



## Box Jenkins Method Applied to forecast the Federal Savings Bank monthly demand of Clients

### Uso da metodologia box-Jenkins para previsão da demanda mensal de clientes em uma agência da Caixa Econômica Federal

SILVA, Jonas Ferreira da<sup>(1)</sup>; NEPOMUCENO, Thyago Celso Cavalcante<sup>(2)</sup>; RODRIGUES, Naialy Patricia<sup>(3)</sup>

<sup>(1)</sup> 0000-0002-3009-8876; Universidade Federal de Pernambuco. Caruaru, Pernambuco (PE), Brasil. [jonas.ferreirasilva@ufpe.br](mailto:jonas.ferreirasilva@ufpe.br)

<sup>(2)</sup> 0000-0001-8327-6472; Universidade Federal de Pernambuco. Caruaru, Pernambuco (PE), Brasil. [thyago.nepomuceno@ufpe.br](mailto:thyago.nepomuceno@ufpe.br)

<sup>(3)</sup> 0000-0002-5954-9458; Universidade Federal de Pernambuco. Caruaru, Pernambuco (PE), Brasil. [naialy.rodrigues@ufpe.br](mailto:naialy.rodrigues@ufpe.br)

O conteúdo expresso neste artigo é de inteira responsabilidade dos/as seus/as autores/as.

#### ABSTRACT

Queue management in the banking service is a current and recurring challenge that service units face daily in modern times, and especially in Caixa Econômica Federal branches, due to their profile and social mission, this problem is quite relevant as it affects the quality of lives of customers and employees, leading to effects at scale where simple demands can result in hours of waiting, a situation even more critical considering the current pandemic and post-pandemic context. The present work aims to estimate the service demand in a typical branch of Caixa Econômica Federal (CEF) for strategic planning and distribution of resources over time and sectors of the bank. Data were collected on the times and quantities of service in the years 2019 and 2020. After collection, the data were processed and the Box - Jenkins methodology was applied to find the best forecast models for the data reported by the establishment. It was found that by using Box-Jenkins models, varying their parameters, it is possible to obtain important results when it comes to demand forecasting. The method has proven to be very effective, even with the limited amount of data.

#### RESUMO

A gestão de filas no serviço bancário é um desafio atual e recorrente que unidades de atendimento enfrentam diariamente em tempos modernos, e em especial nas agências da Caixa Econômica Federais, devido seu perfil e missão social, tal problemática é bastante relevante ao afetar a qualidade de vida de clientes e empregados, levando a efeitos em escala onde demandas simples podem resultar em horas de espera, situação ainda mais crítica levando em consideração o contexto pandêmico e pós-pandêmico atual. O presente trabalho objetiva estimar a demanda de atendimento em uma agência típica da Caixa Econômica Federal (CEF) para o planejamento estratégico e distribuição de recursos ao longo do tempo esesores do banco. Foram levantados dados sobre os tempos e quantidades de atendimento nos anos de 2019 e 2020. Após a coleta, os dados foram tratados e foi aplicado a metodologia box Jenkins para encontrar os melhores modelos de previsão para os dados informados pelo estabelecimento. Constatou-se que a partir da utilização de modelos Box-Jenkins, variando seus parâmetros, é possível encontrar resultados importantes, quando se trata de previsões de demanda. O método se mostrou muito eficaz, mesmo com a quantidade limitada de dados.

#### INFORMAÇÕES DO ARTIGO

##### Histórico do Artigo:

Submetido: 09/11/2022

Aprovado: 02/12/2022

Publicação: 01/12/2023



##### Keywords:

Data analysis, Times series econometrics, ARIMA, Prediction Models, Banking, Brazilian Federal Savings Bank

##### Palavras-Chave:

Análise de dados, Econometria de séries temporais, ARIMA, Modelos de previsão, Banco, Caixa Econômica Federal

## **Introdução**

Filas estão sempre presentes na sociedade, seja para conseguir ingresso em um cinema como para resolver problemas em unidades de serviço como bancos e repartições públicas. A caixa Econômica Federal, também conhecida como Caixa ou CEF, é uma instituição financeira brasileira sob a forma de autarquia de direito privado, com patrimônio próprio e autonomia administrativa, com importância de destaque na formação, consolidação e no desenvolvimento do Estado brasileiro.

Os serviços oferecidos pela instituição sempre estiveram alinhados com os objetivos de sustentabilidade, inclusão social e suporte a populações em situação de vulnerabilidade social, com destaques para o auxílio Brasil, PIS, FGTS dentre outros programas sociais que fazem parte do cotidiano da classe trabalhadora, além do monopólio das operações de penhor e gestão das loterias federais, uma das principais fontes de fomento do governo federal.

Desta forma, temas como gestão da capacidade produtiva são de grande relevância, pois, o planejamento das operações de acordo com o processo ou conceito do serviço, pode garantir que as metas de qualidade estabelecidas sejam atendidas de maneira que o cliente receba o que está esperando.

Nos últimos anos as agências da Caixa Econômica tiveram aumento significativo no número de filas em todos os estados do Brasil para o saque do FGTS e outros programas sociais., impossibilitando que o lead time, isto é, o tempo entre a chegada e a saída do estabelecimento bancário seja pequeno pelo fato de que na maioria das vezes a instituição não está pronta para suprir a demanda. Dessa forma, o quantitativo de pessoas a espera por atendimento se torna uma informação estratégica relevante para instituição, possibilitando soluções ágeis para que não haja grandes aglomerações em filas.

Diante disso, o objetivo desta proposta é construir um modelo de previsão baseado em séries temporais capaz de apresentar a quantidade mensal de demanda por determinados tipos de atendimento nas agências da Caixa Econômica Federal, permitindo encontrar maneiras de atender à demanda de forma mais eficiente. Essa abordagem tem sido amplamente utilizada em estudos sobre previsão de demanda em diversos setores, como o estudo de Chalvatzis et al. (2016), que aplicou a metodologia Box-Jenkins para prever a demanda de eletricidade na Grécia, obtendo resultados promissores [5]. Outros estudos, como o de Ouyang e Zhang (2019), também utilizaram a metodologia Box-Jenkins para prever a demanda de energia elétrica em uma região da China, destacando sua eficácia na previsão de demanda de curto prazo [6]. Além disso, o estudo de Sánchez-Torres et al. (2018) propôs um modelo de previsão de demanda para o setor de turismo na Colômbia, também com base em séries temporais, obtendo resultados satisfatórios.

Além disso, outro estudo realizado por Almeida et al. (2018) buscou analisar a influência das variáveis externas na demanda por serviços bancários em agências da Caixa Econômica Federal. Os resultados mostraram que fatores como nível de emprego, inflação e

taxa de juros impactam significativamente na demanda por serviços bancários, o que reforça a importância da previsão de demanda para o planejamento estratégico da instituição.

