



ConBRepro

XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



IA nas Engenharias

29 nov. a 01
de dezembro 2023

Análise preditiva de vendas em uma indústria de alimentos com uso do Software R: Modelos de Regressão Linear Simples, Suavização Exponencial Dupla e ARIMA

Eduardo Westphal

Programa de Pós-Graduação em Engenharia de Produção – UFPR

Mizael Vieira de Souza

Programa de Pós-Graduação em Engenharia de Produção – UFPR

Mariana Kleina

Programa de Pós-Graduação em Engenharia de Produção – UFPR

Agnelo Denis Vieira

Programa de Pós-Graduação em Engenharia de Produção – UFPR

Izabel Cristina Zattar

Programa de Pós-Graduação em Engenharia de Produção – UFPR

Resumo: A imprevisibilidade é um fator de risco dentro das empresas, pois impossibilita o dimensionamento adequado de recursos que serão necessários para cumprir demandas previamente estabelecidas. A utilização de softwares de código aberto é uma alternativa cada vez mais usual na análise de dados, uma vez que é possível a visualização e interatividade de dados, que muitas vezes não são possíveis por meio de outras formas de análises que não tenham uma abordagem técnica estruturada. As análises apresentadas são realizadas no RStudio, uma ferramenta do *software* R, onde a apresentação dos dados é feita de maneira mais ordenada e intuitiva. A programação vem sendo utilizada cada vez mais para que dados possam ser analisados de maneira mais simples e direta e promova uma tomada de decisão mais rápida e assertiva no dia a dia das empresas. A análise apresentada é um conjunto de resultados de análises da aplicação de três métodos de previsão de demanda, aplicada sobre uma série de dados de uma empresa do ramo de alimentos de grande porte, que atualmente não utiliza nenhuma técnica para projeção orçamentária, e foi limitada a apenas um segmento de atuação da empresa. Os métodos utilizados realizaram o ajuste da série e em base ao método que retornou o menor erro, foram projetadas as vendas mensais do ano subsequente.

Palavras-chave: Previsão de vendas, *Software* R, Análise de Dados, Métodos quantitativos.

Predictive analysis of sales in a food industry using R Software: Simple Linear Regression, Double Exponential Smoothing and ARIMA Models

Abstract: Unpredictability is a risk factor within companies as it hinders the adequate sizing of resources that will be needed to meet a predetermined demand. The use of open-source software is an increasingly common alternative in data analysis, as it allows for data visualization and interactivity, often not achievable through other non-technically structured analytical approaches. The analyses presented are performed in RStudio, a tool of the R software, where data presentation is carried out in a more orderly and intuitive manner. Programming is being increasingly employed to simplify and streamline data analysis, enabling faster and more accurate decision-making in day-to-day business operations. The analysis presented comprises a set of results from the application of three demand forecasting methods on a large-scale food industry's dataset. This particular company currently does not employ any budget projection techniques and the analysis was limited to a specific business segment. The methods used corrected the dataset, and based on the method returning the lowest error, monthly sales for the subsequent year were projected.

Keywords: Sales Forecasting, R Software, Data Analysis, Quantitative methods.

1. Introdução

A importância da previsão no faturamento e nas vendas não pode ser subestimada. É fundamental entender por que é crucial realizar previsões de vendas. Afinal, isso vai muito além de evitar déficits ou excessos de mercadoria. No cenário altamente competitivo do mercado, uma empresa precisa contar com uma previsão de demanda precisa e assertiva para se manter confiável e bem-sucedida. Será explorada detalhadamente como essa previsão desempenha um papel vital em todos os aspectos do negócio, desde o planejamento estratégico até a gestão eficiente de recursos, garantindo assim sua competitividade e confiabilidade a longo prazo.

Segundo Krajewski *et al.* (2012), a previsão representa uma antecipação de eventos vindouros, cujo propósito é o planejamento e a orientação na tomada de decisões, com o intuito de ampliar as chances de êxito ou assegurar resultados favoráveis. De acordo com Battisti (2008), a aplicação dos métodos estatísticos para previsão de demandas pode contribuir para otimizar a eficiência dos procedimentos cotidianos.

Nesse cenário, Sadaei *et al.* (2014) destacam a relevância da análise de séries temporais, fundamentada predominantemente nas observações passadas dispostas sequencialmente ao longo de intervalos temporais específicos. Isso a torna uma abordagem quantitativa para a projeção de valores futuros de uma variável.

Por fim, Box *et al.* (2015) indicam que, para compreender as relações entre os dados analisados em séries temporais, pode-se recorrer à construção de modelos econométricos, eliminando a necessidade de uma teoria subjacente específica para sua interpretação.

