

MODELO SARIMA APLICADO À PREVISÃO DE DEMANDA DE USUÁRIOS DE TRANSPORTE PÚBLICO ENTRE CAMPOS DOS GOYTACAZES E SÃO JOÃO DA BARRA

Camila Romão da Silva Sanches¹

Bacharela em Engenharia de Produção

Gilza Santos Simão Ferreira²

Mestre em Engenharia de Produção

Geísa Pereira Marcilio Nogueira³

Mestre em Engenharia de Produção

Niander Aguiar Cerqueira⁴

Doutor em Engenharia Civil

Resumo

A previsão de demanda é uma atividade importante para auxiliar na determinação dos recursos necessários para a empresa. Nesse sentido faz-se necessário uma ferramenta que auxilie o tomador de decisões neste processo. O objetivo desta pesquisa é realizar uma previsão de demanda de usuários de uma linha de transporte coletivo intermunicipal, utilizando a metodologia Box & Jenkins. Foram reunidos dados históricos de janeiro de 2014 a dezembro de 2017 de uma empresa prestadora de serviços de transporte rodoviário coletivo, com itinerário fixo, intermunicipal, com exclusividade entre os municípios de Campos dos Goytacazes (CG) & São João da Barra (SJB). A obtenção dos modelos mais adequados foi baseada na análise de gráficos e em testes estatísticos próprios da

¹Centro Universitário Redentor, Engenharias, Campos dos Goytacazes-RJ, camilaromaos@gmail.com.

²Centro Universitário Redentor, Engenharias, Campos dos Goytacazes-RJ, gilzasimao@hotmail.com.

³Universidade Candido Mendes, Programa de Pós-Graduação em Engenharia de Produção, Campos dos Goytacazes-RJ, isamarcilio@yahoo.com.br.

⁴Centro Universitário Redentor, Engenharias, Campos dos Goytacazes-RJ, coord.niander@hotmail.com.

metodologia, os quais subsidiaram a decisão de adotar o modelo SARIMA (0,1,0)(2,1,1)₁₂ para realizar as previsões.

Palavras-chave: previsão de demanda; modelos Box-Jenkins; modelos SARIMA.

Abstract

Forecasting demand is an important activity to assist in determining the resources needed by the company. The objective of this research is to perform a demand forecast of users of an intermunicipal collective transportation line, using the Box & Jenkins methodology. Historical data were gathered from January 2014 to December 2017 of a company providing collective road transport services, with fixed itinerary, intermunicipal, with exclusivity between the municipalities of Campos dos Goytacazes (CG) & São João da Barra (SJB). The most suitable models were based on the analysis of graphs and statistical tests of the methodology, which subsidized the decision to adopt the SARIMA model (0,1,0) (2,1,1)₁₂.

Keywords: forecast demand; Box-Jenkins models; SARIMA models.

INTRODUÇÃO

O nível de desenvolvimento econômico e social de uma sociedade tem relação direta com o como ela consegue se locomover e transportar suas cargas, ou seja, com a possibilidade de ir e vir. Essa capacidade de o cidadão se locomover possibilita atividades cotidianas tão essenciais, capazes de proporcionar qualidade de vida em sua totalidade. (FERRAZ & TORRES, 2004).

A prestação do serviço de transporte, quando feita com qualidade, beneficia a toda a população e produz satisfação, fazendo com que todos escolham este tipo de transporte: o coletivo. Com isso, é possível diminuir os congestionamentos das vias, reduzir a quantidade de poluentes emitidos na queima do combustível e aumentar os espaços físicos nas cidades, porque evita-se os incontáveis estacionamentos. Enfim, o transporte coletivo é essencialmente importante para uma sociedade (VASCONCELLOS, 2013).

Um fator importante na efetividade da prestação de um serviço, particularmente o que está sendo tratado neste trabalho, é prever a demanda de usuários do transporte coletivo. Conforme relatado por Werner *et al* (2006), a imprecisão da demanda pode acarretar indisponibilidade do serviço para os seus usuários, ou quando se disponibiliza o produto/serviço em excesso, pode acarretar desperdícios, ociosidades, custos elevados com a mão-de-obra etc. Assim o objetivo gerencial é otimizar os recursos e aplicar métodos adequados para a previsão de demandas; visando atender, da melhor forma possível, a necessidade real dos usuários, proporcionando a adequação do serviço à população.

Para elaborar previsões de demanda podem-se usar técnicas qualitativas e quantitativas ou, mesclar as duas técnicas. De acordo com Pellegrini & Fogliatto (2000), os métodos quantitativos empregam séries temporais para prever a demanda em períodos futuros. Entre as técnicas quantitativas aplicadas a previsões de dados históricos a metodologia de Box-Jenkins é uma das técnicas mais difundida na literatura. A modelagem Box-Jenkins usa a ideia de que cada valor da série (temporal) pode ser [descrito](#) por valores passados, com base no uso da estrutura de correlação temporal que normalmente existe entre os valores da série.

Nesse contexto, esse artigo elege um modelo de previsão de demanda de usuários de uma linha de transporte coletivo intermunicipal, com o intuito de poder contribuir em futuras ações por parte de entidades governamentais. Para gerais tais previsões aplica-se a metodologia de Box-Jenkins.

DESENVOLVIMENTO

Modelos Box-Jenkins

Segundo Morettin & Tolo (2006), uma série temporal é qualquer conjunto de observações ordenadas no tempo. Descreve-se uma série temporal da seguinte forma: z_t , $t = 1, 2, \dots, n$, onde n representa o número de observações da série. Uma série temporal é composta por quatro elementos:

Tendência (T_t): verifica o sentido de deslocamento da série, ao longo de vários anos, sendo possível aumentar, diminuir ou estabilizar.