Diante desse contexto, a previsão de demanda se mostra como uma ferramenta importante para a Caixa Econômica Federal, permitindo a identificação de padrões e tendências na demanda por serviços bancários e o planejamento de capacidade produtiva adequado para atender essa demanda. O uso da metodologia de Box-Jenkins se apresenta como uma opção viável e eficaz para realizar essas previsões, uma vez que permite a análise de séries temporais e a identificação de padrões sazonais e tendências. Com essas informações em mãos, a instituição pode tomar decisões mais acertadas e garantir a satisfação do cliente ao reduzir o tempo de espera em filas e agilizar o atendimento.

Alguns estudos têm utilizado essa metodologia para prever a demanda em sistemas de filas em diferentes contextos. Em um estudo realizado em um hospital em Taiwan, por exemplo, a metodologia Box-Jenkins foi utilizada para prever o número de pacientes em um departamento de emergência [13]. Já em outro estudo, a metodologia foi aplicada para prever a demanda de um sistema de transporte de passageiros em uma cidade na Índia.

No contexto da Caixa Econômica Federal, um estudo recente utilizou a metodologia ARIMA para prever a demanda por atendimento nas agências em diferentes regiões do Brasil [15]. Os resultados mostraram que a metodologia apresentou boa precisão na previsão da demanda e pode ser utilizada como ferramenta para apoiar a gestão da capacidade produtiva nas agências da Caixa.

Em resumo, a gestão eficiente de filas é um desafio para instituições públicas e privadas em todo o mundo. A Caixa Econômica Federal, como uma das principais instituições financeiras do Brasil, enfrenta esse desafio diariamente e pode se beneficiar da aplicação de técnicas de previsão para melhorar a gestão da demanda em suas agências. A construção de um modelo de previsão baseado em séries temporais é uma iniciativa promissora nesse sentido, que pode trazer benefícios para a instituição e seus clientes.

## **Desenvolvimento**

### ***Referencial Teórico***

Uma Série Temporal pode ser definida como uma sequência de valores observados ao longo do tempo, em intervalos iguais. Um dos principais objetivos do estudo de Séries Temporais é criar modelos que demonstrem o comportamento do fenômeno estudado e, a partir daí, gerar previsões. Modelos de previsão são aplicados em diversas áreas da Engenharia, Economia, Ciências Médicas, dentre outros, servindo como base para o planejamento, permitindo avaliar antecipadamente a demanda, projetar capacidade e necessidade de recursos, além de outras atividades.

O primeiro trabalho sobre modelos ARIMA foi publicado em 1970, box e Gwilym Kenning popularizaram o ARIMA (modelo autorregressivo integrado com médias móveis) no livro texto, *time series analysis: Forecasting and control* (Box and Jenkins, 1970).

Modelos ARIMA inicialmente geraram muito entusiasmo na comunidade acadêmica, muito devido a seus fundamentos teóricos. Se suas premissas são atendidas, eles normalmente fornecem ótimas previsões, isso significa que os erros do modelo não contêm informações que poderiam melhorar as previsões. Metodologistas chamam esse fenômeno de ruído branco. Isso não se aplica, no entanto, que os modelos ARIMA são necessariamente superiores as outras opções de previsão, especialmente se os dados não estão em conformidade com as premissas necessárias.

Um estudo de caso em uma indústria eletroeletrônica [8] utilizou a metodologia Box-Jenkins para desenvolver modelos de previsão. O processo envolveu coleta e organização dos dados, validação e formulação dos modelos, avaliação e geração das previsões, e análise e comparação dos resultados. Foram comparados modelos ARIMA, suavização exponencial simples e Holt Winters, utilizando critérios de erro de previsão.

Os erros de previsão considerados foram o MAE (desvio absoluto médio), que mede a acurácia da previsão em relação aos erros absolutos médios; o MAEP (erro absoluto médio percentual), que indica a proporção dos erros em relação aos valores atuais da série; e a estatística U de Theil, que avalia o desempenho das previsões em relação à previsão ingênua.

Após definir os parâmetros de cada método de previsão e utilizar os erros de previsão como ranking, foi constatado que o modelo ARIMA foi a melhor escolha para prever a demanda de eletroeletrônicos, enquanto o SES foi a pior escolha. Isso evidencia que, quando os dados estão bem-comportados, o modelo ARIMA é capaz de explicá-los de forma adequada.

O modelo ARIMA é especialmente adequado para dados bem-comportados devido à sua capacidade de capturar padrões de séries temporais, tendências e sazonalidades. Sua flexibilidade e adaptabilidade permitem ajustar-se a diferentes cenários de previsão, tornando-o uma escolha confiável quando os dados apresentam comportamento estável.

O segundo artigo trata da aplicação do modelo Box-Jenkins na previsão da produção de automóveis [6]. Primeiramente foram recolhidos os dados da produção de automóveis no período entre jan/2007 a jun/2018 e então foi plotado em forma de series temporais, daí iniciou-se a análise dos correlogramas para encontrar os melhores parâmetros para o modelo ARMA, e após essa análise foi testado se o modelo é estacionário, o que de fato aconteceu, ou seja, não é necessário a utilização do modelo ARIMA. E então foi utilizado um teste conhecido como Kruskal-Wallis para entender se há sazonalidade nos dados, e o teste provou que sim. Logo já foi possível encontrar todos os elementos para modelar o problema, após encontrar o modelo ideal ele foi testado e passou no teste de autocorrelação e também de normalidade dos resíduos, e por fim foi plotado as suas previsões. E então foi concluído que de fato o modelo arima mostrou uma ótima opção para a previsão da produção, já que não apresentou tendência

nos dados, ou seja, eles eram aleatórios, o que de fato acontece na vida real, já que a demanda no setor de compras, principalmente de automóveis varia muito ao passar dos meses.

Tais artigos mostraram que se usado corretamente, respeitando os pressupostos e fazendo o diagnóstico dos dados de maneira coerente é possível encontrar ótimas previsões com o método de box-jenkins.

### Procedimento Metodológico

O Lócus da pesquisa se deu em uma agência da CEF, localizada em Jaboatão dos Guararapes, um município do estado de Pernambuco, cidade essa que de acordo com a estimativa do Instituto Brasileiro de Geografia e Estatística é o segundo município do estado com maior número no Índice de Desenvolvimento Humano (IDH) com 0,717, e uma renda média de dois salários-mínimos para os trabalhadores formais, com uma população estimada de 711.330 habitantes e possuindo 5 agências CEF [CEF; IBGE, 2022].

Apesar de um bom IDH frente aos demais municípios do mesmo estado, as pessoas que necessitam utilizar-se dos atendimentos tanto para o Auxílio Emergencial quanto para a retirada do FGTS precisavam se enquadrar em alguns requisitos econômicos dos quais os mesmos não são contribuidores para um IDH elevado. Outro ponto é a localização desta agência, que possibilitava um deslocamento de pessoas de diversas comunidades carentes e inclusive de cidades circunvizinhas próximas que não possuíam agências e com um IDH menor.