Essas interpretações enfatizam a relevância da estatística e dos métodos estatísticos em diversas aplicações, indo desde a tomada de decisões até a previsão de eventos futuros, demonstrando como a estatística se apresenta como uma ferramenta instrumental para a gestão e análise de dados em contextos práticos.

Para atingir esse objetivo, técnicas como a Regressão linear simples, Suavização exponencial dupla (Holt) e o Modelo ARIMA, são frequentemente utilizadas no processo de previsão de demanda.

2. Referencial teórico

2.1 Regressão linear simples

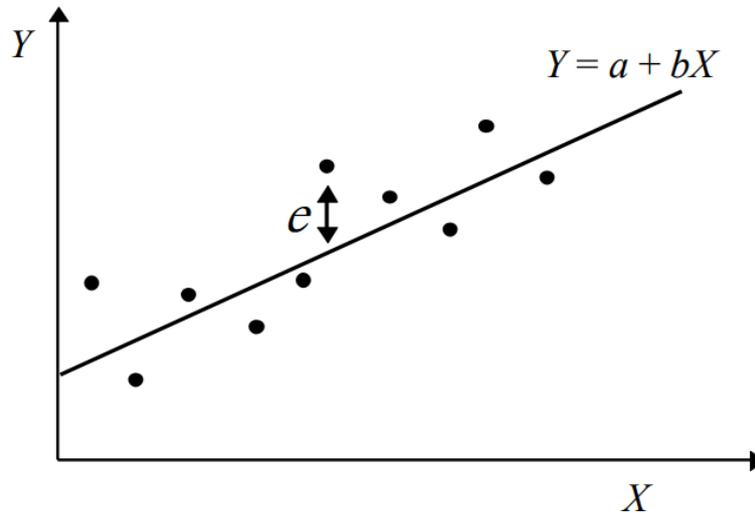
Segundo Krajewski, Ritzman e Malhotra (2009), a regressão linear simples é uma técnica de modelagem estatística usada para entender e prever a relação entre duas variáveis: uma variável dependente (geralmente representada por Y) e uma variável independente (geralmente representada por X). O objetivo principal é encontrar uma equação linear que

descreva a relação entre essas duas variáveis. Segundo Correa (2003) a equação linear é dada pela Equação 1.

$$Y = a + bX \quad (1)$$

Onde Y representa a variável dependente que deseja prever ou explicar, X representa a variável independente que é usada para prever Y , a é o intercepto da linha de regressão, que representa o valor de Y quando X é igual a zero e o b , o coeficiente de inclinação da linha de regressão, que representa a mudança em Y para uma unidade de mudança em X . Na Figura 1, observa-se a formação da reta no sistema cartesiano.

Figura 1 – Representação da equação da reta encontrada por Mínimos Quadrados



Fonte: Tubino (2007)

Gaither e Frazier (2006) descrevem as Equações (2) e (3) para se encontrar os valores de a e b . O cálculo dos coeficientes a e b têm como finalidade minimizar a soma dos desvios quadrados dos dados reais da linha do gráfico.

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2} \quad (2)$$

$$a = \frac{\Sigma Y - b(\Sigma x)}{n} \quad (3)$$

Onde Y representa a variável dependente, a o intercepto no eixo da variável independente Y , b o coeficiente angular, X a variável independente e n o número de períodos observados.

2.2 Suavização exponencial dupla (Holt)

O método de suavização exponencial dupla, também chamado de método de Holt, é indicado para séries com tendência linear, porém sem sazonalidade (CLIFF; RAGSDALE, 2012). A função de previsão do método de Holt é representada pelas Equações (4), (5) e (6).

$$P_{t+h} = Lt + hT_t \quad (4)$$

$$L_t = \alpha R_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (6)$$

Onde L_t é o componente de nível no período t , T_t é o componente de tendência no período t , o h é o número de saltos entre períodos considerados (horizonte de previsão), em que $h = 1, 2, 3$ e assim por diante. A constante de suavização do componente de nível L_t é o α , e o β é a constante de suavização do componente de tendência T_t tanto o α quanto o β estão no intervalo $[0, 1]$. As equações (5) e (6) são responsáveis pela estimativa de nível e inclinação da série temporal, já a equação (4) calcula a previsão de demanda nos próximos períodos.

Conforme descrito por Fernandes e Godinho Filho (2010), o método atribui um peso maior aos dados mais recentes. Isso implica que o método é mais sensível às mudanças na tendência dos dados, uma vez que reage de forma mais rápida e adaptativa às observações mais recentes. Essa característica é fundamental para lidar com séries temporais que apresentam tendências em evolução, permitindo que o modelo se ajuste prontamente às alterações nas condições subjacentes.