Ciclo (C_t): movimento ondulatório que, ao longo de vários anos, tende à periodicidade, ou seja, é o movimento da série que se reafirma ao longo dos períodos de tempo, tendo uma longa duração, a qual muda de ciclo para ciclo ou se repete, em períodos muito longos.

Sazonalidade (S_t): descreve as flutuações periódicas de comprimento constante, repetindo em períodos fixos, o comprimento do período é denotado por “S”, associado, na maioria dos casos, a mudanças climáticas.

Ruído aleatório ou erro (a_t): é tudo aquilo que não é explicado pelas outras componentes da série, ou seja, é o que o modelo estimado não consegue captar.

Os modelos utilizados para representar as séries temporais, são processos estocásticos, ou seja, são processos controlados por leis probabilísticas (MORETTIN & TOLOI, 2006).

A criação destes modelos de séries temporais consiste, de vários fatores, tais como o comportamento do fenômeno ou o conhecimento *a priori* que temos de sua natureza e do objetivo da pesquisa. Na prática, depende também, da existência de métodos adequados de estimação, e da disponibilidade de *softwares* próprios.

Visto que uma série temporal dispõe de dados obtidos continuamente ao longo do tempo, pressupõem-se que ela apresente correlação ordenada no tempo. Os modelos de Box-Jenkins, em geral são conhecidos por ARIMA- Autorregressivos Integrados de Médias Móveis (*Auto Regressive Integrated Moving Averages*), esses modelos matemáticos objetivam [capturar](#) o comportamento da autocorrelação existente entre os valores da série temporal, baseado nesse comportamento é possível realizar previsões futuras. As previsões geradas serão boas se a correlação existente na série for adequadamente modelada.

De acordo com Fava (2000), os modelos ARIMA resultam da combinação de três componentes denominados “filtros”: o componente autorregressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA). No entanto a série pode ser modelada pelas três componentes ou por combinações dos mesmos, derivando em alguns modelos que serão abordados a seguir.

Modelos estacionário

Os modelos estacionários são aqueles que assumem que o processo está em equilíbrio. Neste caso a série temporal desenvolve-se no tempo em torno de um valor constante, não necessitando de nenhuma transformação matemática para estacionarizá-la. Um processo é considerado fracamente estacionário ou fortemente estacionário. Será fracamente estacionário se suas médias e variância se mantêm constantes, ao longo do tempo. Um processo é fortemente estacionário se todos os elementos conjuntos são invariantes a translações do tempo.

Modelo Autorregressivo – AR (*auto regressive*)

Um modelo autorregressivo a série histórica Z_t , é determinado por seus valores passados regredidos e mais o ruído aleatório a_t . O modelo que contém $\phi(B) = 1$ é denominado como modelo autoregressivo, sendo indicado por AR(p), onde ‘p’ denota a ordem do modelo, ou seja, o número de defasagens.

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad 4$$

Onde:

$\tilde{Z}_t = Z_t - \mu$, onde o sinal \tilde{Z}_t representa os valores subtraídos da média do processo, ϕ_p , são as constantes da função autorregressiva; p , número de instantes anteriores usados na regressão; e a_t , ruído aleatório.

A série no instante t depende das observações passadas até o instante $(t - p)$. Assim, o processo autorregressivo é dito de ordem p .

Definindo-se o operador de retardo, B , tal que:

$$\tilde{Z}_{t-1} = B\tilde{Z}_t \quad 5$$

E, conseqüentemente:

$$\tilde{Z}_{t-m} = B^m \tilde{Z}_t \quad 6$$

A equação da série pode ser dada por:

$$\phi(B)\tilde{Z}_t = a_t \quad 7$$

Onde:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad 8$$

O modelo autorregressivo de ordem 1 e 2 pode ser representado pelas expressões 9 e 10, respectivamente:

$$AR(1) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t \quad 9$$

$$AR(2) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t \quad 10$$

O modelo será denominado estacionário quando $\phi(B)$ convergir para $|B| \leq 1$ como $\phi(B) = 0$, intitulada condição de estacionariedade. Esta condição do AR(p), determina que todas as raízes da equação precisam cair fora do círculo unitário. E, o processo de ordem finita, sempre será inversível.

Tem-se, portanto, que as autocovariâncias não dependem do t , mas sim da ordem p . Tendo $|\phi| \leq 1$, pelo requisito de estacionariedade, quanto maior o valor de p , isto é, quanto maior o intervalo entre as observações, menor a autocovariância.

A identificação do modelo se dará através das funções de autocorrelação e autocorrelação parcial, que mostram a ordem do modelo.

Modelo de Médias Móveis – MA (*moving average*)

Neste modelo Z_t será uma função soma algébrica ponderada dos a_t , que se deslocam no tempo. O valor da série é uma função linear dos valores atuais e passados de uma série de ruídos brancos $a_t, a_{t-1}, \dots, a_{t-q}$, definido pelos erros de previsão ou resíduos das observações disponíveis. Nota: MA (q) sendo 'q' o indicador da ordem do modelo. Desta forma, tem-se:

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad 12$$

E, similar ao método autorregressivo pelo operador B, tem-se:

$$\tilde{Z}_t = \theta(B)a_t \quad 13$$

Onde:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad 14$$

Sendo θ um parâmetro do modelo.