A escolha da agência específica se deu pelo fato de um dos autores ser um ex funcionário, portanto conseguiu acesso ao sistema e assim, foi possível a obtenção dos dados, referentes aos anos de 2019 e 2020, pois eram os períodos disponíveis no momento da coleta, as informações foram separadas mensalmente, sendo um total de vinte e quatro observações e foram separados em algumas categorias, para esse presente trabalho serão discutidas três delas, Caixa, Fgts e expresso, pois são as que representam o maior volume de atendimento. O objetivo desse artigo é, a partir da metodologia Box-Jenkins conseguir a quantidade de atendimentos que serão realizados por esses três setores e, assim a caixa poderá alocar recursos de capacidade de atendimento da melhor maneira a atender essa demanda.

Abaixo é apresentado a tabela das estatísticas descritivas dos dados utilizados, para situar o leitor da distribuição dos mesmos.

**Tabela 1.**

#### *Estatística Descritivas dos dados estudados*

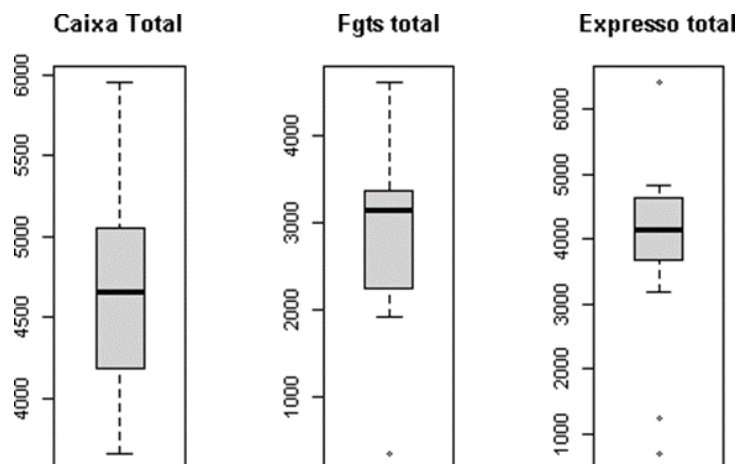
Ano	Grupo	Min	1º Qu	Mediana	Média	3º Qu	Max	Sd
2019	Caixa	3658	4191	4661	4660	4692	5953	641.6
2019	Expresso	693	3688	4136	3867	4588	6421	1442.08
2019	FGTS	358	2237	3148	2810	3365	4606	1048.52
2020	Caixa	995	1151	1821	2325	3538	5314	1321.93
2020	Expresso	657	3160	4634	4611	6161	8128	2414.01
2020	FGTS	1	18.5	77	1194.7	2199	3850	1557.63

*Nome: Caixa Econômica Federal*

Com base nos valores apresentados na tabela 1, é possível perceber uma discrepância entre os dados de 2019 e 2020, isso se dá pelo fato de que em 2020 houve a pandemia e, conseqüentemente o lockdown, isso fez com que as pessoas ficassem incapacitadas de ir às agências, aumentando muito as incertezas de demanda nesse ano, como pode ser visto no desvio padrão dos dados.

**Gráfico 1.**

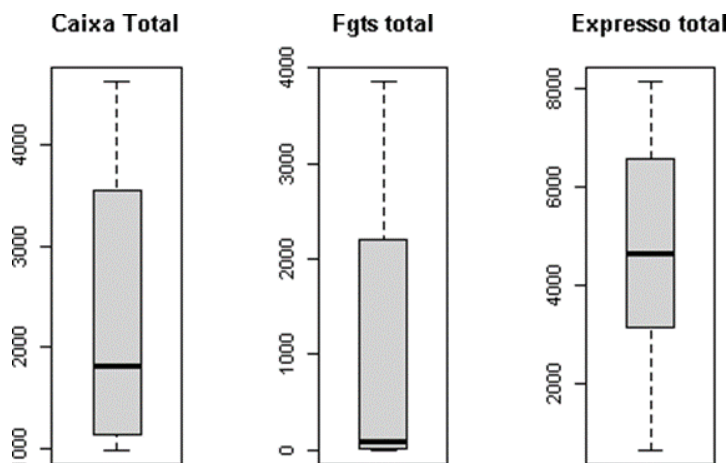
*Box-Plot dos dados para o ano de 2019*



*Nota: Autores com base nos dados da Caixa*

**Gráfico 2.**

*Box-Plot dos dados para o ano de 2020*



*Nota: Autores com base nos dados da Caixa*

Para uma melhor visualização, é apresentado nas figuras 1 e 2 um gráfico de box-plot que é um tipo de gráfico estatístico usado para representar a distribuição dos dados em um conjunto de observações [McGill et al., 1978]. Por meio dele, pode-se perceber como os dados variam, especialmente em 2020, por conta da pandemia, onde houve maiores variações de atendimento, o que dificulta a criação de um modelo assertivo para o cenário, porém pode preparar a local, para algum evento futuro que possa variar a demanda.

O trabalho seguiu várias etapas. Primeiro, os dados foram tratados para se tornarem séries temporais. Em seguida, foram avaliados utilizando estatística descritiva para entender seu comportamento. A análise dos correlogramas ajudou a identificar os modelos a serem considerados. Usando o critério de Akaike (AIC), os melhores parâmetros foram selecionados para explicar os dados, minimizando o AIC. No entanto, apenas o valor mínimo de AIC não é suficiente para afirmar que o modelo é adequado. A presença de sazonalidade nos dados também foi investigada. Depois, o diagnóstico foi realizado para testar se o modelo realmente explica os dados. Isso envolveu a avaliação dos resíduos, incluindo a análise dos correlogramas para identificar correlações e o uso do qqplot para verificar a distribuição normal dos resíduos. Se o modelo passasse por todas essas etapas, com o menor AIC, a melhor sazonalidade e resíduos bem-comportados, ele seria escolhido para prever o futuro da distribuição.

### Modelo ARIMA

Esses modelos consideram três processos, são eles o autorregressivo, de médias móveis e a quantidade de diferenças necessária para que os dados sejam estacionários, essas diferenças são obtidas da seguinte forma  $y_t = y_t - y_{t-1}$ , e isso será feito para todos os dados até que os mesmos sejam estacionários.

Se  $Y_t$  é modelado como  $(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t$  (I)

Em que  $\delta$  é a média de  $Y$  e  $u_t$  é um erro aleatório não correlacionado com média zero e variância constante  $\sigma^2$  (trata-se de um ruído branco), então  $Y_t$  segue um processo autorregressivo estocástico de primeira ordem ou AR(1). Nesse caso o valor de  $Y$  no período  $t$  depende do seu valor no período anterior e de um termo aleatório; os valores de  $Y$  são expressos como desvios com base em um valor médio. Em outras palavras, esse modelo informa que o valor previsto de  $Y$  no período  $t$  é simplesmente alguma proporção ( $=\alpha_1$ ) mais um choque aleatório ou perturbação no período  $t$ ; novamente os valores de  $Y$  são expressos em torno dos seus valores médios.