2.3 Modelo ARIMA

O modelo ARIMA (*Autoregressive Integrated Moving Average*) é um modelo estatístico amplamente utilizado para a previsão de séries temporais. Ele é especialmente útil para séries temporais que exibem padrões de autocorrelação, tendências e sazonalidades. O modelo baseia-se no ajuste dos valores observados, que visa reduzir a lacuna entre os valores gerados e os valores observados para um nível próximo de zero (BOX; JENKINS, 2015; SATO, 2013).

Segundo Hyndman e Athanasopoulos (2018), os parâmetros do modelo ARIMA são AR (ordem autorregressiva), I (ordem de diferenciação) e MA (ordem de média móvel). Esses parâmetros são selecionados com base na análise dos dados da série temporal, como a autocorrelação e a autocorrelação parcial. O objetivo é encontrar o conjunto de parâmetros que melhor descreve a estrutura subjacente da série e permite fazer previsões precisas.

$AR(p)$: ordem autorregressiva, também conhecido como ordem de atraso, onde p representa o número de observações de atraso incluídas no modelo. A parte autorregressiva do modelo ARIMA é expressa pela Equação (7).

$$R_t = c + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \dots + \phi_p R_{t-p} + e_t \quad (7)$$

Onde R_t é o valor da série temporal no tempo t , c é uma constante, e_t é o erro no tempo t e $\phi_1, \phi_2, \dots, \phi_p$ são os coeficientes autorregressivos.

$I(d)$: ordem de diferenciação, onde d representa o número de vezes que as observações brutas são diferenciadas. É dada pela Equação (8).

$$R'_t = R_t - R_{t-d} \quad (8)$$

Onde R'_t é a série diferenciada e d é a ordem de diferenciação.

$MA(q)$: ordem de média móvel, onde q representa o tamanho da janela da média móvel. A parte de média móvel do modelo ARIMA é dada pela Equação (9).

$$R_t = \mu + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \quad (9)$$

Onde R_t é o valor da série temporal no tempo t , μ é uma constante, e_t é o erro no tempo t e $\theta_1, \theta_2, \dots, \theta_p$ são os coeficientes de média móvel.

2.4 Testes de tendência e estacionariedade para séries temporais

Para saber se a série temporal possui tendência, utiliza-se um teste de hipótese chamado de teste Mann-Kendall proposto por Mann (1945) e Kendall (1975) que é um teste não-paramétrico utilizado para determinar se uma série de dados possui tendência temporal estatisticamente significativa. Pelo fato de ser um método não paramétrico, não há necessidade de que os dados apresentem distribuição normal (YUE *et al.*, 2002).

A hipótese nula H_0 considerada pelo teste é que não existe tendência na série analisada, enquanto a hipótese alternativa H_1 afirma que existe tendência, seja ela positiva ou negativa. Para uma série ter tendência é preciso que o p-valor < valor crítico (ex. 5%), assim rejeita-se H_0 (MANN, 1945).

Para saber se a serie temporal possui estacionariedade, pode-se utilizar testes estatísticos (de hipótese) que indicam se uma série é estacionária. O primeiro apresentado é o teste *Augmented Dickey-Fuller* (ADF). A hipótese nula H_0 indica que a série não é estacionária, já a hipótese alternativa H_1 afirma que a série é estacionária. Assim, para uma série ser estacionária é preciso que o p-valor < valor crítico (ex. 5%), assim rejeita-se H_0 (DICKEY, FULLER, 1979).

O segundo teste que pode ser utilizado para verificar estacionariedade é o teste *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS), porém diferente do ADF, nesse caso a hipótese nula H_0 indica que a série é estacionária, já a hipótese alternativa H_1 afirma que a série não é estacionária. Assim, para uma série ser estacionária pelo teste KPSS, é preciso que o p-valor > valor crítico (ex. 5%), assim aceita-se H_0 (KWIATKOWSKI *et al.*, 1992).

2.5 Medida de erro

Raiz do erro médio quadrático (REMQ) dada pela Equação 10 é uma medida estatística comumente usada para avaliar a precisão de um modelo de previsão ou regressão em comparação com os valores observados. Essa métrica é especialmente útil quando você deseja medir o quão próximo a previsão de um modelo está do valor real.

$$REMQ = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} \quad (10)$$

Onde n é o número de períodos de previsão e e_t representa o erro para a observação ou ponto de dados t . Geralmente, é calculado pela diferença entre o valor real R_t e o valor previsto P_t para o mesmo ponto de dados. A Equação 11 apresenta essa diferença no tempo t .