O modelo MA(1) e MA(2) podem ser observados a seguir. O \tilde{Z}_t corresponde a valores do processo subtraídos de sua média.

$$MA(1) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} \quad 15$$

$$MA(2) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} \quad 16$$

Modelos Autorregressivos de Médias Móveis – ARMA (*autoregressive moving average*)

O modelo autorregressivo de médias móveis tem ordem (p,q) com p termos autorregressivos, que representam a ordem de defasagem dos termos autorregressivos, e q termos médias móveis, sendo que este representa a ordem de defasagem do termo de média móveis. Deste modo, tem-se:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad 17$$

E pode-se escrever:

$$\phi(B)Z_t = \theta(B)a_t \quad 18$$

Modelos Não-Estacionários

Como os processos vistos no dia-a-dia são, na sua maioria, não estacionários, é necessário então, encontrar algum tipo de operador que, através de transformações matemáticas, converta a série de dados para estacionários:

Modelo não-estacionário – *Arima (autoregressive integrated moving average)*

A teoria para séries temporais lida com processos estocásticos estacionários, que são séries temporais onde não há uma mudança sistemática na sua média (sem tendências), não há mudanças sistemática na sua variância e sem variações periódicas.

Um processo estacionário, que se caracteriza por possuir um comportamento geral de sua estrutura probabilística, invariante no tempo, pressupõe que uma realização amostral $x_t, x_{t+1}, \dots, x_{t+n}$, para qualquer t , embora aleatória, apresente a mesma forma geral de suas observações, o que possibilita fazer previsões.

Se o processo estocástico não for estacionário, este pode se tornar estacionário por meio de sucessivas diferenciações da série original. Para isso, diferencia-se a série temporal através de um operador ∇ tal que:

$$\nabla Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \quad 19$$

$$\nabla^d = (1 - B)^d Z_t \quad 20$$

Que fornece o modelo:

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \quad 21$$

Onde:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad 22$$

O processo acima é de um modelo ARIMA com ordem (p,d,q) .

A construção do modelo ARIMA parte da concepção de que as séries temporais envolvidas na análise são geradas por um processo estocástico estacionário.

Modelo Sazonal Autorregressivo Integrado Médias Móveis – *SARIMA (Seasonal Integrated Autoregressive Moving Average)*

Os modelos Box & Jenkins analisam a correlação existente entre as observações da série em um determinado instante do tempo e seus antecessores, mas existem observações que apresentam correlação com observações de defasagem superior a 1. Os modelos que contemplam as séries com essa característica são conhecidos como SARIMA.

O modelo SARIMA $(p, d, q)(P, D, Q)$ é composto por uma parte não sazonal (ARIMA) com parâmetros (p, d, q) e uma parte sazonal com parâmetros (P, D, Q) , denotado pela equação:

$$\phi(B)\Phi(B^m)\nabla^d\nabla_m^D Z_t = \theta(B)\Theta(B^m)a_t \quad 30$$

Onde: $\Phi(B^m) = 1 - \Phi_1(B_1^m) - 1 - \Phi_2(B_2^{2m}) - \dots - 1 - \Phi_p(B_p^{pm})$ é o operador autorregressivo sazonal de ordem p .

$\Theta(B^m) = 1 - \Theta_1(B_1^m) - 1 - \Theta_2(B_2^{2m}) - \dots - 1 - \Theta_p(B_p^{pm})$ é o operador de médias móveis sazonal de ordem Q .

O operador $\nabla_m^D Z_t = Z_t - Z_{t-m}$, onde:

$$\nabla_m^D = (1 - B^m)^D, \text{ onde:}$$

Onde: D , é o número de diferenças sazonais e m é o período sazonal.

Para identificação, verificação e estimação de modelos sazonais não há, em princípio, nenhuma dificuldade adicional. A diferença é que temos que distinguir a série em relação ∇ e ∇_m a fim de produzir estacionariedade. Com isso obtemos os valores de d e D , que na maioria das vezes assumem valores iguais no máximo, 2 (MORETTIN & TOLOI, 2006).

A obtenção das ordens p, P, q e Q é feita, como no modelo não sazonal, através das funções de autocorrelação e autocorrelação parcial da série diferenciada.

Etapas da metodologia Box & Jenkins

O objetivo da metodologia Box & Jenkins é determinar os três componentes que configuram qualquer estrutura a saber: p parâmetros autorregressivos, d processos de diferenciação (integração) e q parâmetros de médias móveis. O processo de construção desses modelos propostos por Box & Jenkins é baseado em um ciclo iterativo, composto dos seguintes estágios: identificação, estimação, análise do modelo e previsão.

Identificação do Modelo – Nesta etapa onde se estabelecem as características básicas da série, tais como o tipo do processo gerador (estacionário ou não-estacionário); a existência ou não de componentes sazonal e tendencial; e as estimativas iniciais para os valores dos parâmetros envolvidos.

O processo de identificação do modelo ARIMA deve se iniciar com o parâmetro d . Sua definição é muito simples, e pode ser feita pela ótica da estacionariedade, onde o parâmetro d nada mais é do que o grau de diferenciação necessário para tornar os dados estacionários. A definição dos parâmetros deve ser feita através da análise das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) dos dados históricos.

Estimação dos parâmetros - Esta etapa consiste em determinar se a série estacionária, original ou transformada, é gerada por um processo autorregressivo ($p \neq 0$), de médias móveis ($q \neq 0$), e/ou mista. Neste caso é necessário determinar o valor de p e/ou q . As características da FAC e da FACP em cada caso, é que indicarão qual o possível processo gerador da série.

