Além disso, ela requer um estoque amplo de peças e acarreta custos elevados para as empresas. Dhillon [2002] afirma que a manutenção corretiva é uma ação imprevista, acionada quando o equipamento falha ou quebra, exigindo reparos imediatos, fora do cronograma de manutenção planejado. Essa modalidade de manutenção é realizada para corrigir falhas, panes ou quebras em máquinas, quando estas não executam suas funções conforme projetado [Branco Filho 2008]. Em termos de prejuízo na produção, uma pausa para realizar a manutenção corretiva pode acarretar altos custos, muitas vezes considerados aceitáveis pela equipe de gestão dos equipamentos [Xenos 2004].

No entanto, se essa prática se tornar recorrente e não for devidamente equilibrada, os custos podem aumentar substancialmente, ultrapassando o orçamento planejado pela empresa. A

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

imprevisibilidade dessas interrupções pode resultar no desperdício de recursos de produção e manutenção a qualquer momento [Branco Filho 2008].

Kardec e Nascif [2009] salientam que cerca de 20% dos problemas originados por falhas contribuem para o aumento do tempo de inatividade do equipamento, acarretando consequências que podem representar até 80% dos prejuízos financeiros. Portanto, é fundamental que a equipe responsável pelo planejamento e controle da manutenção estabeleça um cronograma a ser rigorosamente seguido e busque ao máximo evitar paradas não planejadas.

3. Metodologia

O presente trabalho classifica-se como uma pesquisa científica com abordagem empírico-analítica, fundamentando-se em métodos quantitativos e estatísticos [Gamboa 2012]. Originou-se da seguinte indagação: Existem resultados significativos na previsão de dados utilizando um modelo combinado, que emprega o método de Regressão Linear Múltipla associado à modelagem ARIMA?

Para responder a essa questão, foi conduzida uma revisão bibliográfica inicial, explorando as metodologias estatísticas de RLM e ARIMA, considerando sua aplicabilidade no contexto da análise do setor de manutenção industrial. Os dados foram obtidos de uma célula de fundição de uma indústria automotiva especializada na produção de peças e acessórios para veículos de passeio. Com o objetivo de realizar o estudo, foram analisados dados referentes à manutenção corretiva (paradas não programadas) do primeiro semestre de 2023. Esses dados foram tratados para reduzir ruídos, como erros de digitação/apontamento e outliers. Com a aplicação de ferramentas de estatística descritiva, foi possível construir um gráfico de Pareto que possibilitou a classificação dos principais tipos de paradas e, consequentemente, a seleção das variáveis a serem incorporadas no estudo.

No contexto da RLM foram estruturadas as variáveis para análise. A variável dependente (Y) foi definida como Tempo de Parada Total (em minutos) decorrente de paradas não planejadas. As variáveis independentes consideradas são Manutenção Elétrica (X_1), Manutenção Mecânica (X_2) e Manutenção Ferramental (X_3).

Na modelagem dos dados desta pesquisa, foi adotada uma abordagem integrada que combina o método de RLM e a modelagem ARIMA. Inicialmente, realizou-se uma análise exploratória dos dados para identificar padrões e tendências no tempo de parada total em manutenção corretiva. Também foram examinadas as correlações entre as variáveis independentes — Manutenção Elétrica (X_1), Manutenção Mecânica (X_2) e Manutenção Ferramental (X_3) — e a variável dependente, Tempo de Parada Total (Y). Em seguida, verificou-se a estacionaridade dos dados referentes ao Tempo de Parada Total, utilizando os gráficos ACF e PACF.

Posteriormente, construiu-se o modelo por meio da RLM, utilizando 63 dados, sendo X_1, X_2 e X_3 as variáveis preditoras. Na construção do modelo, foram consideradas interações de até terceira ordem, e o método *Stepwise* foi empregado. Esse método automático identifica os termos e/ou preditores úteis para o modelo, adicionando a variável mais significativa ou removendo a variável menos significativa com base no nível de confiança de 95% ($\alpha = 0{,}05$).