$$e_t = R_t - P_t \quad (11)$$

3. Metodologia

A base de dados utilizada foi obtida com o consentimento da empresa e os dados reais foram analisados considerando um fator de conversão, medida para promover uma privacidade na divulgação de dados considerados estratégicos pela empresa. Os dados apresentados referem-se a janeiro de 2021 até dezembro de 2023, sendo que valores considerados nos meses de setembro a dezembro do ano de 2023 (meses futuros a escrita desse artigo) são valores firmes, com pedidos já em carteira, pois a empresa trabalha em um sistema contratual de abastecimento e como são produtos fabricados em unidades no exterior, os pedidos são colocados com antecedência e por isso a previsibilidade destes valores como reais. Esta base foi analisada utilizando o *software* R (BATTISTI e SMOLSKI, 2019). A instalação de algumas bibliotecas foram requeridas para possibilitar iniciar com os testes de estacionariedade, com a biblioteca *tseries* e a função *kpss.test()*, e o teste de tendência, com a biblioteca *trend* e a função *mk.test()*. Para a análise de sazonalidade, realizou-se uma observação direta na plotagem dos dados por meio do gráfico da série temporal.

Para a aplicação do método de regressão linear simples, foi implementado código próprio no *software* R. Para realizar a análise de suavização exponencial dupla, foi necessário a instalação da biblioteca *forecast* e utilizada a função *holt()*, onde o próprio *software* encontra os valores ideais das duas constantes do método. Em relação a análise ARIMA, ela também requer a instalação da biblioteca *forecast* e utilizada a função *auto.arima()*, função esta que faz a otimização automática dos parâmetros no modelo. Os erros de cada modelo foram analisados utilizando a Equação (10) que apresenta a raiz do erro médio quadrático.

Após a identificação do modelo que retornou o menor erro, o método foi aplicado para projeção de vendas do ano subsequente, proporcionando uma projeção com base técnica e uma maior assertividade na projeção.

4. Resultados

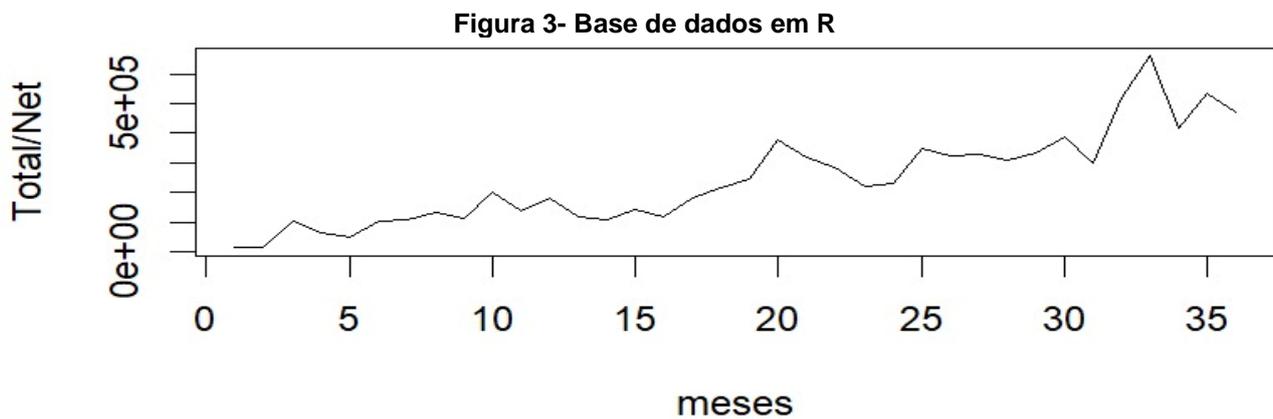
Na Figura 2 são apresentados os 36 dados referente a série temporal utilizada, onde a imagem é oriunda da captura de tela do *software* R.

Figura 2- Base de dados em R

[01]	013263.0	016454.0	101849.0	063564.0	047194.0	103566.0	108300.0	132810.0
[09]	114922.0	198925.0	136291.0	180302.0	117608.0	107963.0	144249.0	115644.0
[17]	183099.0	216788.0	244679.0	378678.0	318592.0	284780.0	222337.0	228525.0
[25]	348298.0	325001.0	327517.0	309720.0	332168.0	386436.0	298141.0	520469.0
[33]	662408.5	418223.8	536307.7	470331.0				

Fonte: Os autores (2023)

A Figura 3 apresenta o gráfico da série temporal utilizada no estudo, onde a variável *Total/Net* representa a série temporal referente ao faturamento realizado.



Após a plotagem da série original de dados, foi analisado por meio da observação do comportamento da série, que não há um período sazonal definido. Com a aplicação do teste de estacionariedade KPSS (Kwiatkowski-Phillips-Schmidt-Shin) foi possível identificar que se trata de uma série não estacionária, onde foi identificado um p-valor $<0,05$, conforme Figura 4.