Análise Do Modelo - Uma vez identificado e estimado o modelo Box & Jenkins, a próxima etapa será a análise para verificar se o mesmo é adequado ou não. Se o modelo for adequado, pode-se adotá-lo para fazer previsão. Mas se o mesmo não é adequado, então, outra especificação deverá ser escolhida para modelar a série, o que resulta refazer as etapas de identificação e estimação. Normalmente, considera-se a análise do resíduo e avaliação da ordem do modelo como meios para verificar a adequação do mesmo.

Previsão - Após o pesquisador haver selecionado, entre os modelos estimados, aquele que se mostrar mais adequado e que tiver o menor AIC e BIC, chega-se à última etapa da metodologia de Box & Jenkins, que consiste na realização de previsões da série em instantes de tempo posteriores.

Estudo de Caso

Para a execução deste trabalho foi realizada uma pesquisa de campo em uma instituição pública de transporte da cidade de Campos dos Goytacazes, o IMTT, para obter os dados referentes número de passageiros que utilizaram o transporte público entre Campos dos Goytacazes e São João da Barra entre janeiro de 2014 e dezembro de 2017. Neste período, a empresa responsável pelo trecho, contava com uma frota de aproximadamente 21 veículos e os dados compõem uma série temporal mensal.

Ressaltamos, aqui, que esta linha de ônibus é realizada por uma única empresa, o transporte alternativo não foi considerado nesta pesquisa. A escolha por essa linha de

ônibus foi motivada pela importância econômica das duas cidades, Campos dos Goytacazes e São João da Barra, principalmente após a instalação do Porto do Açú na cidade de São João da Barra.

A cidade de Campos dos Goytacazes, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), possui uma população estimada de 490.288 habitantes para o ano de 2017, definido pelo último censo, e uma área de aproximadamente 4.026,696 Km² de extensão. Dispõe de 34,2% de suas vias públicas urbanizadas, até o último censo no ano de 2010.

Já a cidade de São João da Barra possui uma estimativa pelo IBGE, de 35.174 habitantes para o ano de 2017, com área extensiva de 455.044 Km². Dispõe de 20,5% de suas vias públicas urbanizadas, até o último censo no ano de 2010.

No tratamento dos dados usou-se o pacote estatístico Gretl. Optou-se por este pacote, pois o mesmo é livre e de fácil acesso e manipulação dos comandos.

Resultados

Para a aplicação da metodologia de Box & Jenkins, conhecidos como Modelos ARIMA (modelos autorregressivos de médias móveis), utilizou-se a série temporal do número de usuários de transporte coletivo, atendidos pela linha Campos dos Goytacazes (CG) X São João da Barra (SJB), entre janeiro de 2014 a dezembro de 2017.

Inicialmente, analisou-se o gráfico de tempo da série em estudo com objetivo de identificar os modelos adequados. Na análise gráfica, é possível observar a presença de tendência e/ou sazonalidade, o que demonstra se a série é ou não estacionária.

Em relação à série temporal do número de atendimentos de usuários de transporte público entre CG&SJB, mostrado na Figura 1, percebe-se uma tendência decrescente. Logo, há indícios de não-estacionariedade da série em virtude dessa tendência.

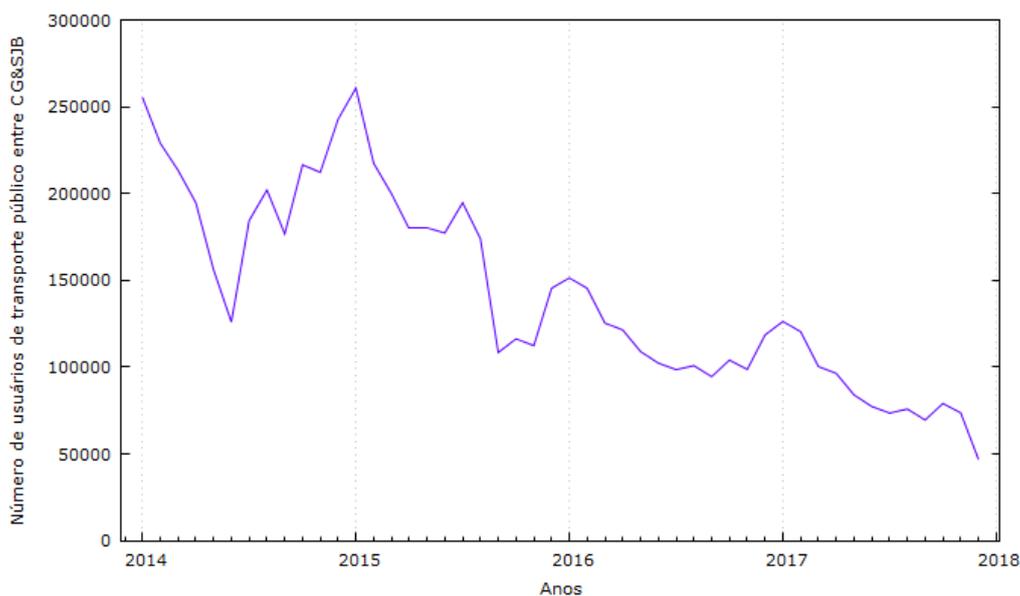


Figura 1 - Série temporal dos usuários de transporte público entre CG&SJB de 2014 à 2017, e sua tendência

Fonte: EVIEWS, versão 10

Outro fator observado na série é a sazonalidade, pois ocorrem variações regulares de picos e vales no número de passageiros transportados. Isso deve ser causado pelas oscilações, devido a feriados, períodos de férias escolares, etc., que estão geralmente relacionados às estações do ano, e que se repetem, com maior ou menor intensidade.