Após a construção do modelo, foi realizada a análise dos resíduos através do estudo da autocorrelação. Em seguida, ajustou-se o modelo ARIMA para modelar os resíduos, definindo os parâmetros (p, d, q) e aplicando os gráficos ACF e PACF, e a diferenciação.

Por fim, para a validação, realizou-se a previsão do Tempo de Parada Total utilizando os 10 dados reservados para esse propósito. Empregou-se o modelo combinado, que consiste na soma das previsões obtidas pela regressão linear com a previsão dos resíduos obtida pelo modelo ARIMA. A comparação com os valores reais foi feita utilizando o MAPE como principal métrica de avaliação dos resultados.

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

4. Análises e Discussões

4.1 Análise exploratória dos dados

Na análise exploratória dos dados, foram realizadas duas abordagens principais: a visualização da variável dependente Tempo de Parada Total em uma série temporal (Figura 1) e a análise de correlação de Pearson entre as variáveis independentes (elétrica, mecânica e ferramental) e a variável dependente (Tabela 1).

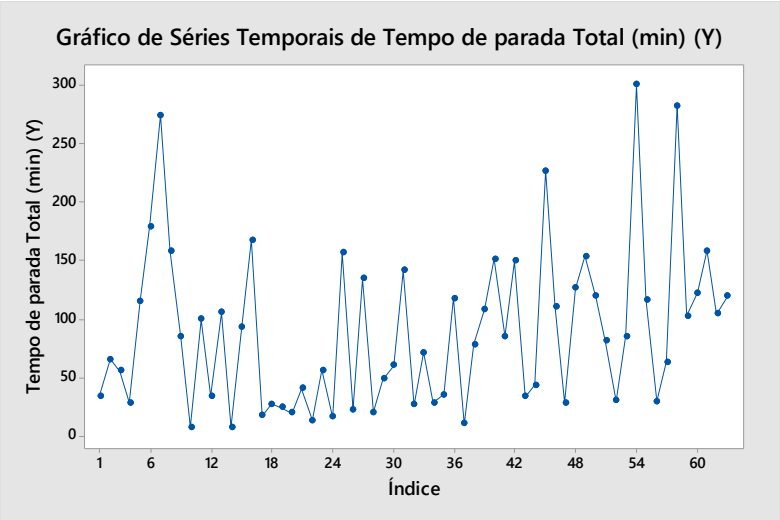


Figura 2: Gráfico de Séries Temporais do Tempo de Parada Total

Tabela 1: Coeficientes de Correlação

Coeficiente de Correlação de Pearson	
Tempo de parada total e Manutenção Elétrica	0,426
Tempo de parada total e Manutenção Mecânica	0,232
Tempo de parada total e Manutenção Ferramental	0,729

A análise do gráfico de séries temporais revela uma alta variabilidade e a presença de diversos outliers, sem uma tendência clara ou padrão sazonal. Esse comportamento não inviabiliza a aplicação do modelo ARIMA ao conjunto de dados; pelo contrário, as análises podem auxiliar na identificação de padrões subjacentes e possíveis causas dos tempos de parada, permitindo melhorias no processo, na gestão de manutenções e na detecção de padrões ocultos na variabilidade dos dados.

Os coeficientes de correlação de Pearson foram de 0,426 para a manutenção elétrica, 0,232 para a manutenção mecânica e 0,729 para a manutenção ferramental, indicando a força das relações entre cada variável independente e a variável dependente. Esses coeficientes sugerem correlações moderadas a fortes, sendo a manutenção ferramental a variável com a correlação mais robusta. Tais resultados corroboram a adequação do modelo de RLM, destacando a relevância das variáveis analisadas para explicar a variação observada.

4.2 Estacionaridade dos dados

A estacionaridade dos dados para a variável Tempo de Parada Total foi analisada por meio dos gráficos ACF e PACF. Os resultados sugerem que os *lags* estão dentro dos limites de significância de 95%, conforme exibido na Figura 3.