Figura 4- Teste de estacionariedade KPSS

```

KPSS Test for Level Stationarity
data: R
KPSS Level = 0.92524, Truncation lag parameter = 3, p-value = 0.01

```

Fonte: Os autores (2023)

Após o teste de estacionariedade, aonde retornou um p-valor <0.05 , ou seja, confirma que é uma serie não estacionária, foi aplicado um teste para verificar outra premissa da série de dados, que é a característica de tendência. Para tal teste, foi utilizado o teste de hipótese de Mann-Kendall e o resultado foi que a série possui o fator tendência, já que o resultado do p-valor foi $<0,05$, conforme Figura 5.

Figura 5- Teste de tendência de Mann-Kendall

```

Mann-Kendall trend test
data: R
z = 6.8513, n = 36, p-value = 7.318e-12
alternative hypothesis: true S is not equal to 0
sample estimates:
  S      varS      tau
504.0 5390.0    0.8

```

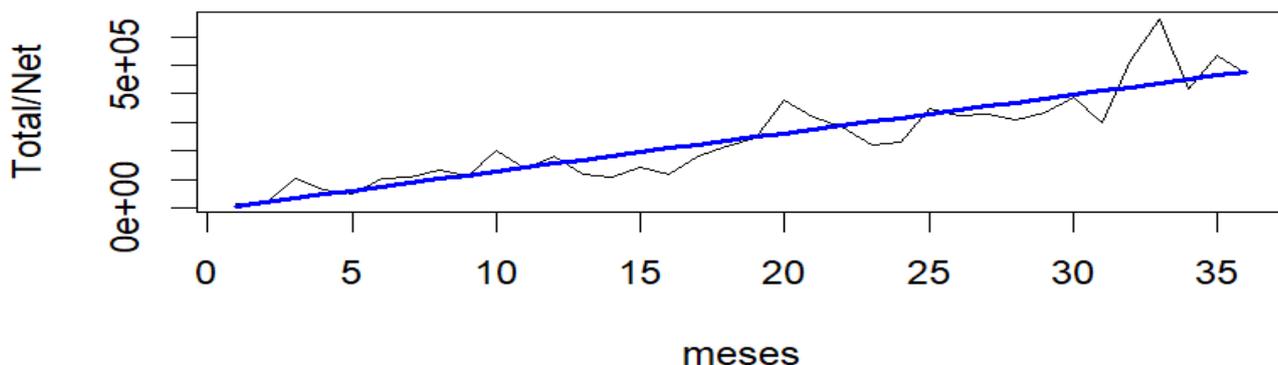
Fonte: Os autores (2023)

A aplicação destes testes, se fez necessária para definição dos métodos adequados que consideram os fatores de tendência e estacionariedade como premissas básicas do modelo matemático e utilizam fatores para corrigi-los, gerando assim previsões com o menor erro possível.

Foi realizada a aplicação de cada um dos métodos, primeiramente apresenta-se a Figura 6 considerando o método de regressão linear simples (RSL). O método tem bom desempenho para séries temporais que apresentam comportamento linear, podendo ter tendência de crescimento ou decrescimento, e foram identificadas estas características na

série temporal em estudo. O REMQ para este método no período de 36 meses foi de 64689,1.

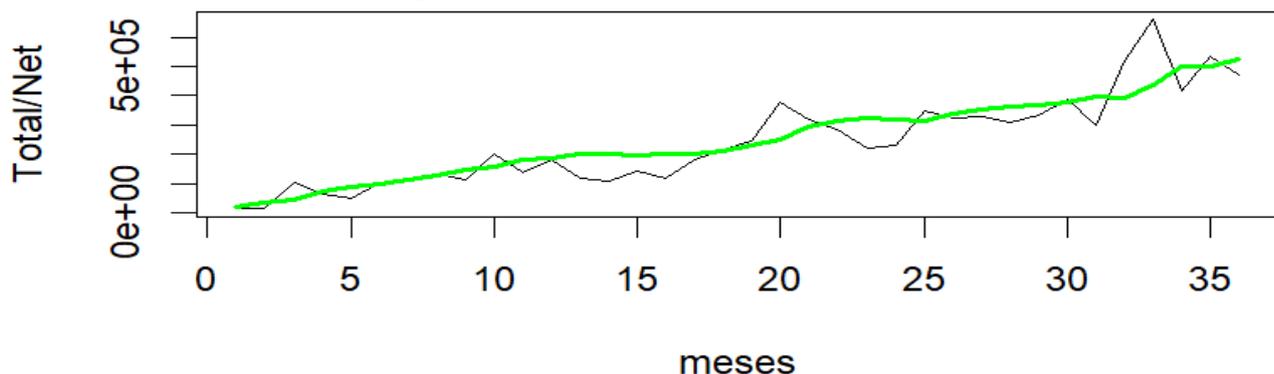
Figura 6- Regressão linear simples



Fonte: Os autores (2023)

