

4º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS

CATEGORIA 1

O TRANSPORTE SOBRE TRILHOS NA REGIÃO METROPOLITANA DE JOÃO

PESSOA/PB: MODELAGEM ESTATÍSTICA DE FATORES QUE INFLUENCIAM A

DEMANDA DE PASSAGEIROS

INTRODUÇÃO

A compreensão a respeito das variáveis que podem afetar a demanda tem sido caracterizada como um desafio recorrente para empresas, independentemente do ramo em que atuam, tendo em vista a importância que um adequado planejamento da capacidade (e oferta) de bens e serviços têm para a sustentabilidade dos negócios. O transporte de passageiros, embora ainda modestamente comentado em termos de previsões estatísticas da demanda, não é excluído desse contexto de análise da demanda.

Trabalhos analíticos, especialmente os estatísticos, assumem não raramente que as observações de um dado fenômeno guardam entre si informações relevantes para compreensão mais detalhada desse fenômeno. Através da estimação de modelos, considerando-se premissas matemáticas, uma série de interpretações e análise podem ser feitas, inclusive no tocante à demanda e a relevância dessa para o planejamento e a tomada de decisões gerenciais.

No contexto metroferroviário, de forma semelhante ao que acontece em outros modais, o transporte de passageiros mostra-se como um serviço público essencial na realidade brasileira. Importante comentar, ainda, que as condições sociais e econômicas, bem como o planejamento urbano, colocam o transporte sobre trilhos como uma via alternativa.

Diante de tudo que foi exposto, o presente trabalho buscou, defender a importância de se modelar (e, por conseguinte, conhecer mais afincado) a demanda de transporte de passageiros. Supondo-se, nesse caso, que se trata de condição essencial para processos de planejamento mais robustos, face às vistas diversas atividades e ações de uma empresa metroferroviária. Além da defesa em torno da necessidade de adoção de métodos quantitativos, objeto de estudo tem a intenção de avançar de forma bastante prática sobre a gestão de transportes, à medida em que busca gerar resultados a respeito da eficiência do transporte, através de um processo analítico que envolve variáveis por vezes pouco consideradas nas decisões organizacionais.

Do ponto de vista da problemática, o presente trabalho busca responder a seguinte questão: Até que ponto fatores internos e externos a um sistema metroferroviário podem

influenciar a demanda de passageiros? Tomando como referência uma sequência de testes estatísticos e diagnósticos, o presente trabalho almeja a investigação, sob uma perspectiva quantitativa, a respeito de fatores como acidentes, existência de campanhas de marketing, notícias, investimentos etc. podem influenciar, de maneira significativa, de demanda de passageiros transportados na grande João Pessoa. As séries utilizadas, nesse sentido, se referiram ao quantitativo de passageiros transportados de 2013 a 2016 e séries das variáveis independentes supracitadas. Do ponto de vista dos objetivos, o objetivo geral foi modelar a demanda, através das teorias estatísticas da correlação, regressão linear e séries temporais, de forma que resultados relevantes pudessem ser efetivamente estudados. No tocante aos objetivos específicos, de forma resumida, foram definidos: i) realizar um diagnóstico da demanda de passageiros transportados pelo sistema de João Pessoa, por meio de técnicas estatísticas; ii) testar o ajustamento de modelos de correlação, regressão e VAR na compreensão dos efeitos influenciadores da demanda de passageiros transportados; e iii) verificar quais os fatores mais relevantes, estatisticamente.

DIAGNÓSTICO

O diagnóstico do presente estudo está dividido em três seções. Primeiro, uma breve revisão da literatura, abordando especificamente discussões sobre a modelagem da demanda em operações de transporte, aspectos teóricos da análise estatística através de modelos de regressão e correlação; e aspectos conceituais relativos à análise por meio de séries temporais. Na sequência, a segunda seção abordou a caracterização do sistema

metroferroviário de João Pessoa, através de informações qualitativas e quantitativas referentes à demanda de passageiros. Por fim, na terceira seção, são apresentados os métodos e procedimentos adotados na pesquisa.

1. REVISÃO DA LITERATURA

1.1 Modelagem da demanda

A demanda pode ser compreendida, conforme explicam Lustosa et al. (2008, p. 50) como "(...) a disposição dos clientes ao consumo de bens e serviços ofertados por uma organização". Os autores explicam, ainda, que essa demanda é influenciada por uma série de fatores que, por sua vez, podem ser definidos por questões externas ou internas da organização, a exemplo de aspectos macroeconômicos e mesmo condições operacionais.

A modelagem da demanda, para Moreira (2011), tem uma relação muito próxima do processo de planejamento, especificamente nos processos de previsão. O autor argumenta que a previsão da demanda é inerente ao planejamento e, da mesma maneira, por se tratar de um processo racional de busca por informações sobre o comportamento futuro, acaba influenciado a dinâmica da tomada de decisão nas organizações.

A demanda é conceituada, em geral, em dois tipos: a demanda pontual e demanda repetitiva (LUSTOSA et al, 2008). A demanda pontual é caracterizada por uma situação denominada "pico da demanda", fator que ocorre de forma concentrada na janela temporal e, algum tempo depois, diminui de forma estatisticamente significativa ou desaparece. Nesse tipo de demanda duração de um dado evento, por vezes, é capaz influenciar de forma

bastante característica o comportamento da série, praticamente sendo o principal motivo do aumento da demanda. O segundo tipo de demanda normalmente conceituada, ainda de acordo com Lustosa et al (2008), a repetitiva, pode ser classificada em dependente ou independente. Na dependente, o que se verifica é a associação facilmente identificada a uma outra demanda ou série, como exemplo pode-se citar a relação entre a demanda de passageiros de ônibus públicos é dependente da demanda geral por transportes públicos. No caso da independente, por sua vez, fatores como a tendência, estacionariedade e sazonalidade, aspectos geralmente relacionados a modelos de séries temporais.

A análise da demanda, para fins de previsões e geração de informações para tomada de decisões, depende de algumas características essenciais. Segundo Moreira (2011), em quaisquer circunstâncias, alguns pontos são constantes, tais como a disponibilidade de dados, tempo e recursos; e o horizonte de previsão ou corte temporal para análise. Importante comentar, ainda de acordo com o autor, que algumas premissas devem ser consideradas, quais sejam: i) a presunção de que as mesmas causas que estiveram presentes no passado, gerando a demanda, continuarão presentes no futuro, servindo assim de base para se inferir sobre comportamento futuro; e ii) a existência de erros, seja por fatores aleatórios ou atributos que não puderam ser captados pelo método de análise.

Segundo Brown (1959), a demanda é considerando-se que a relação complexa entre diversos fatores pode ser compreendida a partir do efeito de cada fator sobre essa demanda, inclusive a interação entre os mesmos. Para o autor é plausível, nesse sentido, que a construção de modelos matemáticos ou estatísticos apresentem estimativas confiáveis sobre

os fatores geradores da demanda. O uso de um modelo satisfatório, complementa o autor, é capaz de permitir a visualização do mecanismo de interação dos fatores junto à série.