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

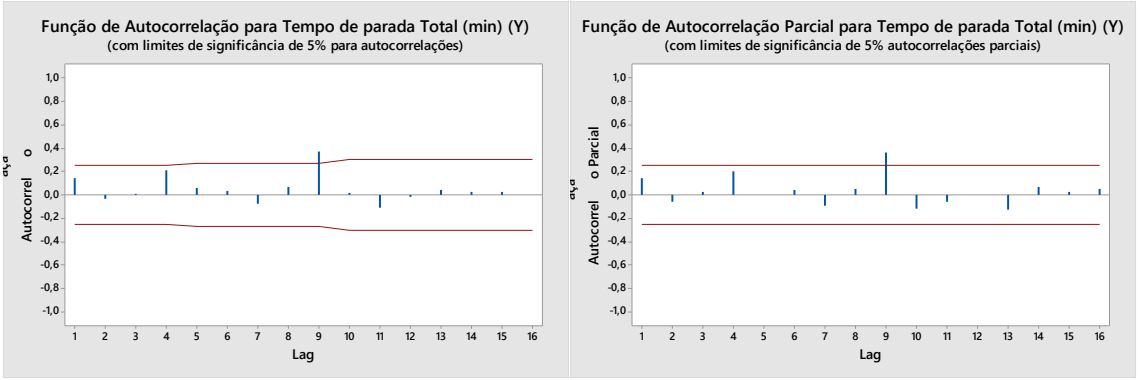


Figura 3: Gráficos para análise de estacionaridade dos dados – ACF (esquerda) e PACF (direita)

4.3 Modelo de Regressão Linear Múltipla

O ajuste do modelo de regressão foi realizado considerando interações até a terceira ordem. Além disso, foi utilizado o método *Stepwise*, uma ferramenta automática que identifica os preditores mais relevantes para o modelo. Esse processo adiciona a variável mais significativa ou remove a menos significativa com base em um nível de confiança de 95% ($\alpha = 0,05$). A Figura 4 apresenta o resultado da modelagem da equação de regressão.

Sumário do Modelo

S	R2	R2(aj)	R2(pred)
15,0959	95,38%	95,14%	94,69%

Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	11,24	2,96	3,80	0,000	
Elétrica	1,0315	0,0586	17,60	0,000	1,02
Mecânica	1,0841	0,0773	14,02	0,000	1,07
Ferramental	1,0383	0,0339	30,62	0,000	1,08

Equação de Regressão

Tempo de parada Total (min) (Y) = **11,24** + **1,0315** Elétrica + 1,0841 Mecânica + 1,0383 Ferramental

Figura 4: Resultados gerados

O método aplicado gerou a equação de regressão com os termos estatisticamente significativos, com base no valor do *p-value*, todos inferiores a 0,05. Os valores de R^2 demonstram a consistência do modelo, com R^2 de 95,38% e R^2 ajustado de 95,14%, indicando a proporção da variação da resposta explicada pelo modelo. O R^2 predito, de 94,69%, reflete a capacidade preditiva do modelo. Os valores previstos pela equação de regressão estão apresentados na Tabela 2, no tópico 4.5 deste trabalho.

A Figura 5 apresenta graficamente os resultados dos resíduos do modelo, cuja análise é essencial para avaliar sua adequação e precisão. Os resíduos do modelo de RLM foram armazenados como uma série de dados, uma vez que foram utilizados na construção do modelo ARIMA.