A seguir, é realizada a análise das funções de autocorrelações (FAC) e autocorrelações parciais (FACP), que auxiliam na verificação da estacionariedade e na proposição de um modelo. A Figura 6 mostra os gráficos, respectivamente, da FAC e FACP do número de usuários de transporte público entre CG&SJB entre janeiro de 2014 e dezembro de 2017. Observando-se a Figura 2, constata-se o decaimento exponencial acompanhado de senos e cossenos das funções de autocorrelações FAC. Esse comportamento é típico do processo autorregressivo MA(1). Já a análise do gráfico da PACF, indica a autocorrelação parcial de ordem um significativa. Assim, há indicação de que a ordem do modelo autorregressivo é um, modelo AR(1).

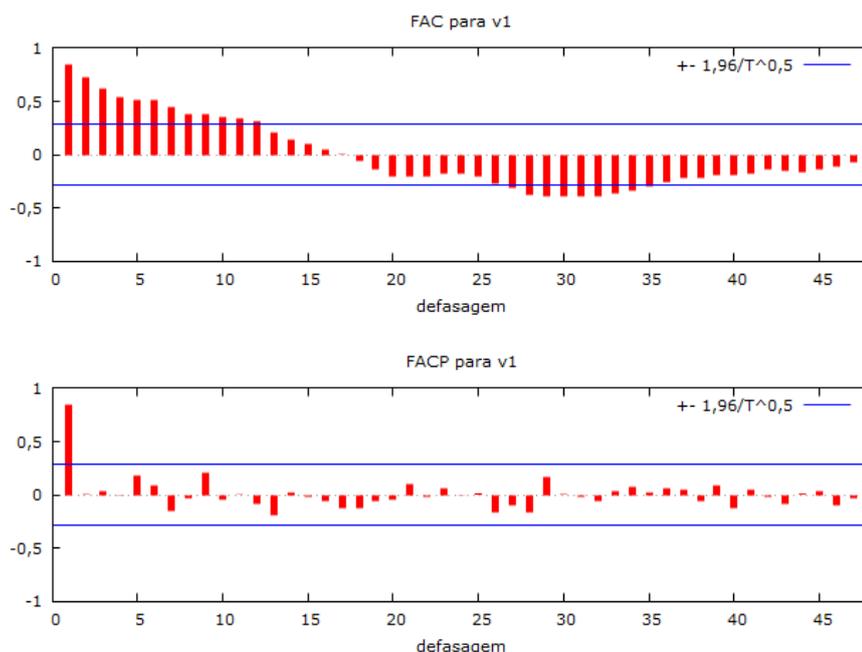


Figura 2 - Função de autocorrelação e função de autocorrelação parcial para o número de atendimentos de clientes de transporte público entre CG&SJB
 Fonte: GRETL, versão 3

Além da análise gráfica da série, foi realizado, também, o teste de raiz unitária, Dickey-Fuller (DF), para verificar se o comportamento da série era estacionário ou não. Para tal, utilizou-se o pacote econométrico GRETL, (versão 3.0), em que a hipótese de não-estacionariedade.

Logo, diante da constatação de não-estacionariedade da série, a mesma foi diferenciada uma vez e novamente o teste DF é realizado. Os resultados deste segundo teste estão resumidos na Figura 3. Observa-se que a estatística τ é de -4,93, e os valores críticos da estatística Dickey_Fuller aos níveis de 1% e 5% são, respectivamente, -2,62 e -1,95. Logo, a estatística τ é menor que os valores críticos e rejeita-se a hipótese nula de raiz unitária; ou seja, a série após uma diferenciação, é estacionária. Assim, tem-se a indicação de que a ordem de integração deve ser um, isto é, $d = 1$.

| | t-Statistic | Prob.* |
|----------------------------------------|-------------|--------|
| Augmented Dickey-Fuller test statistic | -4.930293 | 0.0000 |
| Test critical values: 1% level | -2.617364 | |
| 5% level | -1.948313 | |
| 10% level | -1.612229 | |

*MacKinnon (1996) one-sided p-values.

Figura 3 - Resumo do Teste DF
 Fonte: GRETL, versão 3

O gráfico da série, após uma diferenciação, é mostrado na Figura 4. Verifica-se que a série diferenciada apresenta-se mais estável que a série original. Logo, ela deve ser utilizada para se estimar o modelo que represente a série em estudo.

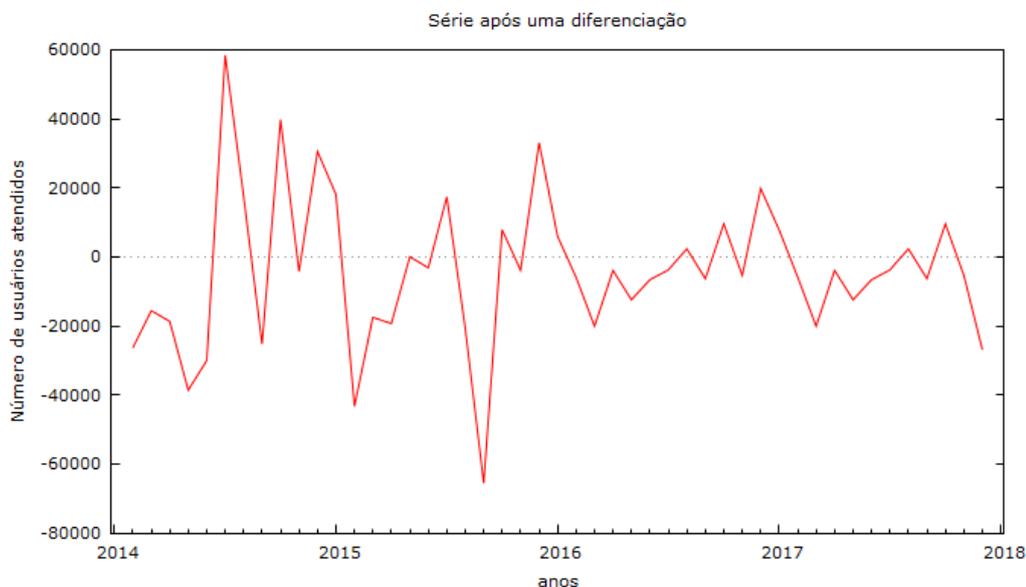


Figura 4 - Série do número de atendimentos de clientes de transporte público entre CG&SJB após uma diferenciação.
 Fonte:GRETl, versão 3.