De forma esquematizada, alguns métodos de análise da demanda são apresentados na Figura 1:

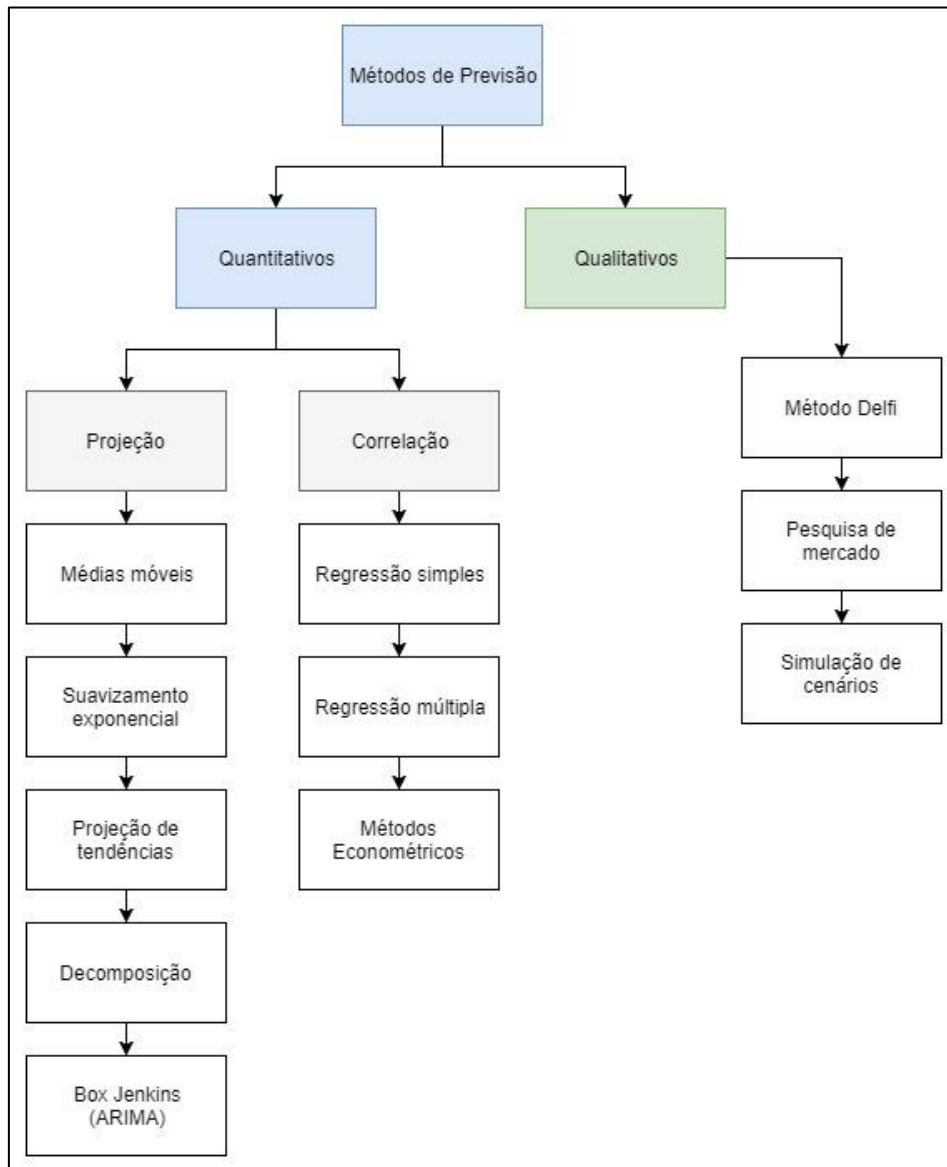


Figura 1. Alguns métodos de análise da demanda.

Fonte: Adaptado de Lustosa et al (2008).

Salienta-se, a respeito dos métodos comumente utilizados para análises relacionadas à demanda, que no caso dos modelos quantitativos, a base conceitual é a utilização de técnicas estatísticas. No caso da adoção de métodos qualitativos, por sua vez, o padrão é a utilização de consensos e opiniões (MOREIRA, 2011; LUSTOSA et al, 2008).

A respeito da análise da demanda, cabe comentar que a mesma é considerada uma temática bastante relevante para o setor de transportes, tendo em vista a sua abrangência teórica e prática. Para Murça e Muller (2014), a análise da demanda no setor de transportes mostra-se como de fundamental importância para os processos de planejamento, principalmente aqueles destinados à garantia da mobilidade e do desenvolvimento do sistema, abarcando também questões relacionadas à alocação de recursos para investimentos de forma eficiente.

Ainda sobre a importância da análise da demanda, de acordo com Serra, Tavares e Santos (2005), deve-se reconhecer a importância desses estudos para que os gestores possam realizar o planejamento. Nessa perspectiva, cada vez mais vêm sendo adotadas técnicas quantitativas que possam compreender com maior robustez os mecanismos formadores da série histórica. Os autores comentam, ainda, que métodos baseados em modelos econométricos, inclusive de análises das séries temporais, têm se mostrado bastante adequados para a modelagem da demanda, a fim de se gerarem informações significativas para que as empresas lidem com a sua demanda.

1.2 Regressão Linear e Correlação

As análises estatísticas que envolvem o estudo de duas ou mais variáveis podem ser feitas através de técnicas de regressão e correlação (BRUNI, 2011). Originalmente, a diferença entre a adoção de modelos matemáticos e modelos estatísticos - sendo esses últimos utilizados na modelagem de regressões e correlações - reside na atribuição de um erro estatístico, que conduz a determinação de um modelo melhor ajustado aos pontos e testes de significância que possam determinar a robustez do modelo. De acordo com Freund (2006), o problema básico a ser respondido, com modelos de correlação ou regressão, é a determinação de quão bom é o ajuste de uma tal reta aos dados.

Diante do exposto, percebe-se que a forma mais simples de se verificar se duas variáveis estão associadas, de acordo com Field (2009), é através da análise da forma como variam conjuntamente. Essa maneira mais simples, por sua vez, se refere exatamente ao conceito de correlação, que de certa forma é preliminar ao de regressão embora compartilhem entre si aspectos comuns nos procedimentos matemáticos. O conceito chave, de acordo com o autor, é entender a ideia de variância, mas de forma que as variáveis em questão sejam associadas, com o propósito de se observar as diferenças entre dos dados observados e as médias das duas variáveis. Matematicamente, a correlação pode ser encontrada a partir da aplicação da Equação 1:

$$(1) \quad \rho = \frac{\sum_{i=1}^n (x_i - X) \cdot (y_i - Y)}{\sqrt{\sum_{i=1}^n (x_i - X)^2} \cdot \sqrt{\sum_{i=1}^n (y_i - Y)^2}} = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$$

Na equação 1 deve-se entender que os valores medidos para x e y são referentes às variáveis que se busca encontrar a correlação, no presente caso, se tratando da modelagem pelo coeficiente de Perason. Essas variáveis, por meio da covariância com os parâmetros das médias de X e Y , por sua vez, permitem a verificação da linearidade da associação entre os conjuntos, gerando valores que variam de -1 a 1, sendo a direção da associação e a força da associação os principais aspectos que merecem observação, após o ajuste do conjunto de dados ao modelo apropriado.