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

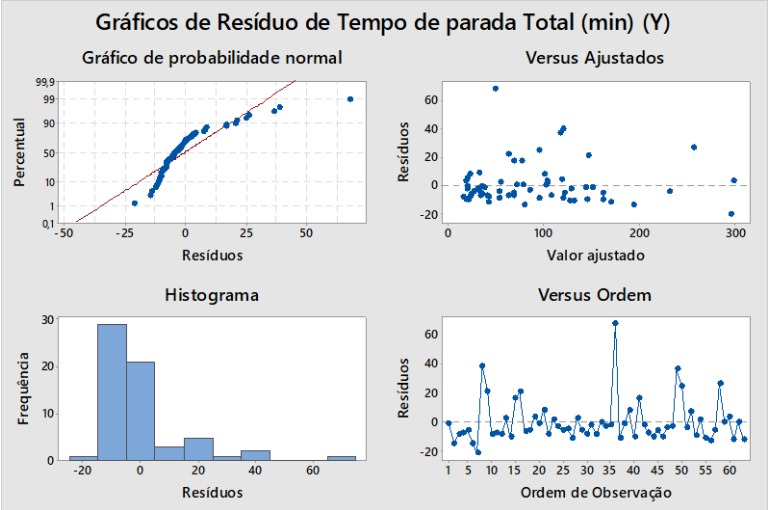


Figura 5: Resultados dos resíduos do modelo

Os resíduos representam as diferenças entre os valores observados e os valores previstos pelo modelo, sendo que uma distribuição dos resíduos próxima à normalidade sugere que o modelo é adequado para explicar a variação da variável *Y*. Ao examinar os resíduos gerados, verifica-se que sua distribuição se aproxima de uma distribuição normal, com a maioria dos valores centrados em torno de zero e uma dispersão uniforme ao longo dos valores previstos.

Além disso, não foram identificados padrões discerníveis nos resíduos, como heterocedasticidade ou autocorrelação, o que indica que o modelo captura adequadamente a estrutura subjacente dos dados. Com base nessa análise, conclui-se que o modelo de regressão linear múltipla, construído pelo método *stepwise*, é apropriado para explicar a relação entre as variáveis independentes e a variável dependente em estudo.

4.4 Modelo de previsão ARIMA para os resíduos

O modelo ARIMA foi aplicado aos resíduos da modelagem por regressão linear, complementando o modelo final de previsão dos dados. Após a aplicação da diferenciação, a série apresentou um comportamento adequado. Os gráficos ACF e PACF (Figura 6) demonstram os resultados obtidos com a diferenciação de ordem 1 (*d = 1*).

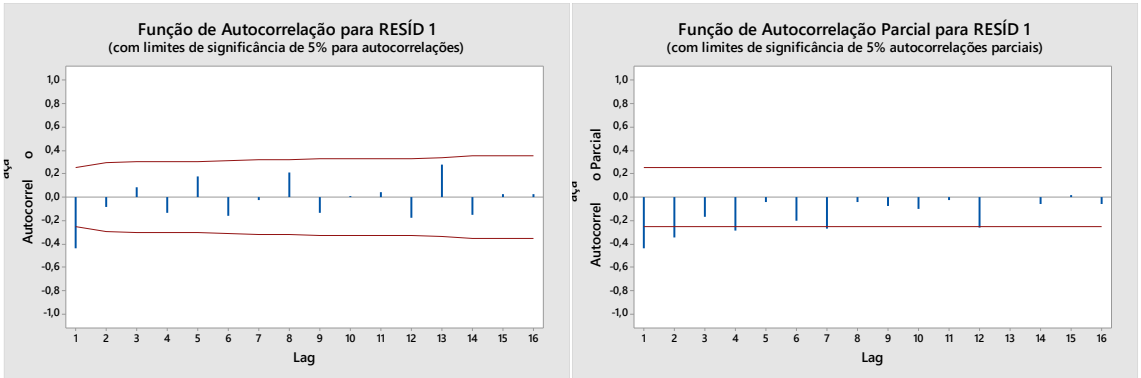


Figura 6: Gráficos de estacionaridade dos resíduos – ACF (esquerda) e PACF (direita)

A partir desses gráficos, os parâmetros (p, d, q) do modelo ARIMA foram definidos. Com uma diferenciação *d = 1*, o gráfico ACF apresentou um *lag* significativo, determinando *q = 1*, enquanto o gráfico PACF mostrou cinco *lags* significativos, resultando em *p = 5*. Portanto, os

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

parâmetros definidos e utilizados para o modelo ARIMA aplicado aos resíduos foram (5, 1, 1). A Figura 7 apresenta as previsões do modelo ARIMA para os resíduos.