Mas se  $Y$  segue o seguinte modelo  $(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t + \alpha_2(Y_{t-2} - \delta) + u_t$  (II)

$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t$  (III)

Nesse caso  $Y_t$  é um processo autorregressivo de ordem  $p$ -ésima, ou AR( $p$ ). Processo de média móvel (MA)

Como já foi citado anteriormente o processo AR não é o único mecanismo que pode ter gerado  $Y$ , supondo que  $Y$  é modelado seguindo a equação  $Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$  (IV) Em que  $\mu$  é uma constante e  $u$ , como antes é um termo de erro estocástico de ruído branco. Nesse caso  $Y$  no período  $t$  é igual a uma constante mais uma média móvel dos termos de erro atuais e passados. Portanto, pode-se dizer que  $Y$  segue um processo de média móvel de primeira ordem, ou um MA(1). Mas se  $Y$  segue a expressão  $Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$  (V). Então é um processo MA(2). De forma mais geral,  $Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q}$  (VI)

É um processo MA( $q$ ). Resumindo, um processo de média móvel é apenas uma

combinação linear de termos de erro de ruído branco.

Processo autorregressivo de médias móveis (ARMA)

Pode acontecer de  $Y$  possuir características tanto de AR quanto de MA e seja, portanto, ARMA.  $Y_t$  segue um processo ARMA(1,1) e pode ser escrito como  $Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1}$  (VII) Portanto há um termo autorregressivo e um termo de média móvel. Na equação VII  $\theta$  representa um termo constante.

Em geral, em um processo ARMA(p,q), haverá termos autorregressivos  $p$  e termos de média móvel  $q$ .

Processo autorregressivo de médias móveis (ARIMA)

Se for preciso diferenciar uma série temporal  $d$  vezes para torná-la estacionária e aplica o modelo ARMA(p,q), pode-se dizer que a série temporal é ARIMA(p,d,q), ou seja, ela é uma série temporal autorregressiva integrada de médias móveis, em que  $p$  denota os números dos termos autorregressivos,  $d$  o número de vezes que a série deve ser diferenciada antes de tornar-se estacionária e  $q$  o número de termos de média móvel. Uma série temporal ARIMA(2,1,2) deve ser diferenciada uma vez ( $d=1$ ) antes de torna-la estacionaria, e a série temporal estacionária (de primeira diferença) pode ser modelada como um processo ARMA(2,2), pois possui dois termos AR e dois MA. Claro, se  $d=0$ , uma série é estacionária para ARMA(p,q). Um processo ARIMA(p,0,0) por exemplo, significa um processo AR(p) puramente estacionário; um ARIMA(0,0,q) significa um processo MA(q) puramente estacionário. Dados os valores de  $p, d$  e  $q$  é possível dizer qual processo está sendo modelado.

Foi mencionado anteriormente que o termo  $u$  é um ruído branco, em termos do arima essa característica pode ser descrita como ARIMA(0,0,0), ou seja é um processo estacionário e que não tem parâmetros Autorregressivos e nem de médias móveis.

## Resultados e Discussões

Utilizando os conceitos apresentados anteriormente será possível encontrar previsões coerentes para os dados, o código para análise será desenvolvido na linguagem R.

### Previsão para o Grupo Caixa

O primeiro grupo a ser analisado será o da caixa, como já foi mencionado a metodologia ARIMA que exige que o conjunto de dados ou suas diferenças seja estacionário, então esse teste será feito primariamente.

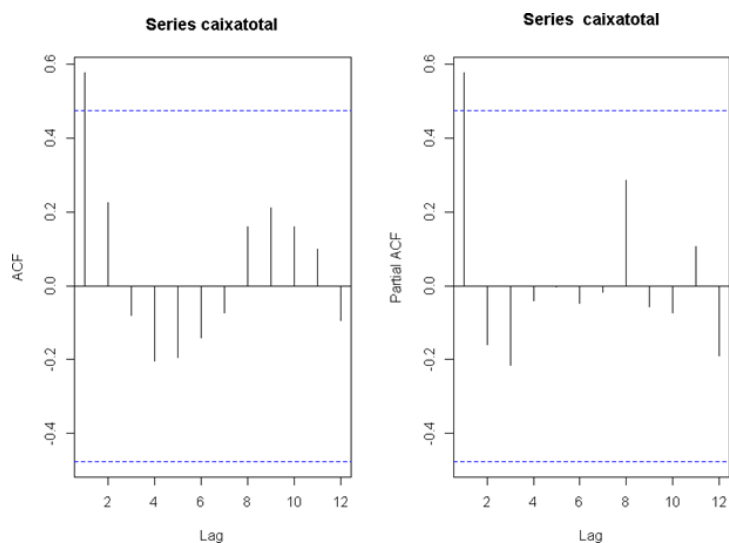
No teste de Dickey-Fuller duas hipóteses são consideradas, a hipótese nula é de que os dados não são estacionários e a alternativa é que são estacionários, então para  $p$ - valores maiores que 0,05 (que foi o valor máximo definido pelo autor) não há evidências suficiente para inferir que os dados são estacionários, logo o resultado gerado pelo teste foi de 0.3869. os dados não são estacionários, portanto, no ARIMA a primeira diferença será considerada.

A próxima etapa será realizar a análise dos correlogramas:



**Gráfico 3.**

*Correlogramas dos dados da caixa*

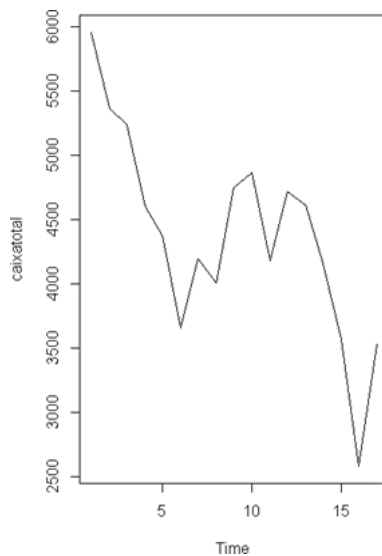


*Nota: Autor (2022)*

Observando o gráfico 3, alguns modelos podem se considerados, são eles  $(1,1,1)$ ,  $(1,1,0)$  e  $(0,1,1)$ , porém isso não é preciso o suficiente, pois os dados dependem de outro fator que é a sazonalidade, mais uma vez deve-se analisar o gráfico para encontrar esse parâmetro. Esse gráfico está representado no gráfico 4.