O segundo método utilizado foi a suavização exponencial dupla, conhecido também como método de Holt. Foi um método escolhido devido sua indicação de aplicação em séries com tendência linear, porém sem sazonalidade. Os parâmetros ajustados pelo *software* R foram $\alpha = 2.164112e - 01$ e $\beta = 1.000143e - 04$. A Figura 7 apresenta o resultado do ajuste realizado. Para esse método o REMQ foi de 68571,79.

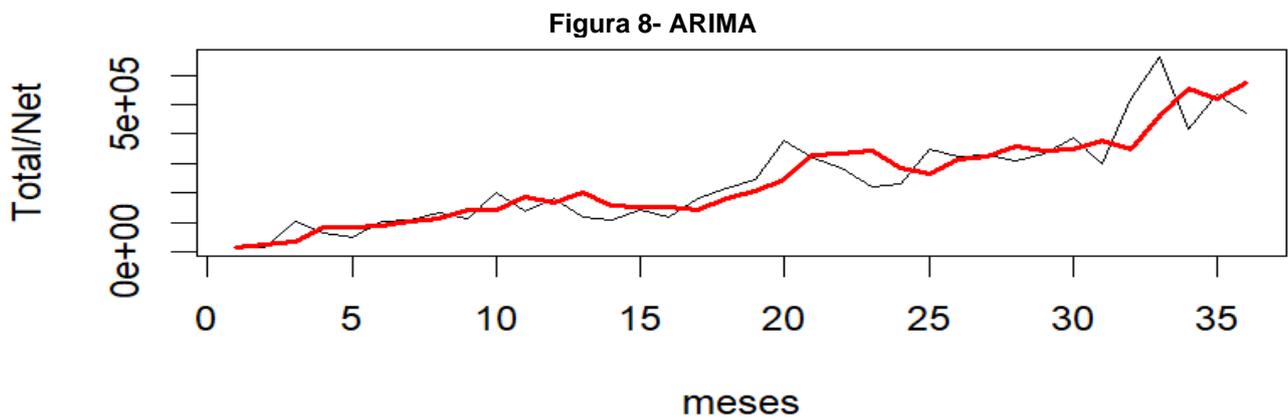
Figura 7- Suavização exponencial dupla



Fonte: Os autores (2023)

O terceiro método aplicado foi o ARIMA apresentado na Figura 8. É um método bastante versátil e simples, usa poucos parâmetros para ajuste dos modelos, permite testar hipóteses sobre os parâmetros, trata autocorrelação de forma explícita e ajusta o modelo por etapas. São três etapas em suma, uma que define a ordem do modelo autoregressivo, uma segunda que define se a série deve ser diferenciada, para que se torne estacionária, e uma terceira, que define a ordem do modelo de médias móveis. O REMQ quando aplicado o método no período amostral foi de 71143,54.

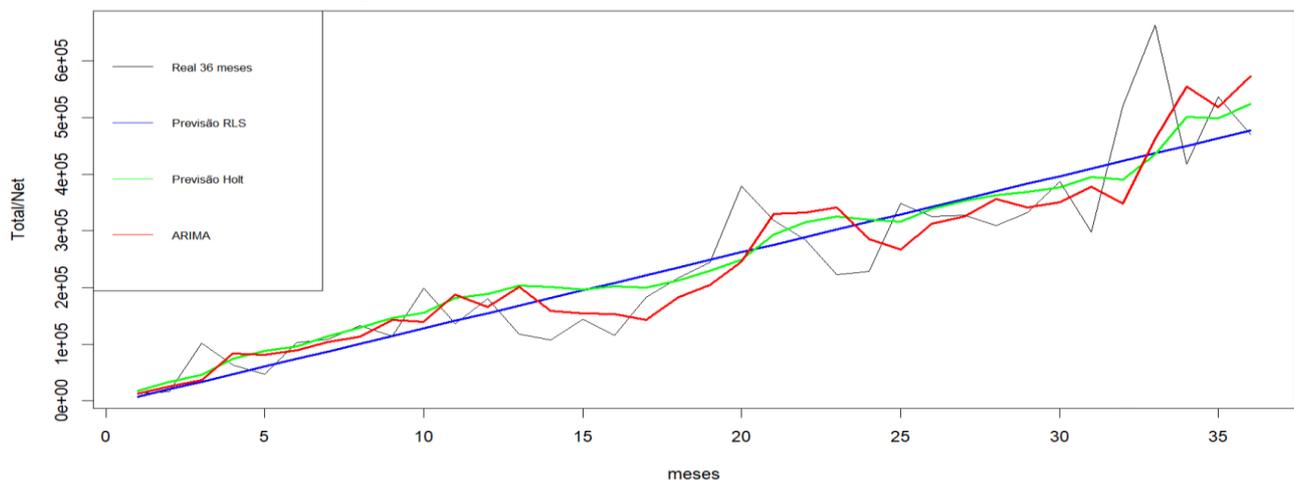
Os parâmetros ajustados foram: $p = 2$, $d = 1$ e $q = 0$. O parâmetro $p = 2$ significa que foram utilizados dois dados passados da série temporal para a parte autoregressiva, resultando nos coeficientes $\phi_1 = -0,4806$ e $\phi_2 = 0,2867$. O parâmetro $d = 1$ significa que a série foi diferenciada uma vez para se tornar estacionária, corroborando com o teste de estacionariedade realizado anteriormente. E o parâmetro $q = 0$ significa que não foram utilizados valores passados de erros de previsão para a parte de médias móveis.