Previsões do período 63

		Limites de 95%	
Período	Previsão	Inferior	Superior
64	-8,16214	-41,9983	25,6740
65	-1,40652	-36,7338	33,9208
66	-9,72610	-45,7060	26,2538
67	-3,30685	-41,0214	34,4077
68	-8,62150	-47,2933	30,0503
69	-5,44540	-47,5918	36,7009
70	-6,22092	-49,6850	37,2431
71	-6,01906	-50,8422	38,8041
72	-6,44488	-52,6674	39,7777
73	-5,92065	-53,5380	41,6967

Figura 7: Resultado da previsão pelo modelo ARIMA dos resíduos

A partir da primeira observação dos dados gerados pela previsão utilizando o modelo ARIMA, é possível notar a presença de valores negativos, o que deve ser contextualizado, uma vez que valores negativos não são esperados para a variável tempo. No entanto, para a análise final da previsão, esses dados não são considerados definitivos. Para essa aplicação, será utilizado o Modelo Combinado (RLM + ARIMA), o qual será abordado no próximo tópico.

4.5. Previsão

Para a previsão dos dados, foi realizada a combinação dos métodos RLM e ARIMA. Os 10 dados reservados inicialmente para a análise final foram utilizados nesta etapa, e o desempenho do modelo foi avaliado por meio do MAPE. O cálculo do MAPE, descrito pela Equação (2), foi empregado para essa avaliação.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right| x\ 100 \quad (2)$$

No qual,
A_i: valor real
P_i: valor previsto
n: número total de observações

Os dados previstos para o Tempo de Parada Total resultaram da aplicação do modelo combinado. A previsão com base nos resíduos, utilizando o modelo ARIMA, abrange a parte dos dados que a regressão não conseguiu explicar. Dessa forma, o modelo combinado proporciona uma previsão mais completa. Os valores das previsões para os 10 dados estão apresentados na Tabela 2.

Tabela 2: Resultados previstos pelo modelo e MAPE

Previsão pelo modelo de RLM	Previsão pelo modelo ARIMA	Tempo de parada total previsto (RLM + ARIMA)	Tempo de parada total real	Diferença (%)
46	-8,16	38	33	14%
156	-1,41	154	170	9%
38	-9,73	29	26	10%
268	-3,31	264	268	1%
177	-8,62	169	218	23%
116	-5,45	111	102	9%
182	-6,22	175	172	2%

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

Previsão pelo modelo de RLM	Previsão pelo modelo ARIMA	Tempo de parada total previsto (RLM + ARIMA	Tempo de parada total real	Diferença (%)
200	-6,02	194	182	7%
113	-6,44	106	110	3%
193	-5,92	187	243	23%
			MAPE	**10%**

A previsão realizada pelo modelo combinado apresentou resultados significativos. O MAPE obtido foi de 10%, indicando que, em média, as previsões do modelo desviam-se dos valores reais em 10%. Esse nível de precisão é considerado aceitável em muitas aplicações práticas, sugerindo uma boa capacidade preditiva do modelo. O MAPE relativamente baixo evidencia que as previsões estão próximas dos valores reais, ressaltando a eficácia do modelo.

A Figura 8 mostra uma representação gráfica da aplicação do modelo, com a linha azul representando os dados reais e a linha vermelha os dados previstos.