Field (2009), explica que pelo menos três coeficientes de correlação são amplamente usados nas pesquisas, a fim de medir a associação entre um conjunto de variáveis (por vezes, inclusive, com a representação gráfica através de uma matriz de correlação), sendo esses modelos: i) o de Pearson; ii) o de Spearman; e iii) o de Kendall.

A correlação de Pearson é a mais comumente utilizada para análise de associações, sob o ponto de vista estatístico, tratando-se, inclusive, de uma formulação assumidamente paramétrica. O modelo mede a relação, sob o pressuposto da linearidade e da normalidade do conjunto amostral. Além desses pressupostos, o modelo considera ainda que existe homocedasticidade, ou seja, variância homogênea na dispersão dos dados perante a regressão de um modelo.

O coeficiente de correlação dos postos de Spearman, diferente da correlação de Pearson, se trata de um modelo não-paramétrico, usado para mensurar o grau de associação entre duas variáveis sem, contudo, assumir suposições sobre a distribuição da amostra. De acordo com Field (2009), o modelo é interessante para situações em que os dados violem, de alguma forma, suposições paramétricas, a exemplo de não-normalidade da distribuição.

De forma resumida, o coeficiente é construído através da ordenação dos dados, para em seguida aplicar-se o coeficiente de Pearson.

O Tau de Kendall, ou simplesmente correlação de Kendall, assim como o modelo de Spearman, trata-se de uma estatística não-paramétrica e por sua vez é indicado para utilização em um conjunto de dados pequeno ou, quando a existência de postos ordenados enseja uma grande quantidade de situações de empate. Na prática, é adotado quando o pesquisador verifica que após a ordenação dos escores, muitos dados apresentam o mesmo posto (FIELD, 2009). Cabe comentar que, de acordo com Howell (1997), embora a popularidade do coeficiente de Spearman seja maior, há de se comentar que a robustez do modelo de Kendall, de forma pragmática, pode levar a melhores estimativas da correlação entre as variáveis, dessa maneira, permitindo generalizações mais precisas do que os modelos de Spearman e Pearson.

De acordo com Anderson, Sweeney e Williams (2011), ao se considerar que decisões e análises gerenciais frequentemente se baseiam na relação entre duas variáveis (note-se, nesse caso, que não se assume como suficiente a associação), por vezes modelos de regressão são mais adequados para a obtenção de resultados mais confiáveis e, por conseguinte, mais úteis.

Na terminologia básica da análise da regressão linear são considerados, inicialmente, que uma variável que é prevista é dita como uma variável dependente. Na mesma linha de conceituação, uma ou mais variáveis utilizadas para previsão da variável dependente são denominadas variáveis independentes (ANDERSON; SWEENEY; WILLIAMS, 2011). Importante conceituar que, matematicamente, um modelo de regressão linear considera, em geral, que

a Y designa a variável dependente e X a variável independente. A equação 2 apresenta a generalização do modelo:

$$(2) \quad y = \beta_0 + \beta_i X + \varepsilon$$

Os betas da equação são considerados os parâmetros do modelo, enquanto ε se trata do erro, considerado como uma variável aleatória. Esse erro, por sua vez, é o responsável pela variabilidade em Y que não pode ser explicada pelo relacionamento linear almejado com a consideração das variáveis independentes X a Xn.

Na prática, os parâmetros amostrais da regressão são estimados, considerando-se o conjunto de dados disponíveis sobre a amostra. Essa ausência de parâmetros, por sua vez, enseja a adoção de um procedimento matemático capaz de gerar os estimadores mais adequados, frente pressupostos estabelecidos e aspectos estatísticos disponíveis. De acordo com Sweeney e Williams (2011), o método dos mínimos quadrados, ou no inglês Ordinary Least Squares (OLS) se trata de um procedimento bastante usado em dados amostrais para que se estimar uma regressão e, igualmente, os parâmetros do modelo.

1.3 Séries Temporais

Uma série temporal, conceitualmente, trata-se de uma série histórica, que por sua vez foi obtida através de dados organizados por meio de intervalos de tempo considerados regulares em um período específico (LATORRE; CARDOSO, 2001). Caso essas observações sem consecutivas e dependentes de outras, torna-se possível a previsão e o fornecimento de

bases para compreender o comportamento do evento ao qual se desenvolvem as análises (SERRA; TAVARES; SANTOS, 2005).

De forma geral, o propósito da análise das séries temporais é o estudo da forma como a estrutura temporal e a dinâmica de interação das variáveis, em um conjunto de dados, considerando as correlações seriais dos resíduos, a fim de que se compreendam os elementos aleatórios e independentes. Segundo Morettin e Toloí (2004), a análise de séries temporais pode ajudar a: i) investigar o mecanismo gerador da séries temporal; ii) realizar previsões de valores futuros da série temporal; iii) descrever o comportamento da séries, em termos de tendências e sazonalidades; e, por fim, iv) investigar periodicidades relevantes nos dados.

Segundo Latorre e Cardoso (2001), os componentes fundamentais de uma série histórica e, por conseguinte, pertinentes à análise de séries temporais podem ser definidos em torno de três aspectos não observáveis: a tendência (T), a sazonalidade (S) e a variação aleatória, conceitualmente definida como ruído branco (a). Em linhas gerais, duas formas de elaboração do modelo podem ser utilizadas, uma versão aditiva ($Z = T + S + a$) ou uma versão multiplicativa ($Z = T \cdot S \cdot a$)

É necessário, ao se estudar séries temporais, que se reconheça que, a exemplo dos métodos tradicionalmente utilizados na econometria/estatística, um modelo de séries temporais pode ser paramétrico ou não paramétrico. Segundo Morettin e Toloí (2004), um modelo paramétrico tem sua análise feita no domínio do tempo, enquanto um modelo, por sua vez, parte da necessidade de se analisar a frequência. Além da definição da abordagem paramétrica do modelo, conforme explicam Serra, Tavares e Santos (2005), verifica-se na literatura estatística, que se sugere que os modelos devem ser simples, com o menor número

12

de parâmetros possíveis e sua utilização deve ser de fácil compreensão e descrição para os usuários interessados nas informações.

Uma modelagem bastante utilizada em séries temporais, e que por sua vez têm uma operacionalização bastante simples (e capaz de gerar resultados interessantes), segundo Cavalcanti (2010), são os modelos tidos como autoregressivos vetoriais ou, simplesmente, VAR. A popularidade desses modelos, segundo o autor, deriva em grande parte de percepção de que tais modelos permitem análises robustas de inter-relações entre múltiplas variáveis a partir de um conjunto razoavelmente mínimo de restrições.