**Gráfico 4.**

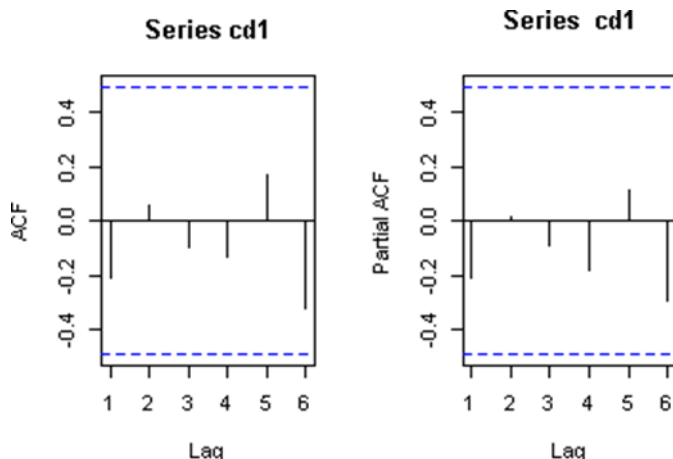
*Plot dos dados da caixa*



*Nota: Autor (2022)*

**Gráfico 5.**

*Correlograma para a sazonalidade da caixa*



*Nota: Autor (2022)*

A partir dessa análise, é possível definir que a melhor frequência para essa sazonalidade será a de 6 períodos. Para a parte sazonal dois modelos foram testados, (1,1,1) e (0,1,0) com base no gráfico 5. A tabela desses modelos é apresentada abaixo.

**Tabela 2.**

*Modelos selecionados para previsão dos dados da caixa*

Modelo	ARIMA	Sazonalidade	AIC
M1	(1,1,1)	(0,1,0)	169.01
M2	(1,1,1)	(1,1,1)	169.41
M3	(0,1,1)	(0,1,0)	167.88
M4	(1,0,0)	(0,1,0)	167.88
M5	(0,1,0)	none	251.06

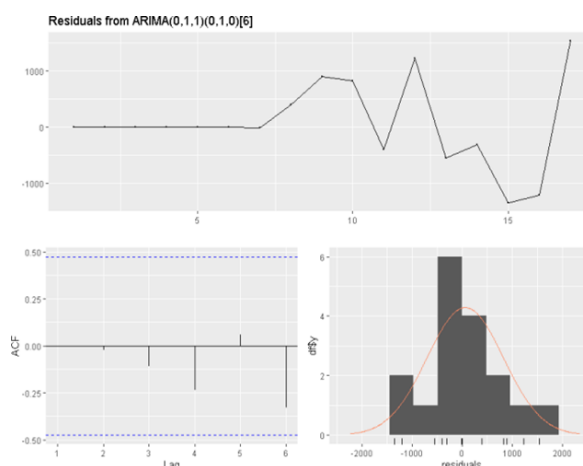
*Nota: Autor (2022)*

Considerando os valores do AIC para tentar encontrar a melhor escolha não é possível analisando simplesmente esse parâmetro, já que os valores de M3 e M4 são iguais, logo deve-se encontrar outra variável que são a soma dos erros conhecida como sigma ao quadrado, para M3 esse valor é de 935767.4 e o do modelo M4 é de 936023.5 como M3 apresenta o menor valor, ele será o modelo escolhido.

Agora que foi encontrado os melhores parâmetros para a explicação dos dados é preciso testar se eles são de fatos bons o suficiente e, para isso é necessário utilizar da etapa de diagnósticos, onde será realizado uma análise sobre os resíduos do modelo, esses plots são encontrados no gráfico 6.

### Gráfico 6.

#### Gráficos para diagnóstico do modelo de previsão para a caixa



Nota: Autores (2022)

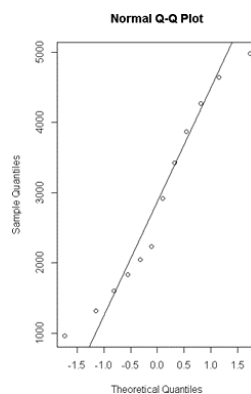
No gráfico 6 será analisado duas coisas, a primeira é se todos os dados estão dentro do intervalo em azul, o que de fato acontece significando que eles não estão autocorrelacionados, isso pode ser confirmado por meio do teste de Ljung-box.

No teste de Dickey – Fuller apresentado anteriormente a ideia era rejeitar a hipótese nula, ou seja, encontrar o valor de p baixo o suficiente, para esse teste o objetivo é diferente, ou seja, a ideia é encontrar um p-value alto o suficiente para que não se tenha todas as evidências para comprovarem que os dados estão autocorrelacionados, que é o que acontece no teste o resultado foi de 0.6399 comprovando a análise do correlograma.

O segundo parâmetro a ser analisado é a distribuição dos resíduos, basicamente a ideia é que ela siga uma distribuição normal, que é mostrado no histograma apresentado no gráfico 4 e ao analisa-lo, parece de fato que eles seguem essa distribuição, porém há duas formas de comprovar tal hipótese, a primeira já mencionada nesse artigo é o qqplot que traz os valores para um intervalo entre -1.5 e 3 para plotar uma reta normal aos dados, se eles estão acima da reta, significa que seguem uma distribuição normal, caso contrário não seguem

### Gráfico 7.

#### Gráfico da dispersão dos dados



Nota: Autor (2022)

Analisando esse plot no gráfico 7 parece de fato que todos estão próximos o suficiente da reta para confirmar a hipótese, porém existe um teste para ter a certeza final e ele é conhecido como teste de Shapiro-wilk, ele se comporta da mesma maneira que o anterior, ou seja a ideia é encontrar um valor de p alto o suficiente para não rejeitar a hipótese nula. O resultado encontrado foi de 0.3033 os resíduos de fato seguem uma distribuição normal.

Após todas essas informações, pode-se concluir que de fato o modelo escolhido explica bem os dados, e então é possível encontrar os valores que são o objetivo desse artigo que são as previsões, que podem ser encontradas na tabela abaixo.

**Tabela 3.**

*Dados das previsões do modelo selecionado*

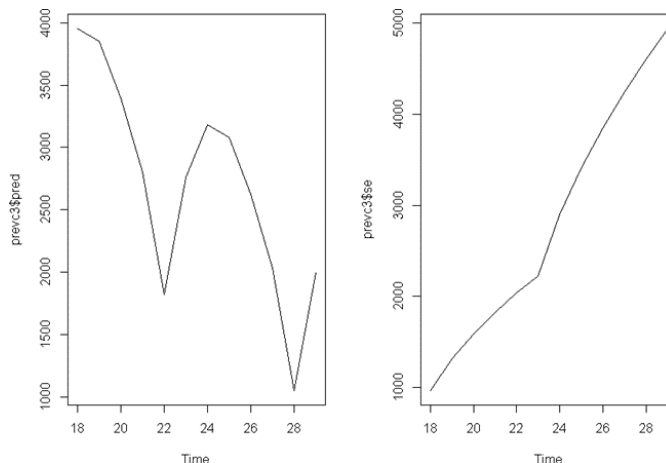
Meses para frente	Previsão
1	3951
2	3847
3	3392
4	2803
5	1821
6	2761
7	3184
8	3080
9	2625
10	2036
11	1054
12	1994

*Nota: Autor, 2022*

E, por último foi plotado o gráfico dessa previsão, juntamente com o gráfico dos resíduos.