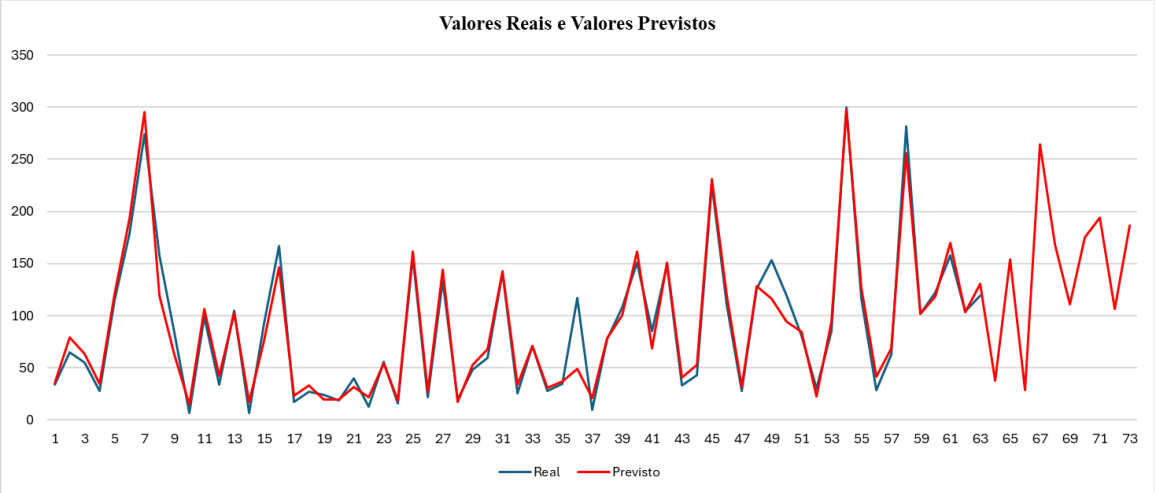


Figura 8: Representação gráfica dos valores reais e valores previstos pela RLM

O gráfico compara os valores reais (linha azul) com os valores previstos (linha vermelha), mostrando que, embora o modelo de previsão acompanhe bem as tendências gerais dos dados, há discrepâncias significativas em alguns picos, indicando superestimação em certos pontos. Em várias áreas, o modelo captura a direção das mudanças, mas em outras há erros de previsão mais elevados. Apesar da volatilidade dos dados, o modelo consegue seguir o padrão, o valor do MAPE em 10% sugere que, no geral, as previsões estão bastante alinhadas com os valores reais ao longo dos dados analisados.

5. Conclusão

Este estudo investigou a viabilidade e eficácia de combinar métodos de RLM e ARIMA para prever o Tempo de Parada Total em uma célula de fundição de uma indústria automotiva. Os resultados obtidos foram significativos e apontam para a utilidade desse modelo combinado na análise e previsão de tempos de paradas não programadas.

Na construção por RLM utilizando o método *Stepwise*, foi construído um modelo de regressão que identificou as variáveis independentes significativas e suas interações. Na análise dos resíduos, estes foram analisados e um modelo ARIMA foi ajustado para capturar a autocorrelação residual, e enriquecer a previsão dos dados.

Em relação aos resultados encontrados, pode-se dizer que as correlações de Pearson indicaram uma relação moderada a forte entre as variáveis independentes e a variável dependente,

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br

especialmente para a manutenção ferramental (r = 0,729). O ajuste do Modelo de RLM apresentou um R^2 de 95,38% e um R^2 ajustado de 95,14%, indicando que a maior parte da variabilidade no Tempo de Parada Total foi explicada pelo modelo. Na validação do modelo combinado, o modelo (RLM + ARIMA) apresentou um Erro Absoluto Médio Percentual (MAPE) de 10% ao prever os tempos de parada total dos 10 dados reservados, demonstrando boa capacidade preditiva.

Os resultados sugerem que o uso combinado de RLM e ARIMA pode melhorar significativamente a previsão de paradas não programadas em ambientes industriais. Esse modelo integrado fornece previsões mais precisas, ajudando na alocação eficiente de recursos de manutenção, além de identificar padrões ocultos a partir da análise ARIMA dos resíduos, que pode revelar padrões subjacentes que não são capturados apenas pelo modelo de regressão.

Uma oportunidade para pesquisas futuras consiste na exploração de modelos ARIMA mais complexos, bem como de outros modelos de séries temporais, com o objetivo de aprimorar a precisão das previsões. Além disso, seria pertinente investigar a aplicação do modelo combinado (RLM + ARIMA) de previsão em diferentes contextos.

Agradecimentos

Os autores agradecem a CNPq, CAPES e FAPEMIG pelo apoio à esta e outras pesquisas.