Considerado como uma extensão da regressão univariada, mas considerado a existência de múltiplas variáveis, um VAR é uma regressão produzida por OLS e pode ser expresso, de forma genérica, conforme apresentado na equação 3:

$$(3) \quad X_t = A_0 + A_1X_{t-1} + \dots A_pX_{t-p} + B_0Z_1 + B_0Z_1 + B_1Z_{t-1} \dots + B_pZ_{t-p} + e_t$$

De forma geral, as matrizes de A e B são matrizes do tipo $n \times x$, de coeficientes e valores que são defasados em relação às variáveis endógenas e exógenas. Para fins de escolha do melhor modelo VAR, por sua vez, comumente utilizam-se os critérios que permitam parcimônia nas análises.

Os modelos VAR, segundo Cavalcanti (2010), de forma reduzida resumem as correlações observadas entre variáveis de interesse e podem ser estimados sem grandes problemas paramétricos. Comenta-se, por fim, que no contexto dos modelos VAR, além do principal objetivo de se gerar previsões, o que se busca é a minimização do erro quadrático

médio, de forma que os estimadores da equação sejam não viciados e, por sua vez, a variância nas previsões seja a mínima possível.

2. O SISTEMA METROFERROVIÁRIO DE JOÃO PESSOA

O sistema metroferroviário de João Pessoa encontra-se situado no litoral do Estado da Paraíba. O sistema possui uma via permanente singela, considerada não segregada, com bitola métrica e dormentes de concreto, totalizando uma extensão de cerca de 30 km. O sistema é operado pela Cia Brasileira de Trens Urbanos (CBTU), através da Superintendência de Trens Urbanos de João Pessoa, e atende os municípios de Santa Rita, Bayeux, João Pessoa e Cabedelo, com 12 estações, sendo duas em Santa Rita, uma em Bayeux, quatro em João Pessoa e cinco em Cabedelo. A Figura 2 apresenta o mapa, com estações e o trecho operado pela CBTU em João Pessoa/PB:

de 2016 a empresa transportou cerca de 1,7 milhões de passageiros, com uma média de 9,41 passageiros por km e uma receita total em torno de 900 milhões de reais.

3. METODOLOGIA

Do ponto de vista metodológico o presente trabalho pode ser considerado: quanto à natureza, como uma pesquisa aplicada; quanto aos objetivos, uma pesquisa exploratória; quanto aos procedimentos um estudo de caso (PRODANOV; FREITAS, 2013). Quanto à abordagem, conforme entendimento de Viana (2001), trata-se de um trabalho predominantemente quantitativo.

A fim de atingir os objetivos definidos para a pesquisa, as técnicas adotadas foram balizadas por modelos estatísticos, tais como correlações e regressões. Preliminarmente, com a intenção de se explorar a amostra, o ponto inicial para geração de resultados consistiu em uma análise exploratória dos dados (AED), comumente utilizada da estatística para apresentação, de forma claramente descritiva, de informações iniciais sobre a amostra e o fenômeno estudado. Na sequência foram modelados, através de correlações e regressões lineares (simples e múltiplas), os fatores que se buscou averiguar na proposta de trabalho. Por fim, de posse dos outputs dos modelos, as interpretações foram conduzidas, sendo sintetizadas ao final dos resultados.

O Quadro 1 apresenta as principais estatísticas e modelos utilizados no trabalho:

Quadro 1 - Principais estatísticas e modelos utilizados no trabalho.

Categoria e/ou tipo	Estatísticas
Análise Exploratória de Dados (AED)	Média, Mediana, Variância, Desvio-padrão, etc.
Regressão Linear	Simples e Múltipla, Ordinary Least Squares (OLS)
Correlação	Coeficientes de Pearson, Spearman e Kendall
Time series / Forecasting	Vetor autoregressivo (VAR)

Cabe salientar que, ao longo da geração de resultados, os testes adicionais necessários para verificação da robustez dos modelos também são conduzidos, a exemplo da verificação da significância através do teste F, critérios Akaike, significância do R^2 ajustado, Fitness, testes de normalidade, a exemplo de Breuch-pagan entre outros, a depender do tipo de modelagem. De forma semelhante, ao final de cada estimação são discutidos os principais resultados da modelagem, bem como as decorrências dos parâmetros e testes dos modelos.

Para composição dos modelos, foram utilizados dados diários do transporte de passageiros no período de janeiro de 2013 a dezembro de 2016. Considerando-se que existe o transporte normal de segunda à sexta-feira e que, nos sábados, a grade é específica, optou-se pela generalização com a exclusão dos sábados. A escolha do corte dos sábados deveu-se à verificação prévia de que a série temporal não respeitaria o mecanismo gerador originalmente existente nos dias úteis normais, tendo em vista que o transporte no sábado refere-se a uma espécie de demanda atípica, exclusiva para aquele dia da semana. Além da utilização da quantidade de passageiros transportados, as regressões consideraram ainda as seguintes variáveis: dia da semana, ocorrência de acidente, evento externo com repercussão

na CBTU e evento interno (campanhas, ações sociais etc.). As variáveis em questão, tidas como independentes, também se destrincharam em uma variável nominal, tipo dummy, utilizada para testar o efeito geral dos eventos (se positivos ou negativos). O Quadro 2 apresenta as variáveis adotadas no estudo:

Quadro 2 - Variáveis e modelo geral do estudo.

Modelo Geral:	$P_{dia} = \alpha + D_i + Ex_i + \varepsilon$ Onde, P é a quantidade de passageiros transportados no dia; α é a constante do modelo; D é o dia da semana, utilizado como variável de controle; Ex é o tipo de evento, podendo ser Ec, Ea, Ee, etc.; e ε é o erro padrão da estimativa.
Variáveis independentes	
Dia da semana	Dias da semana, considerados de segunda a sexta.
Evento Interno	Campanhas, ações sociais, investimentos ou notícias/conteúdos criados pela CBTU. A variável assume as hipóteses de E ("c" para campanhas; e "n" para notícias ou conteúdo).
Evento Externo	"Ee". Refere-se a todo ou qualquer evento, originado externamente (paralisação, crimes nas proximidades etc.) e que podem impactar na demanda.
Acidente	"Ea" Evento tipo acidente que pode acontecer no trecho ou nas estações. Refere-se à situações em que algum acidente repercutiu na realidade da empresa, tendo em vista a menção do nome da CBTU no evento.

No tocante às ferramentas utilizadas para tratamento dos dados, comenta-se que para organização prévia dos dados foram utilizadas planilhas eletrônicas. Para as análises estatísticas, por sua vez, optou-se pela utilização do STATA, por ser considerado um pacote estatístico abrangente, uma vez que traz testes prévios às análises, posteriores e, sobretudo,

modelos de séries temporais. A respeito da série, importante comentar que foi construída com 1218 observações, referentes aos dias de segunda a sexta, excluídos feriados, no período de janeiro de 2013 a dezembro de 2016.

ANÁLISE DOS RESULTADOS

Os resultados do presente trabalho são apresentados em três tópicos, quais sejam: A) Análise Exploratória de dos Dados; B) Análise das Correlações e da Regressão Linear; e C) Modelagem de Séries Temporais.