Com base nas análises do teste DF e das FAC, FACP para verificar a estacionariedade da série, o modelo indicado é um ARIMA(1,1,1).

Nesta etapa de verificação do modelo ARIMA(1,1,1), realiza-se a avaliação por meio da análise de resíduos. O modelo proposto pela análise da FAC e PACF foi o ARIMA(1,1,1). Contudo, como pode ser verificado na Figura 5, a FACP apresenta a defasagem 8 significativa. Esse comportamento indica a presença de sazonalidade na série, pois os resíduos não se comportam como ruído branco, ou seja, não temos uma série de resíduos com média zero e variância constante.

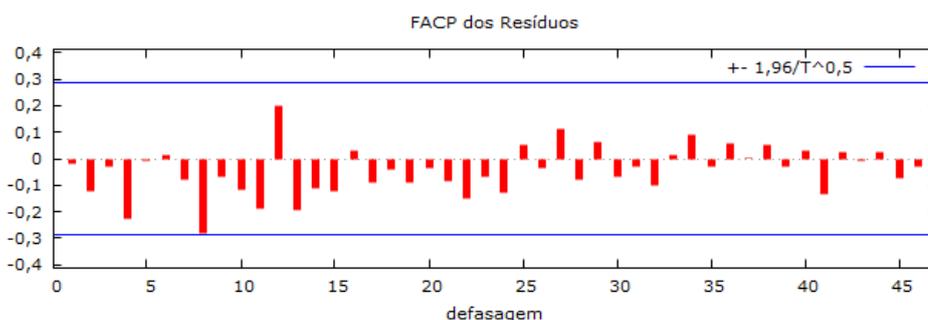


Figura 5 - Função de autocorrelação residual e função de autocorrelação parcial residual para o número de atendimentos de clientes de transporte público entre CG&SJB.
 Fonte: GRETl, versão 3.

Avaliação da ordem do modelo

Como os resíduos do modelo ARIMA(1,1,1) indicaram a presença de sazonalidade, aplica-se uma diferença sazonal, $s=12$, à série diferenciada, visando remover a influência do componente de sazonalidade.

Após esse processo, retomamos novamente o procedimento de identificação das possíveis ordens, através da análise gráfica das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série com uma diferenciação normal e uma sazonal.

A Tabela 1 apresenta alguns dos modelos analisados e os critérios que levaram à escolha do modelo SARIMA(0,1,0)(2,1,1)₁₂. Dentre os modelos sazonais abordados, apenas o SARIMA(0,1,0)(2,1,1)₁₂ e SARIMA(0,1,0)(1,1,0)₁₂ apresentam coeficientes significativos. Os demais modelos, coeficiente não-significativo e possuem AIC, MAPE, RME e U-Theil maiores que o modelo SARIMA(0,1,0)(2,1,1)₁₂ escolhido. Os resíduos do modelo SARIMA(0,1,0)(2,1,1)₁₂ não apresentam autocorrelação como mostra a Figura 6. Aplicou-se o teste de Box-Pierce nos resíduos do modelo (p -valor = 0,5595) ; logo, o teste foi não significativo, e assim, não se rejeita a hipótese de que os resíduos são estacionários, com média zero e variância constante, ou seja, apresentam um ruído branco. Conclui-se, desta forma, que o modelo se ajusta bem aos dados e, é um bom modelo estimado, sem memória tanto pelo correlograma como pelo teste Ljung-Box.

Tabela 1 - Critérios de comparação para verificar o melhor modelo

| Modelo | p-value dos coeficientes | Critério AIC | MAPE | RME | U – Theil |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|--------------|--------------|------------------|
| SARIMA(0,1,0)(1,1,0) ₁₂ | $\varphi_1 = 0,0755$ | 800,6029 (T = 35) | 12,833 | 21157 | 1,4156 |
| SARIMA(0,1,0)(1,1,1) ₁₂ | $\varphi_1 = 0,7187$ $\theta_1 = 0,6425$ | 802,3791 (T = 35) | 12,938 | 21086 | 1,4128 |
| SARIMA(0,1,0)(2,1,1)₁₂ | $\varphi_1 = 4,79e - 013$ $\varphi_2 = 0,0183$ $\theta_1 = 1,56e - 015$ | 801,6837 (T = 35) | 12,12 | 19723 | 1,3295 |
| SARIMA(0,1,0)(1,1,2) ₁₂ | $\varphi_1 = 0,8282$ $\theta_1 = 0,5093$ $\theta_2 = 0,5183$ | 804,1350 (T = 35) | 12,652 | 21006 | 1,4108 |
| SARIMA(0,1,0)(3,1,1) ₁₂ | $\varphi_1 = 0,1810$ $\varphi_2 = 0,5436$ $\varphi_3 = 0,3562$ $\theta_1 = 0,4412$ | 803,9591 (T = 35) | 12,53 | 20304 | 1,3884 |