Referências

Abdullah, S., Nasir, N.H., Ismail, M., Ahmed, A.N., Jarkoni, M.N., 2019. Development of ozone prediction model in urban area. Int. J. Innov. Technol. Explor. Eng. 8 (10), 2263–2267. https://doi.org/10.35940/ijitee.J1127.0881019.

Baldez, M. B., de Almeida Teixeira, M., and Junior, J. A. M. (2022). A importância da manutenção preventiva com foco na redução da manutenção corretiva em equipamentos médicos do Hospital Universitário de Vassouras/RJ. Revista Eletrônica TECCEN, 15(1), 59-65.

Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Branco Filho, G. (2008). A organização, o planejamento e o controle da manutenção. Rio de Janeiro: Ciência Moderna.

Chimuco, F. T. (2015). Modelos de regressão linear: fatores que influenciam a esperança de vida em Angola (Doctoral dissertation).

Dhillon, B. S. (2002). Engineering maintenance: a modern approach. Boca Raton: CRC PRESS.

Furletti, L. A., De Carvalho, Í. L., Da Silva, H. B., e de Castro Peixoto, L. (2022). Aplicação de regressão linear múltipla para analisar a relação entre buscas por palavras-chave na internet e casos de COVID-19 no Brasil. Exacta.

Gamboa, S.S. Pesquisa em educação: métodos e epistemologias. 2. ed. Chapecó, SC: Argos, 2012. https://doi.org/10.18675/2177-580X.vol2.n1.p09-32

Işikdag, Ümit et al. Estimating Construction Material Indices with ARIMA and Optimized NARNETs. Computers, Materials and Continua, v. 74, n. 1, 2023.

Kardec, A. e Nascif, J. (2009). Manutenção: função estratégica. (3a. ed.). Rio de Janeiro: Qualitymark.

Lin, S., Fu, Y., Jia, X., Ding, S., Wu, Y., & Huang, Z. (2020). Discovering Correlations between the COVID-19 Epidemic Spread and Climate. International Journal of Environmental Research and Public Health, 17(21), 7958.

Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. 3rd ed. New York: Wiley; 1998.

McLinden, C.A., Fioletov, V., Shephard, M.W., Krotkov, N., Li, C., Martin, R.V., Moran, M.D., Joiner, J., 2016. Space-based detection of missing sulfur dioxide sources of global air pollution. Nat. Geosci. 9 (7), 496–500. https://doi.org/10.1038/ngeo2724.

Morettin, P., A.; TOLOI, Clélia MC. Análise de séries temporais: modelos lineares univariados. Editora Blucher, 2018.

Oliveira, L. De, Tonin, P. C., e Vicenzi, S. L. (2020). Comportamento dos custos totais de produção no segmento da avicultura de postura no estado do Paraná: estudo baseado na análise de regressão linear múltipla. Revista Produção Online, 20(1), 28-46.

Pourmohammad, P., Jahani, A., Chahooki, M.A., Meigooni, H., 2020. Road impact assessment modelling on plants diversity in national parks using regression analysis in comparison with artificial intelligence. Model. Earth Syst. Environ. 6, 1281–1292. https://doi.org/10.1007/s40808-020-00799-6.

Rabajante, J. F. (2020). Insights from early mathematical models of 2019-nCoV acute respiratory disease (COVID-19) dynamics. arXiv preprint arXiv:2002.05296.

SILVA, Heric Ramos. Análise de dados de vida aplicada a um componente aeronáutico na otimização do seu intervalo de manutenção. 2016. 41. Monografia Especialização em Engenharia da Confiabilidade - Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

Xenos, H. G. Gerenciando a manutenção produtiva. (2004). Nova Lima: INDG Tecnologia e Serviços Ltda.

Yorifuji, T., Kashima, S., Suryadhi, M.A., Abudureyimu, K., 2019. Acute exposure to sulfur dioxide and mortality: historical data from Yokkaichi, Japan. Arch. Environ. Occup. Health 74 (5), 271–278. https://doi.org/10.1080/19338244.2018.1434474

DOI: 10.59254/sbpo-2024-193513
https://proceedings.science/p/193513?lang=pt-br