**Gráfico 8.**

*Plot da previsão e do erro*



*Nota: Autor (2022)*

## Previsão para o Grupo Expresso

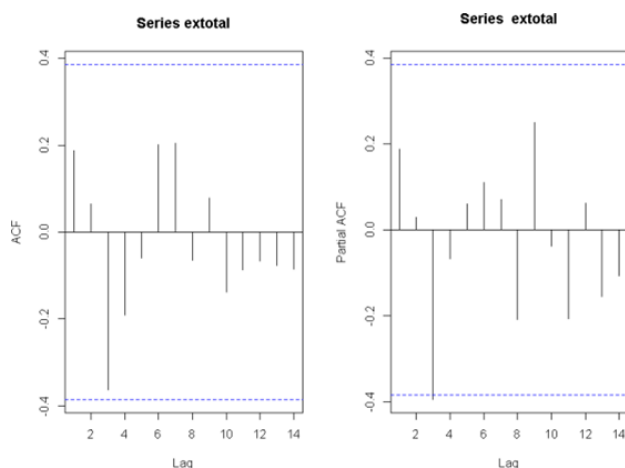
Mais uma vez o primeiro passo é testar se os dados são estacionários, e como já foi mostrado, isso pode ser feito por meio do teste de Dickey-Fuller.

Como o p-valor é maior que 0.05 (0,07238), então o conjunto é dito não estacionário e a primeira diferença será considerada para o modelo ARIMA.

O Próximo passo consiste em analisar o correlograma dos dados.

### Gráfico 9.

*Correlograma dos dados do Expresso*

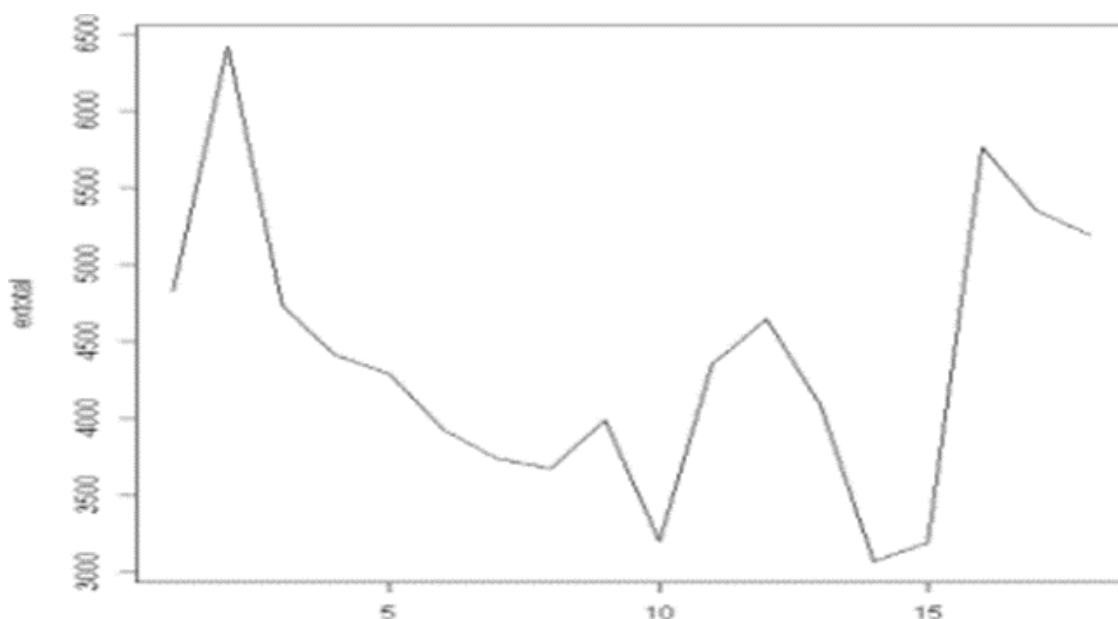


*Nota: Autor (2022)*

Para esse caso, como mostrado no gráfico 9, vários modelos podem ser considerados, mas os principais são: (0,1,0), (0,1,3), (0,1,1) e (1,1,1). Porém, como já foi dito, isso não será suficiente para garantir uma previsão eficiente, então deve-se também analisar a sazonalidade dos dados, se houver.

### Gráfico 10.

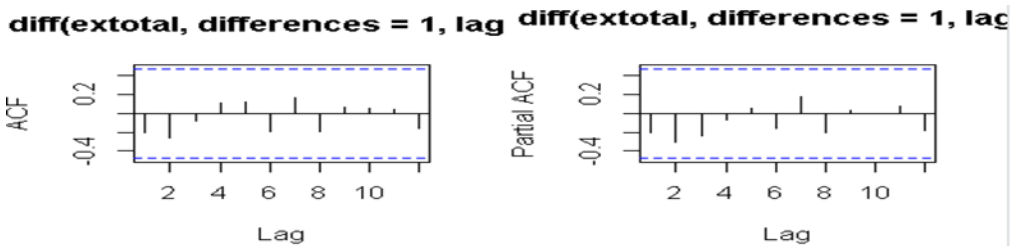
*Plot dos dados do expresso*



*Nota: Autor (2022)*

**Gráfico 11.**

*Correlograma para a sazonalidade*



Nota: Autor (2022)

O período de sazonalidade será de 12, já para os parâmetros será testado apenas (0,10), pois se trata de um passeio aleatório. Com base em todas essas informações é possível selecionar modelos, para tentar prever tais dados.

**Tabela 4.**

*Modelos selecionados para prever os dados do expresso*

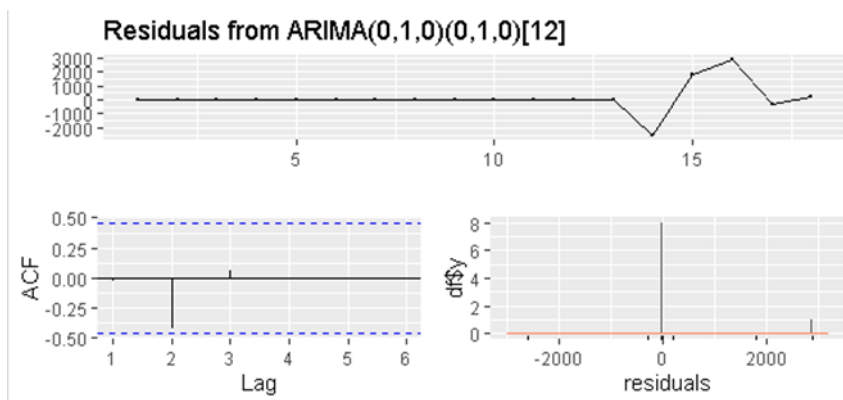
Modelo	ARIMA	Sazonalidade	AIC
M1	(0,1,1)	(0,1,0)	91.78
M2	(0,1,0)	(0,1,0)	89.8
M3	(0,1,3)	(0,1,0)	93.65
M4	(1,1,1)	(0,1,0)	93.4
Auto arima	(0,0,1)	none	297.48

Pelo critério do menor AIC, o M2 apresenta os melhores parâmetros para explicar os dados.