Fonte: Os autores (2023)

A Figura 9 apresenta graficamente um comparativo entre os três métodos aplicados à série de dados.

Figura 9- Comparativo entre os modelos aplicados



Fonte: Os autores (2023)

Após a aplicação de todos os métodos e análise comparativa do REMQ, foi proporcionado a empresa um método para previsão do ano subsequente. O método com menor erro e consequentemente o que mais se adapta as características da série de dados é o método de regressão linear simples (RSL), conforme pode-se notar na Tabela 1.

Tabela 1- Resultado dos REMQ para cada método aplicado

Método	REMQ
Regressão Linear simples	64689,10
Suavização Exponencial Dupla	68571,79
ARIMA	71143,54

Fonte: Os autores (2023)

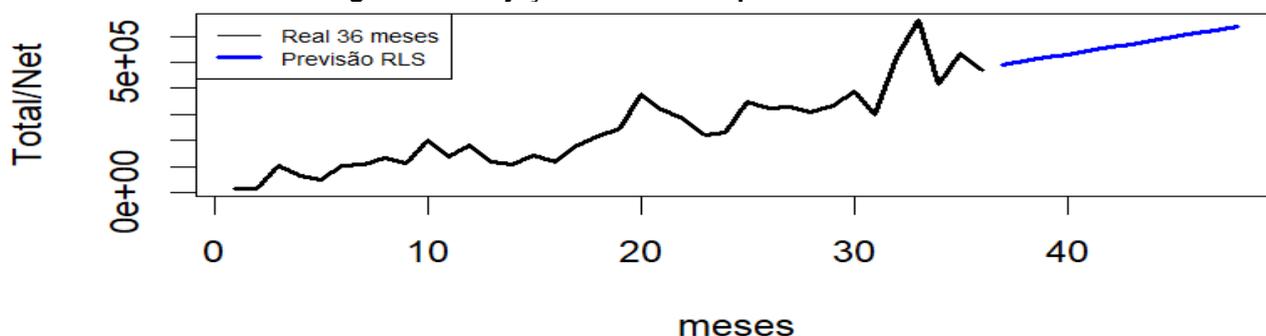
Portanto, realizando a projeção para o ano de 2024, os resultados obtidos estão descritos na Figura 10 e apresentados graficamente na Figura 11.

Figura 10- Valores projetados para o ano de 2024

[1]	490469.4	503895.1	517320.8	530746.5	544172.1	557597.8
[7]	571023.5	584449.2	597874.8	611300.5	624726.2	638151.8

Fonte: Os autores (2023)

Figura 11- Projeção dos valores para o ano de 2024



Fonte: Os autores (2023)

Com isso, a empresa a partir de agora terá uma nova ferramenta, que proporciona de maneira eficiente, acessível do ponto de vista monetário e técnico, que poderá servir como baliza para tomada de decisões nos diversos níveis gerenciais e operacionais.

5. Conclusões

Utilizando a programação em R, foi possível analisar uma base dados externa e tirar conclusões com base nessas análises, e o mais importante, realizar uma projeção dos valores de faturamento, para os próximos períodos.

O *software* R, se mostra uma ferramenta de fácil obtenção por ser de código aberto e os comandos utilizados para realizar a análise de dados são simples e intuitivos, com isso é possível analisar vários dados de uma única vez e muitas vezes com apenas um comando.