A) Análise exploratória dos dados

Com o propósito de se conhecer, preliminarmente, os dados referentes à demanda e as variáveis estudadas no presente trabalho, foi realizada uma breve análise exploratória dos dados a série da quantidade de passageiros transportados, conjuntamente com os eventos estudados, quais sejam: campanhas (ou eventos internos promovidos em prol dos passageiros, em quantidade de dias); eventos de origem externa, a exemplo de campanhas de saúde, segurança etc., que chegaram até a CBTU; acidentes (ou situações envolvendo crimes, situações de risco etc., ocorridas no trecho operado pela CBTU ou envolvendo passageiros da CBTU); e por último, notícias publicadas (seja pela CBTU ou órgãos relacionados), envolvendo algum anúncio positivo em relação à empresa, tais como investimentos, reformas, recursos, soluções de problemas etc.

Com o propósito de se conhecer, graficamente, a série de dados, a Figura 2 apresenta o comportamento da variável quantidade de passageiros, com frequência diária:

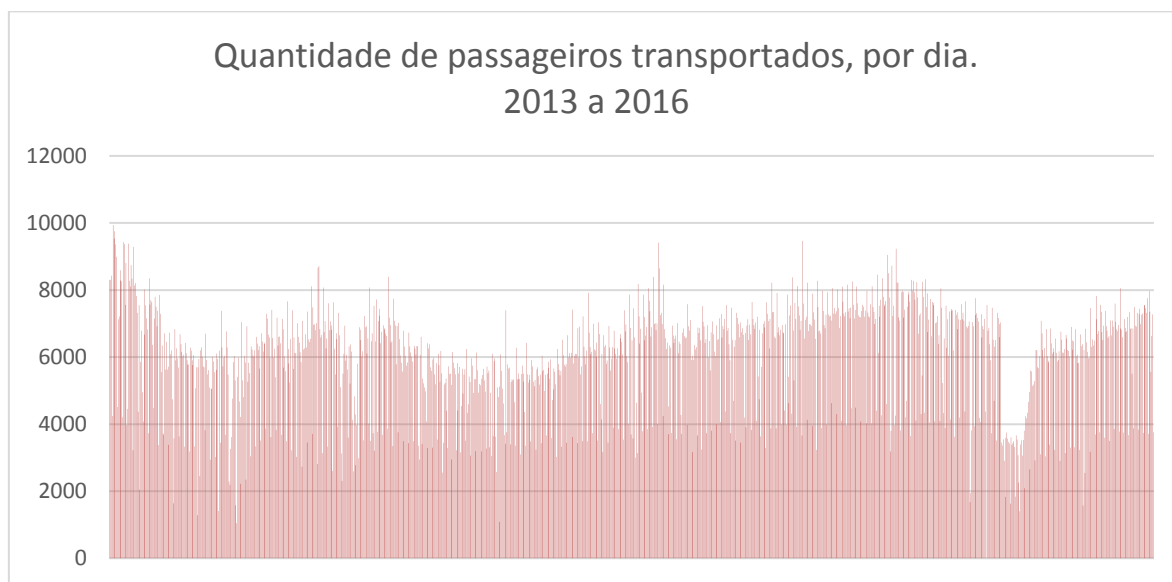


Figura 2 - Gráfico da quantidade de passageiros transportados, diariamente, ao longo do período.

O gráfico, a princípio, permite que se percebam picos no transporte de passageiros (de forma pontua e, em alguns casos, de forma mais densa) e, da mesma maneira, quedas súbitas no transporte (quando analisados pontos do gráfico em que a demanda cai de maneira vertiginosa). Interessante comentar, inclusive, que em quedas mais abruptas e densas (no final do período, por exemplo), parecem ter relação com eventos ou situações exógenas que aparentemente podem ter influenciado a demanda. Em alguns casos, por exemplo, ao se analisar com maior detalhamento a série, pode-se verificar que existem coincidências de momentos de queda da demanda com, por exemplo, um acidente metroferroviário.

Embora o gráfico plotado permita algumas considerações a respeito da série e, de forma rústica, algumas interpretações, há de se colocar em discussão que o simples conhecimento do comportamento da série a partir de uma análise gráfica é limitado e por vezes pode levar a vieses de análises. A fim de minimizar considerações equivocadas, as seções seguintes apresentam informações relativas a testes estatísticos e análises exploratórias.

A Tabela 1 apresenta alguns dados, referentes à análise exploratória:

Tabela 1 - Análise Exploratória dos Dados.

Passageiros transportados (por dia)			
Ano/Período	Média	Mediana	Desvio-padrão
2013	5865,84	6152,50	1707,40
2014	5405,38	5645,00	1223,82
2015	6455,93	6876,00	1407,31
2016	5898,89	6538,00	1725,15
Período (2013 a 2016)	5905,42	6240,50	1573,87
Quantidade média de eventos por ano (em número de dias)			
Campanhas (ações sociais, eventos de marketing etc.)			15,75
Eventos Externos (ações municipais, estaduais, instituições etc.)			13,25
Acidentes (e situações de risco, colisões, violência etc.)			14,00
Notícias (publicações da CBTU, matérias, anúncios de investimentos etc.)			13,00
Duração do efeito de um evento sobre a demanda			
Campanhas (ações sociais, eventos de marketing etc.)			2,3 dias (+)
Eventos Externos (ações municipais, estaduais, instituições etc.)			1,1 dias (+)
Acidentes (e situações de risco, colisões, violência etc.)			2,8 dias (-)
Notícias (publicações da CBTU, matérias, anúncios de investimentos etc.)			1,9 dias (+ ou -)

A respeito das estatísticas geradas, interessante comentar que a média de passageiros transportados caiu de 2013 para 2014, recuperando-se em 2015 e novamente voltando ao patamar equivalente ao de 2014 apenas no ano de 2016. Uma análise por meio da medida de

tendência central, por sua vez, parece insuficiente para algumas considerações. A fim de facilitar a interpretação, através da análise da mediana dos dados juntamente com o desvio padrão, pode-se perceber que a série de dados parece ter sofrido choques, em determinados momentos, que levaram a variações de quase 30% na quantidade de passageiros transportados.

Outra informação relevante, condicionada por sua vez às suposições de que variáveis independentes - que ainda serão discutidas com detalhamento, diz respeito aos eventos definidos como relevantes para série (teoricamente), aconteceram ao longo do período. De acordo com os dados obtidos, em média entre 13 e 15,75 dias eventos aconteceram por ano no período considerado. Importante comentar que não se trata de quantidade relativa de eventos, como acidentes ou campanhas de marketing, mas, na verdade, de dias em que situações enquadradas, de forma genérica nas referidas categorias, aconteceram. Assim, não se fala de 14 acidentes, mas, na verdade, de 14 situações que podem ter sido acidentes com passageiros, crimes de maior repercussão, colisões, etc.