A próxima etapa consiste em diagnosticar o modelo, ou seja, como mostrado anteriormente essa parte serve para mostrar se de fato o modelo escolhido explica bem os dados.

**Gráfico 12.**

*Gráficos para o diagnóstico do modelo escolhido para o expresso*



Nota: Autor (2022)

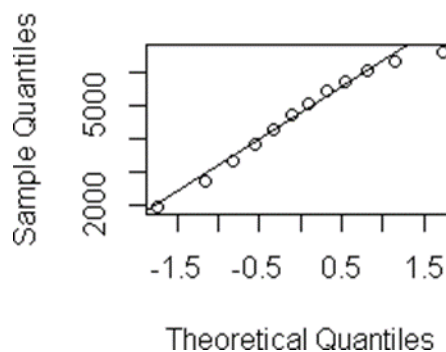
Os resíduos não estão correlacionados, já que no correlograma todos estão dentro da

zona em azul, no gráfico de linhas há 4 pontos que não estão comportados, mas de resto está ok e para avaliar a suposição de que os erros seguem uma normal é necessário plot o gráfico qq.

**Gráfico 13.**

*Teste de normalidade*

**Normal Q-Q Plot**



*Nota: Autor, 2022*

Como é possível notar, os pontos não estão muito próximos à reta (gráfico 13), mas estão próximos o suficiente o que é um indicativo de que de fato os erros seguem uma normal e para confirmar isso é mostrado mais uma vez o teste de Shapiro, cujo resultado foi de 0.7729) confirmando assim a hipótese.

Portanto tendo a confirmação de que o modelo é fato bom o suficiente para explicar bem os dados, pode-se partir para a previsão cuja tabela pode ser encontrada abaixo.

**Tabela 5.**

*Dados com o resultado da previsão para o expresso*

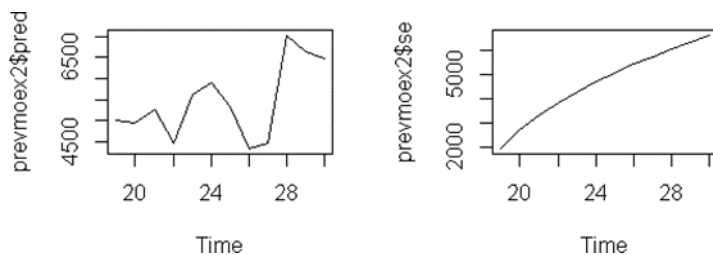
Meses para frente	Previsão
1	5011
2	4943
3	5259
4	4470
5	5624
6	5918
7	5346

*Nota: Autor (2022)*

E, no gráfico 14 é encontrado o plot com os dados dessa previsão juntamente com o erro.

**Gráfico 14.**

*Plot dos dados da previsão e do erro*



*Nota: Autor (2022)*

**Previsão para o Grupo FGTS**

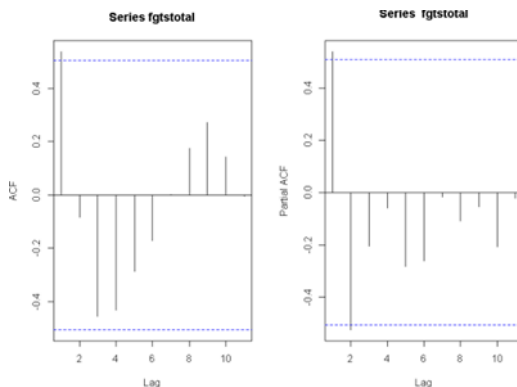
**Teste de estacionariedade**

O conjunto não é estacionário, pois o p-valor foi de 0.2314 então deve-se aplicar as primeiras diferenças.

Partindo para a análise dos correlogramas para a parte ARIMA.

**Gráfico 15.**

*Correlograma para os dados do FGTS*

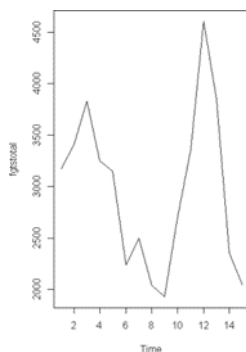


*Nota: Autor (2022)*

Para esse caso os modelos considerados serão, (0,1,0), (1,1,1), (0,1,2), (0,1,1) e (1,1,0). Já para parte da sazonalidade.

**Gráfico 16.**

*Plot dos dados do FGTS*



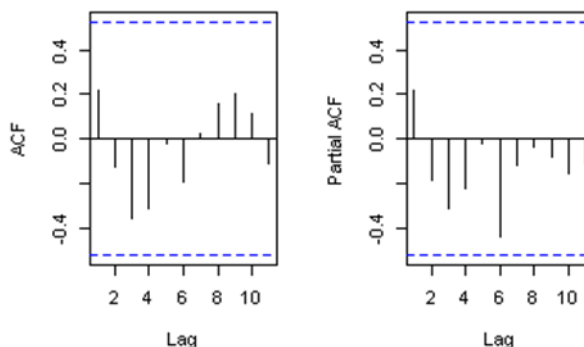
*Nota: Autor (2022)*



**Gráfico 17.**

*Correlograma para a sazonalidade*

$\text{diff}(\text{fgtstotal}, \text{differences} = 1, \text{lag} = 1)$   $\text{diff}(\text{fgtstotal}, \text{differences} = 1, \text{lag} = 1)$



Nota: Autor (2022)

O período da sazonalidade será de 9 e dados o correlograma apenas o conjunto (0,1,0) será testado, já que não há lags fora da zona em azul.