Com isso, pode-se dizer que o trabalho concluiu seu objetivo e foi possível tirar conclusões sobre as vendas futuras da empresa objeto de estudo, e com a utilização de métodos de previsão de demanda, definidos em base a alguns testes de hipóteses, foi possível identificar o método que retornava o menor erro, e assim realizar uma previsão mais assertiva e metodológica do valor de faturamento da empresa para os próximos 12 meses. Para futuras pesquisas, sugere-se uma análise mais aprofundada por categorias, visando compreender se o comportamento de cada segmento é congruente com o método escolhido para a projeção global de vendas.

Referências

BATTISTI, I. D. E.; SMOLSKI, F. M. S. **Software R: Análise estatística de dados utilizando um programa livre 1a**. Edição BAGÉ EDITORA FAITH 2019.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.

CLIFF, T.; RAGSDALE. **Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science**, 6 ed. 2012.

DICKEY, D.; FULLER, W. **Distribution of the Estimator for Autoregressive Time Series With a Unit Root**. Journal of the American Statistical Association, v. 74, p.427-431, 1979.

CORREA, S. M. **Probabilidade e estatística**, 2ª edição. - Belo Horizonte: PUC Minas Virtual, 2003.

GAITHER, N; FRAZIER, G. **Administração da produção e operações**. 8. ed. São Paulo: Thomson Learning, 2006.

FERNANDES, F. C. F.; GODINHO FILHO, M. **Planejamento e Controle da Produção. Dos Fundamentos ao Essencial**. São Paulo: Atlas S.A, 2010.

HYNDMAN, R.J. ; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Libgen Librarian, 2018.

KRAJEWSKI, L. J.; RITZMAN, L. P.; MALHOTRA, M. **Administração da produção e operações**. 8. ed. São Paulo: Pearson Prentice Hall, 2009.

KENDALL, M. G. R. **Correlation Methods**. 4a edição. Charles Griffin, Londres, 1975.

KWIATKOWSKI, D.; PHILLIPS, P. C.; SCHMIDT, P.; SHIN, Y. **Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?**. Journal of econometrics, v. 54, n. 1-3, p. 159-178, 1992.

MANN, H. B. **Non-parametric tests against trend**. Econometrica: Journal of the econometric society, p. 245-259, 1945.

SADAEI, H. J.; ENAYATIFAR, R.; ABDULLAH, A. H.; GANI, A. **Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search**, International Journal of Electrical Power & Energy Systems 62: 118–129. 2014.

SATO, R. C. **Gerenciamento de doenças utilizando séries temporais com o modelo ARIMA**. Einstein, v. 11, n. 1, p. 128-131, 2013.

YUE, S.; PILON, P.; CAVADIAS, G. **Power of the Mann-Kendall and spearman's rho tests for detecting monotonic trends in hydrological series**. Journal of Hydrology, Amsterdam, v. 259, n. 1-4, p. 254-271, 2002.

ANEXO

Códigos em R utilizados no trabalho.

```
# LEITURA DO DADOS
R<-read.csv("TotalNet.csv")
```

```
# TESTES DE ESTACIONARIEDADE E TENDÊNCIA, RESPECTIVAMENTE
library("tseries")
kpss.test(R)
```

```

library("trend")
mk.test(R)

# RAIZ DO ERRO MÉDIO QUADRÁTICO
REMQ<-function(real,previsto){
  erro<-(real-previsto)^2
  calculo<-sqrt(mean(erro,na.rm=TRUE))
  return(calculo)
}

# CÓDIGOS DE REGRESSÃO LINEAR SIMPLES
Y<-R
n<-length(R)
X<-1:n
b<-(n*sum(X*Y)-sum(X)*sum(Y))/(n*sum(X^2)-sum(X)^2)
a<-(sum(Y)-b*sum(X))/n
rl<-function(X){return(a+b*X)}
PrevRL<-rl(X)
REMQ(R,PrevRL)
plot(X,R,type="l",xlab="meses",ylab="Total/Net",lwd=2)
lines(X,PrevRL,lwd=2,col="blue")

# CÓDIGOS DE SUAVIZAÇÃO EXPONENCIAL DUPLA
library("forecast")
PrevHOLT<-holt(y=R,h=1,initial = "optimal")$fitted
REMQ(R,PrevHOLT)
plot(X,R,type="l",xlab="meses",ylab="Total/Net",lwd=2)
lines(X,PrevHOLT,col="green",lwd=2)

# CÓDIGOS DO ARIMA
library("forecast")
PrevARIMA<-auto.arima(R)$fitted # otimização automática dos parâmetros
REMQ(R,PrevARIMA)
plot(X,R,type="l",xlab="meses",ylab=" Total/Net",lwd=2)
lines(X,PrevARIMA,col="red",lwd=2)

```