Além de se conhecer a quantidade média de eventos supostamente influenciadores da série, com intenção de se apresentar dados mais facilmente interpretáveis, uma das fases da análise exploratória consistiu em se verificar, em média, por quantos dias os efeitos desses eventos parecem influenciar a série. Assim, campanhas (ações sociais, campanhas de saúde, marketing etc.) foram verificadas com um fator capaz de influenciar a série, de forma linear e discreta, por cerca de 2,3 dias (de forma positiva), por outro lado acidentes (situações envolvendo colisões, crimes, etc.) tendem a influenciar negativamente a demanda por cerca de 2,8 dias. Além dessas, um fato interessante é que notícias (anúncios de investimentos,

notícias publicadas pela CBTU, reportagens etc.) podem ter um efeito dual: se, por um lado, podem influenciar positivamente a demanda, por outro também podem influenciar negativamente, sugerindo-se, portanto, que o teor das notícias pode ser um indicador relevante na estimação do efeito.

B) Correlações e Regressão Linear

A fim de se conhecer as associações de fatores relacionados à demanda de passageiros, considerando-se o período de 2013 a 2016, fez-se uso de três modelos de correlação. No primeiro momento, o coeficiente de Pearson, que é um modelo paramétrico. No segundo momento, a correlação de Spearman, que relaxa suposições relativas à normalidade. Por fim, e talvez mais importante, estimou-se a correlação das variáveis através do modelo de Kendall, que, conforme sinaliza a teoria, pode ser considerado o mais adequado, principalmente enquanto proposta não-paramétrica. A Tabela 2 apresenta a Correlação de Pearson:

Tabela 2 - Correlação de Pearson.

	quantidade	dia	ecampanha	eexterno	eacidente	enotcia
quantidade	1.0000					
dia	-0.5375 0.0000	1.0000				
ecampanha	0.0695 0.0152	-0.0301 0.2941	1.0000			
eexterno	-0.1533 0.0000	0.0634 0.0269	-0.0135 0.6386	1.0000		
eacidente	-0.2343 0.0000	0.0606 0.0346	-0.0336 0.2417	0.0108 0.7059	1.0000	
enotcia	0.1311 0.0000	-0.0623 0.0298	-0.0310 0.2799	-0.0450 0.1161	-0.0464 0.1058	1.0000

Por meio da análise da correlação de Pearson pode-se, agora a partir de uma perspectiva estatística que inclui a significância (por meio de teste de hipótese), analisar a associação de fato entre as variáveis definidas para o estudo. Interessante comentar que, por se tratar tão somente de associação, não se pode inferir a respeito da causalidade, contudo, pode-se conhecer a forma como as variáveis, duas a duas, caminham ao longo da série de dados. Tomando como ponto de partida a variável quantidade de passageiros, foco do presente estudo e definida como variável dependente no modelo geral apresentado na metodologia, percebe-se pela Tabela 2, que as variáveis campanha, evento externo, evento acidente e evento notícia são todos significantes a 5%. Comenta-se, ainda, que a variável dia, originalmente colocada como de controle, apresentou o maior poder explicativo da associação, inclusive com correlação negativa junto à quantidade de passageiros. Importa explicar, nesse sentido, que por se tratar de uma variável ordinal, na verdade o que se tem é

que à medida que os dias se aproximam da sexta-feira, a quantidade média de passageiros parece cair. Igualmente, no caso do início da semana, o que se tem é uma provável associação com o aumento da quantidade de passageiros transportados.

Semelhantemente ao que foi feito através da correlação de Pearson, que é paramétrica, optou-se também pelo modelo de Spearman, não paramétrico, a fim de que pressupostos de normalidade possam ser razoavelmente relaxados. Apresenta-se na Tabela 3 o modelo:

A Tabela 3 - Correlação de Spearman.

	quantidade	dia	ecampanha	eexterno	ecidente	enotcia
quantidade	1.0000 1218					
dia	-0.5181 1218 0.0000	1.0000 1218				
ecampanha	0.0769 1218 0.0073	-0.0295 1218 0.3029	1.0000 1218			
eexterno	-0.1188 1218 0.0000	0.0639 1218 0.0258	-0.0135 1218 0.6386	1.0000 1218		
ecidente	-0.2034 1218 0.0000	0.0610 1218 0.0332	-0.0336 1218 0.2417	0.0108 1218 0.7059	1.0000 1218	
enotcia	0.1404 1218 0.0000	-0.0628 1218 0.0283	-0.0310 1218 0.2799	-0.0450 1218 0.1161	-0.0464 1218 0.1058	1.0000 1218

A partir da análise da matriz de Spearman, observa-se que, a partir do corte por meio do erro fixado (5%) para significância, novamente todas as variáveis são significantes para explicação das associações. Destaca-se, nesse sentido, que novamente as variáveis evento

campanha e evento notícia apresentam uma associação positiva (à medida que acontecem, aparentemente o número de passageiros sobe) e os eventos externos e acidentes, por sua vez, tem uma carga associativa negativa, permitindo-se, dessa forma, a inferência de que à medida que estes acontecem, supostamente tem-se uma variação negativa no transporte.

Dando continuidade às análises, a fim de se ter uma maior precisão na modelagem, foi realizada a análise pelo Tau de Kendall. A Tabela 4 apresenta a correlação de Kendall:

Tabela 4 - Correlação através do Tau de Kendall

	quantidade	dia	ecampanha	eexterno	eacidente	enotcia
quantidade	0.9998 1218					
dia	-0.3581 1218 0.0000	0.8351 1218				
ecampanha	0.0197 1218 0.0073	-0.0075 1218 0.3028	0.0982 1218			
eexterno	-0.0280 1218 0.0000	0.0149 1218 0.0258	-0.0012 1218 0.6387	0.0833 1218		
eacidente	-0.0492 1218 0.0000	0.0146 1218 0.0333	-0.0031 1218 0.2418	0.0009 1218 0.7062	0.0878 1218	
enotcia	0.0328 1218 0.0000	-0.0145 1218 0.0284	-0.0028 1218 0.2800	-0.0037 1218 0.1162	-0.0039 1218 0.1059	0.0818 1218

Considerado, teoricamente, o modelo de correlação que produz resultados mais robustos (embora ainda seja pouco utilizado, frente aos tradicionais Pearson e Spearman), o Tau de Kendall produziu resultados que ajudaram a confirmar o que já se imaginava: as

associações das variáveis evento notícia e campanha, tem sinal positivo (junto à quantidade transportada) enquanto as variáveis evento externo e acidente são capazes de influenciar negativamente a série.