**Tabela 6.**

*Modelos selecionados para prever os dados do FGTS*

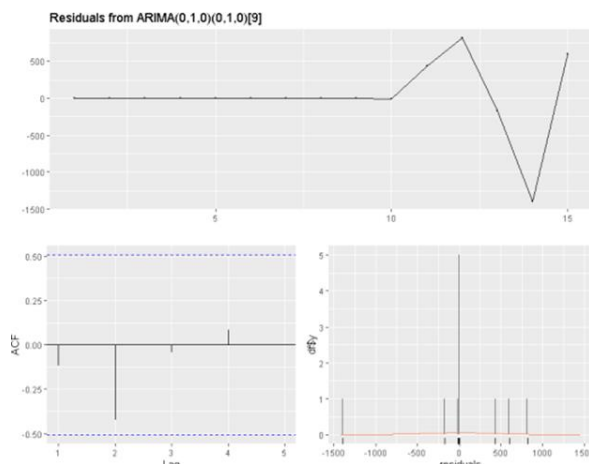
Modelo	ARIMA	Sazonalidade	AIC
M1	(0,1,0)	(0,1,0)	81.0252
M2	(1,1,1)	(0,1,0)	83.7655
M3	(1,1,0)	(0,1,0)	82.9552
M4	(0,1,1)	(0,1,0)	82.1
Auto.Arima	(0,0,1)	none	240.76

Fonte: Autores (2022)

Menor AIC pertence ao modelo M1, logo será ele o escolhido para a parte do diagnóstico. Oplot dos resíduos é encontrado no gráfico 18.

**Gráfico. 18**

*Diagnóstico do modelo escolhido para prever o FGTS*

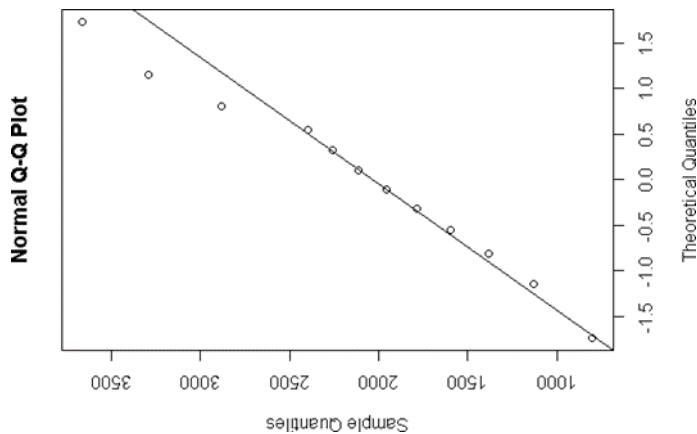


Nota: Autor (2022)

Todos os lags estão dentro da zona estatisticamente aceitável, logo não estão autocorrelacionados, mas mais uma vez não é possível a partir do histograma afirmar que eles seguem uma distribuição normal, logo o qqplot (gráfico 17) e o Shapiro (figura 7) serão usados.

**Gráfico 19.**

*Plot de normalidade*



*Nota: Autor (2022)*

Como pode ser visto no gráfico 17, há 3 pontos de alavancagem, porém estão na mesma direção dos outros pontos, logo são pontos de alavancagem bons e como foi encontrado no teste de Shapiro, o p-valor é de 0.9661, não interferem na normalidade dos resíduos.

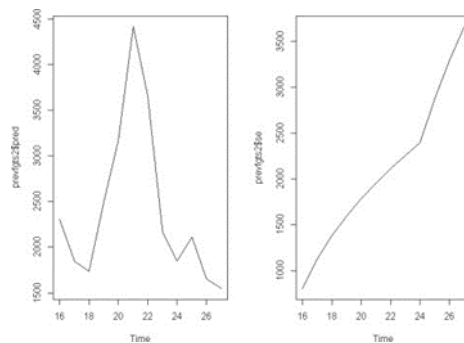
**Tabela 6.**

*Dados com a previsão para o FGTS*

Meses para frente	Previsão
1	2304
2	1845
3	1738
4	2500
5	3172
6	4413
7	3657
8	2161
9	1851
10	2111
11	1652
12	1545

*Nota: Autor (2022)*

Logo, o modelo explica bem os dados e suas previsões podem ser encontradas na tabela abaixo. No gráfico 20 é encontrado o plot da previsão e do erro.

**Gráfico 20.***Plot com os dados da previsão e do erro**Nota: Autores (2022)***Conclusões**

O estudo de caso apresentado, mostrou diversos modelos de previsão baseados na metodologia box Jenkins, onde foi possível encontrar bons resultados acerca dos dados fornecidos por uma agência da caixa econômica federal, todos os modelos seguiram métodos necessários para que não haja erros exorbitantes quando se é comparado com os originais, isto é, passaram pela etapa de diagnóstico e todos se mostraram bons o suficiente para a explicação dos dados.

Dito isso o objetivo do artigo foi alcançado, ou seja, foi possível mostrar quanto será a demanda de clientes para a instituição para que assim eles procurem maneiras de suprir tal demanda.

O estudo de caso apresenta limitações nos dados, isto é, não havia tantos para entender suficientemente bem o real comportamento dos clientes, já que vários elementos acerca do ano de 2020 foram retirados devido à pandemia.

Por fim, recomenda-se pesquisas futuras sobre a metodologia ARIMA, já que como foi demonstrado é um modelo de fácil entendimento, fácil aplicação e devido as suas etapas é também muito confiável, também é recomendado o uso de variáveis exógenas, em outras palavras são variáveis que estão fora do modelo, mas que tem relação com ele e, assim é possível encontrar melhores regressores para o modelo final, sempre utilizando critério de AIC para determinar se não variáveis boas ou não para o modelo.

**REFERÊNCIAS**

- Silveira, F., Vieira, M. M. F., & De Castro, D. C. (2013). A Presença do estado no setor financeiro Brasileiro: o caso da Caixa Econômica Federal. *GESTÃO. Org*, 11(1), 132-159.
- Santos, A. C., EPA, U., do Espírito Santo, R. S., & da Rocha, R. P. Um estudo sobre a Previsão de Demanda (PD) para gestão da Capacidade Produtiva (CP) e Planejamento Agregado da Produção (PAP) em uma empresa do setor de biotecnologia.

- Castel-Branco, C. N. (2015). Desenvolvimento empresarial e da capacidade produtiva no crescimento económico e na industrialização. Questões sobre o desenvolvimento produtivo em Moçambique, 13.
- Marques, C. F. (2012). Estratégia de gestão da produção e operações. IESDE BRASIL SA.
- Monteiro, E. B., Lopes, F. L., & Silva, M. R. (2017). Capacidade produtiva na gestão de serviços: revisão sistemática da literatura. Revista Brasileira de Gestão e Inovação, 5(2), 1-22.
- Sampaio, P., Cabral, J., & Costa, J. (2019). The impact of waiting lines on service quality and customer satisfaction in healthcare. International Journal of Healthcare Management, 12(2), 89-94.
- Cruz, F. R. B., & Delamaro, M. E. (2019). Análise de séries temporais: um estudo sobre modelos auto-regressivos integrados de médias móveis. Revista Eletrônica de Iniciação Científica em Computação, 19(1), 47-61.
- Gujarati, D. N., & Porter, D. C. (2011). Econometria básica-5. Amgh Editora.
- Ehlers, R. S. (2007). Análise de séries temporais. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná, 1, 1-118.
- Morettin, P. A. (2017). Econometria financeira: um curso em séries temporais financeiras. Editora Blucher.
- Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.