Fonte: Gretl, versão 3.0

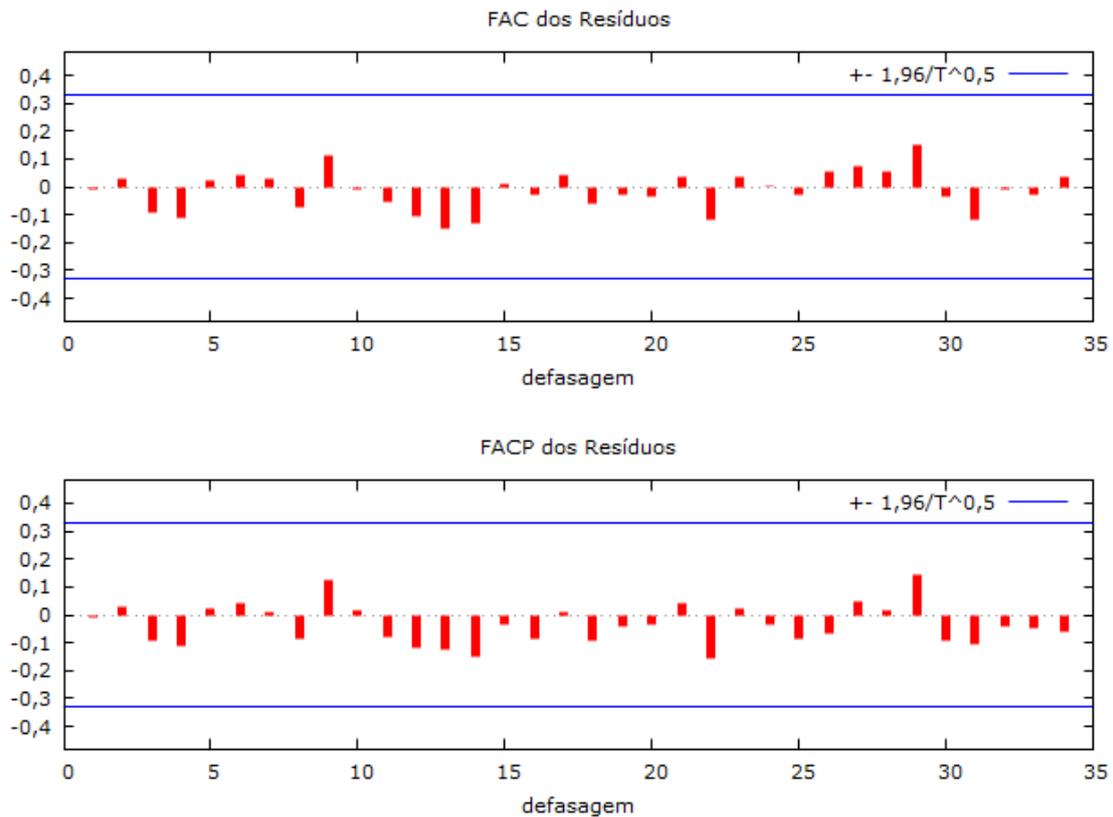


Figura 6 - Função de autocorrelação e autocorrelação parcial residual do modelo SARIMA(0,1,0)(2,1,1)₁₂
 Fonte: GRETL, versão 3.

Para verificar se os resíduos apresentam o comportamento de uma distribuição Normal, ou seja, uma distribuição com média zero e variância constante, analisa-se os resultados dos testes de Doornik-Hansen e ARCH, respectivamente. A Figura 7 mostra o histograma com o resumo do teste Doornik-Hansen, em que o p-valor=0,4823 indica a não rejeição da hipótese nula de normalidade dos resíduos. O teste ARCH com um p-valor de 0,973924 não rejeita da hipótese nula de variância constante (homocedasticidade). Portanto, os resíduos seguem o padrão de um ruído branco.

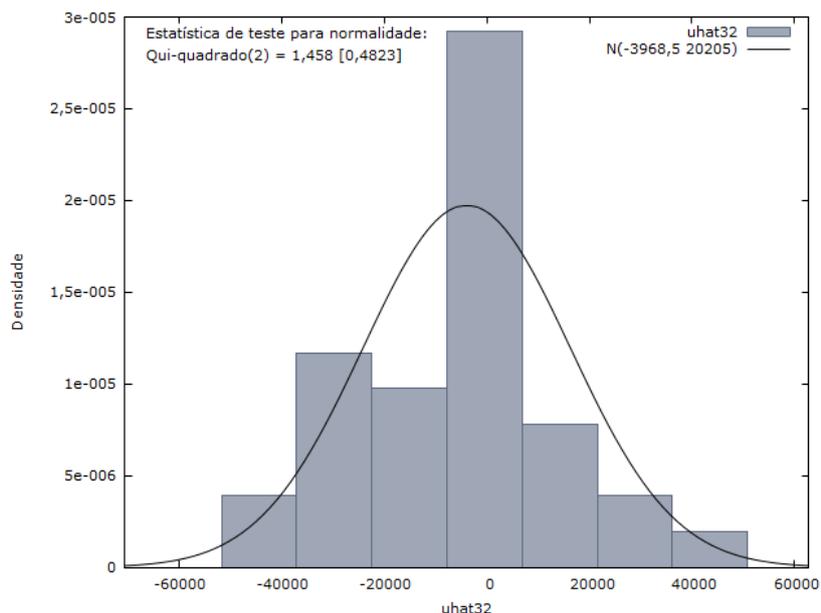


Figura 7 - Histograma e teste de normalidade dos resíduos do modelo SARIMA(0,1,0)(2,1,1)₁₂
 Fonte: GRETL, versão 3

As previsões geradas para o ano de 2014, a partir do modelo SARIMA(0,1,0)(2,1,1)₁₂, podem ser visualizadas na Figura 8. Para 2018, a previsão é de redução no número de usuários do transporte coletivo da linha Campos dos Goytacazes X São João da Barra. Verifica-se que as maiores variações ocorreram nos anos de 2015 e 2016, período este em que crise na Petrobras afetou grupos no Porto do Açú, em São João da Barra.