Tendo em vista que a análise por meio de correlações encontra seus limites na capacidade preditiva e na explicação da causalidade, optou-se, na sequência, pela modelagem através de regressão linear. Fixando-se, dessa forma, como variável dependente a quantidade de passageiros, e testando-se a significância das variáveis campanha, evento externo, acidente e notícia, gerou-se um modelo simples, mas útil às análises que se propõem no presente estudo. A modelagem por meio de Regressão Linear é apresentada na Tabela 5:

Tabela 5 - Análise da Regressão Linear

Source	SS	df	MS	Number of obs = 1218		
Model	1.0649e+09	5	212972126	F(5, 1212) = 132.39		
Residual	1.9498e+09	1212	1608711.19	Prob > F = 0.0000		
Total	3.0146e+09	1217	2477090.05	R-squared = 0.3532		
				Adj R-squared = 0.3506		
				Root MSE = 1268.3		

quantidade	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dia	-467.7036	21.24473	-22.02	0.000	-509.3841	-426.023
ecampanha	345.987	164.3652	2.10	0.035	23.51505	668.459
eexterno	-880.1926	178.6685	-4.93	0.000	-1230.726	-529.6588
eacidente	-1475.483	174.1031	-8.47	0.000	-1817.06	-1133.906
enotcia	673.4723	180.5434	3.73	0.000	319.26	1027.685
_cons	8073.035	103.5601	77.96	0.000	7869.858	8276.212

A análise da regressão, conforme apresentado na Tabela 5, revelou de maneira mais robusta algumas informações úteis para se compreender como a demanda por passageiros sofre influência das variáveis que teoricamente foram fixadas para análise no presente trabalho. A priori, o que se observa é que o modelo de regressão proposto é significativo, para

todos os coeficientes estimados. Em segundo lugar, a partir da verificação do R^2 ajustado, percebe-se que o modelo tem uma capacidade preditiva baixa, mas consistente com o que já se esperava, tendo em vista que a perspectiva do trabalho é entender a capacidade de influência (efeitos causados) pelas variáveis propostas. Interessante estudar, na sequência, a relevância dos coeficientes: sendo todos significativos a 5%, tem-se que eventos externos e acidentes têm uma influência negativa, estatisticamente, no resultado do modelo. A título de exemplificação, considerado os dados estimados, um acidente parece causar um efeito mais prolongado e, ao mesmo tempo, de maior força na série de passageiros transportados, comparativamente com a ocorrência de eventos de origem externa à CBTU. No tocante aos efeitos positivos, inesperadamente, percebeu-se que a ocorrência de notícias (anúncios, investimentos, matérias etc.) geram um impacto positivo quase que duas vezes superior à variável campanha. Assim, tem-se que a demanda é claramente mais influenciada positivamente pela ocorrência de notícias e, na contramão, negativamente influenciada principalmente por acidentes. Notadamente, há de se comentar que a variável acidente causa um impacto negativo muito mais duradouro e intenso, aparentemente, do que as demais variáveis, o que chama atenção durante a modelagem.

Diante das análises, interessante destacar que a presunção de normalidade precisa ser validade para que o modelo seja adequadamente analisado. Assim, realizou-se o teste de Shapiro-Wilk para o conjunto de variáveis. A tabela 6 apresenta o teste de Shapiro-Wilk:

Tabela 6 - Teste de Shapiro-Wilk

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
quantidade	1218	0.95557	33.541	8.768	0.00000
dia	1218	0.98927	8.100	5.222	0.00000
ecampanha	1218	0.96992	22.702	7.794	0.00000
eexterno	1218	0.96406	27.131	8.239	0.00000
eacidente	1218	0.96603	25.645	8.098	0.00000
enotcia	1218	0.96336	27.661	8.287	0.00000

O intuito do teste, como dito anteriormente, foi apenas confirmar se as inferências podem ser teoricamente justificadas, tendo em vista que o modelo de regressão linear presume a normalidade da série de dados. Com base nos resultados obtidos na tabela 6, por sua vez, pode-se chegar à conclusão de que o modelo é válido, tendo em vista o atendimento à hipótese de normalidade por todas as variáveis.

Na sequência, a fim de se evitar complicações com análises que desconsideram a heterocedasticidade, optou-se pelo teste de Bresch-Pagan, como forma de se investigar a robustez do modelo a partir dos resíduos do modelo. A Tabela 7 apresenta o teste de Breusch-Pagan:

Tabela 7 - Teste de Brescu-Pagan

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: dia ecampanha eexterno eacidente enotcia