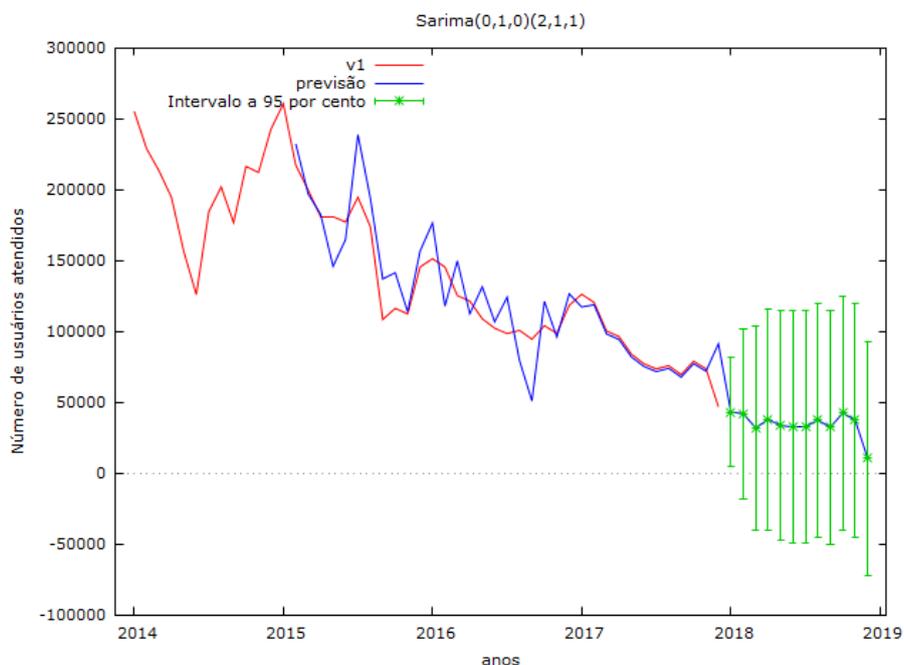


Figura 8 - Número de atendimentos de clientes de transporte público entre CG&SJB
 Fonte: GRETL, versão 3

As previsões apresentadas na tabela 2 referem-se ao período de Janeiro de 2018 a dezembro de 2018, com seus respectivos intervalos de confiança. É importante observar que séries simples (contendo apenas tendência, por exemplo) podem ser modeladas com um número relativamente pequeno de observações, mas séries complexas (contendo termos de sazonalidade com múltiplos períodos, por exemplo) podem exigir 80 ou mais observações para obter boa modelagem. No caso em estudo, foram utilizadas 48 observações, pois representam a totalidade dos dados disponibilizados pela empresa.

Tabela 2 - Critérios de comparação para verificar o melhor modelo.

| Ano: Mês | Previsão | Limite Inferior | Limite Superior |
|----------|----------|-----------------|-----------------|
| 2018:01 | 43155,31 | 4498,42 | - 81812,19 |
| 2018:02 | 41918,61 | -18311,10 | - 102148,31 |
| 2018:03 | 31734,59 | -40661,67 | - 104130,84 |
| 2018:04 | 37835,45 | -40551,49 | - 116222,40 |
| 2018:05 | 33524,37 | -47450,26 | - 114499,00 |
| 2018:06 | 32723,26 | -49224,15 | - 114670,67 |
| 2018:07 | 32749,44 | -49509,65 | - 115008,52 |
| 2018:08 | 37347,73 | -44992,63 | - 119688,09 |
| 2018:09 | 32269,88 | -50085,86 | - 114625,63 |
| 2018:10 | 42421,92 | -39935,23 | - 124779,07 |
| 2018:11 | 37292,44 | -45064,72 | - 119649,60 |
| 2018:12 | 10433,03 | -71924,48 | - 92790,53 |

Fonte: Gretl, versão 3.0.

REFERÊNCIAS

FAVA, Vera Lucia **Manual de econometria**. In: VASCONCELOS, M. A. S.; ALVES, D. São Paulo: Editora Atlas, 2000.

FERRAZ, Antonio Clóvis; TORRES, Isaac Guillermo Espinosa. **Transporte Público Urbano**. 2º ed. Editora Rima. São Carlos, 2004.

MORETTIN, Pedro Alberto; TOLOI, Clélia Maria de Castro; **Análise de Séries Temporais**. 2ª ed. São Paulo: Egard Blucher, 2006.

PELLEGRINI, Fernando Rezende; FOGLIATTO, Flávio Sanson; **Passos para Implantação de Sistemas de Previsão de Demanda – Técnicas e Estudo de Caso**. Revista Produção, v.11 n.1, Nov 2001. Disponível em: <http://www.scielo.br/pdf/prod/v11n1/v11n1a04>

VASCONCELLOS, Eduardo Alcântara de, **Políticas de transporte no Brasil: a construção da mobilidade excludente**. 1 ed. Barueri, SP: Manole, 2013.

WERNER, Liane; LEMOS, Fernando de Oliveira; DAUDT, Tiago; **Previsão de Demanda e Níveis de Estoque uma abordagem conjunta aplicada no setor Siderúrgico**. XIII SIMPEP, Bauru, nov. 2006.