chi2(5)      =    26.25
Prob > chi2  =    0.0001
```

De posse da análise do teste, cabe salientar que em razão do erro máximo tolerado no presente trabalho (desde o princípio das análises, 5%), não se rejeita a hipótese de

homocedasticidade, isto é, o postulado de que a variância dos resíduos é constante. Assim, não se encontram dados menos dispersos em torno da reta da regressão. Esse resultado, por sua vez, ajuda a verificar mais uma vez a robustez do modelo.

B) Modelagem de séries temporais (Basic VAR)

Por fim, para conclusão das análises referentes à demanda de passageiros transportados no sistema metroferroviário de João Pessoa/PB, foram rodados modelos Basic VAR, com quatro lags para iteração, objetivando-se verificar, finalmente, se a série temporal (inclusive a partir de uma perspectiva de forecast/previsão), incorpora as variáveis propostas.

Cada modelo VAR, por sua vez, tem sua análise restrita à interação individual de cada componente gerador, por sua vez, caracterizados a partir das variáveis campanha, notícia, acidente e evento externo. A Tabela 8 apresenta o modelo VAR tendo como variável endógena as campanhas

Tabela 8 - VAR (Evento: campanhas)

Log likelihood	=	-10494.81	AIC	=	17.31929
FPE	=	113952	HQIC	=	17.34776
Det(Sigma_ml)	=	110622.4	SBIC	=	17.39493

Equation	Parms	RMSE	R-sq	chi2	P>chi2
quantidade	9	1514.76	0.0765	100.6056	0.0000
ecampanha	9	.22157	0.0096	11.76947	0.1618

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
quantidade						
quantidade						
L1.	.082366	.0285487	2.89	0.004	.0264116	.1383203
L2.	.1249782	.0284291	4.40	0.000	.0692581	.1806982
L3.	.1268582	.0284232	4.46	0.000	.0711498	.1825666
L4.	.1108931	.0285325	3.89	0.000	.0549704	.1668159
ecampanha						
L1.	130.0513	196.0776	0.66	0.507	-254.2537	514.3563
L2.	-122.5987	196.0773	-0.63	0.532	-506.9031	261.7057
L3.	58.73816	196.0399	0.30	0.764	-325.493	442.9693
L4.	157.5341	195.88	0.80	0.421	-226.3835	541.4518
_cons	3261.65	271.8633	12.00	0.000	2728.807	3794.492
ecampanha						
quantidade						
L1.	6.79e-06	4.18e-06	1.63	0.104	-1.40e-06	.000015
L2.	5.15e-06	4.16e-06	1.24	0.216	-3.00e-06	.0000133
L3.	3.86e-06	4.16e-06	0.93	0.354	-4.29e-06	.000012
L4.	-4.04e-06	4.17e-06	-0.97	0.333	-.0000122	4.14e-06
ecampanha						
L1.	.0058474	.028681	0.20	0.838	-.0503663	.0620611
L2.	.0064878	.0286809	0.23	0.821	-.0497257	.0627014
L3.	.0091414	.0286755	0.32	0.750	-.0470614	.0653443
L4.	.0614132	.0286521	2.14	0.032	.0052562	.1175702
_cons	-.0217709	.0397664	-0.55	0.584	-.0997116	.0561699

Apesar da baixa confiabilidade do modelo, o que se colocou em cheque foi a capacidade preditiva da campanha, na previsão da demanda de passageiros. Diferentemente do que foi encontrado na análise convencional da regressão, que despreza pressupostos de autocorrelação serial ao longo do tempo (inclusive para projeções), verificou-se que a variável campanha não é significativa a 5% na explicação. Por outro lado, a partir da modelagem, que por sua vez não considerou outras variáveis (como a variável controle que era o dia da

semana), pode-se estimar ainda, com base nas iterações, que a variável companhia é claramente capaz de influenciar positivamente a demanda.

Na sequência foi estimado o modelo com a variável eventos externos. A Tabela 9 apresenta o modelo VAR tendo como variável endógena dos eventos externos:

Tabela 9 - VAR (Eventos externos)

Log likelihood	=	-10372.65	AIC	=	17.11803
FPE	=	93179.02	HQIC	=	17.14651
Det(Sigma_ml)	=	90456.4	SBIC	=	17.19367
Equation	Parms	RMSE	R-sq	chi2	P>chi2
quantidade	9	1511.5	0.0805	106.2905	0.0000
eexterno	9	.203592	0.0146	17.96649	0.0215

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
quantidade						
quantidade						
L1.	.0862173	.0289429	2.98	0.003	.0294902	.1429445
L2.	.1324255	.0287896	4.60	0.000	.0759989	.1888522
L3.	.1303228	.0287596	4.53	0.000	.073955	.1866905
L4.	.1062636	.0288663	3.68	0.000	.0496867	.1628405
eexterno						
L1.	300.3411	215.9218	1.39	0.164	-122.8579	723.5401
L2.	390.5546	216.9218	1.80	0.072	-34.6042	815.7135
L3.	165.5352	217.191	0.76	0.446	-260.1514	591.2219
L4.	-98.13376	216.1194	-0.45	0.650	-521.72	325.4525
_cons	3180.359	272.7898	11.66	0.000	2645.701	3715.017
eexterno						
quantidade						
L1.	1.14e-06	3.90e-06	0.29	0.770	-6.50e-06	8.78e-06
L2.	-1.21e-06	3.88e-06	-0.31	0.755	-8.81e-06	6.39e-06
L3.	3.95e-06	3.87e-06	1.02	0.308	-3.64e-06	.0000115
L4.	5.32e-06	3.89e-06	1.37	0.171	-2.30e-06	.0000129
eexterno						
L1.	.0955598	.0290838	3.29	0.001	.0385567	.1525629
L2.	-.0340868	.0292184	-1.17	0.243	-.0913539	.0231803
L3.	-.0196771	.0292547	-0.67	0.501	-.0770153	.0376611
L4.	.0421399	.0291104	1.45	0.148	-.0149154	.0991952
_cons	-.0143607	.0367436	-0.39	0.696	-.0863768	.0576555

O modelo VAR rodado permitiu se verificar que não somente a variável eventos externos é significativa, como o seu coeficiente tende, no longo prazo (para fins de forecast), a

influenciar positiva ou negativamente a série. Essa verificação, por sua vez, leva a crer que o conteúdo do tipo de evento é o principal delimitador desse efeito, cabendo, inclusive, investigação mais detalha sobre a variável, inclusive seu destrinchamento.

Considerando-se, em seguida, a variável acidentes, foi estimado novo modelo VAR, a exemplo dos modelos anteriores, com o objetivo de se verificar a eficácia da variável na previsão da série. A Tabela 10 apresenta o modelo VAR tendo como variável endógena a ocorrência de acidentes

Tabela 10 - VAR (acidentes)

Log likelihood	= -10392.61	AIC	= 17.15092
FPE	= 96294.04	HQIC	= 17.17939
Det(Sigma_ml)	= 93480.41	SBIC	= 17.22656

Equation	Parms	RMSE	R-sq	chi2	P>chi2
quantidade	9	1512.77	0.0790	104.0716	0.0000
ecidente	9	.208285	0.0214	26.48526	0.0009

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
quantidade						
quantidade						
L1.	.0903432	.0291349	3.10	0.002	.0332398	.1474466
L2.	.1255568	.0290889	4.32	0.000	.0685437	.18257
L3.	.1336042	.0291329	4.59	0.000	.0765048	.1907036
L4.	.101923	.0293035	3.48	0.001	.0444892	.1593569
eacidente						
L1.	231.9877	213.2389	1.09	0.277	-185.9529	649.9282
L2.	-6.269502	213.2777	-0.03	0.977	-424.2861	411.7471
L3.	125.063	213.0269	0.59	0.557	-292.462	542.5881
L4.	-379.9712	212.3912	-1.79	0.074	-796.2502	36.30786
_cons	3237.266	296.3393	10.92	0.000	2656.451	3818.08
eacidente						
quantidade						
L1.	-.0000125	4.01e-06	-3.12	0.002	-.0000204	-4.64e-06
L2.	-.0000107	4.01e-06	-2.66	0.008	-.0000185	-2.82e-06
L3.	-4.33e-06	4.01e-06	-1.08	0.281	-.0000122	3.54e-06
L4.	-2.74e-06	4.03e-06	-0.68	0.498	-.0000106	5.17e-06
eacidente						
L1.	-.0075409	.0293598	-0.26	0.797	-.065085	.0500033
L2.	-.036101	.0293651	-1.23	0.219	-.0936556	.0214536
L3.	-.0271461	.0293306	-0.93	0.355	-.084633	.0303409
L4.	.0135379	.0292431	0.46	0.643	-.0437775	.0708533
_cons	.2271935	.0408015	5.57	0.000	.1472241	.3071629

Além de significativa a 5%, o modelo em questão confirmou o que já se havia observado nos modelos anteriores (de regressão e correlação), bem como o que já se esperava desde a análise exploratória dos dados. A ocorrência de acidentes causa um efeito negativo na série, inclusive com repercussão nas observações posteriores (logo, nos dias seguintes). Esse efeito, além de negativo, parece ser constante e duradouro (das as iterações e seus resultados).

Por fim, para conclusão das análises dos fatores que podem influenciar a demanda, foi regredido novo modelo VAR, dessa vez considerando a variável notícias na modelagem. A Tabela 11 apresenta o modelo VAR tendo como variável endógena notícias:

Tabela 11 - VAR (notícias)

Log likelihood =	-10377.11	AIC	=	17.12539
FPE	= 93867.41	HQIC	=	17.15387
Det(Sigma_ml)	= 91124.68	SBIC	=	17.20103

Equation	Parms	RMSE	R-sq	chi2	P>chi2
quantidade	9	1512.59	0.0792	104.3763	0.0000
enotcia	9	.202548	0.0068	8.271241	0.4074

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
quantidade						
quantidade						
L1.	.0867323	.0287328	3.02	0.003	.030417	.1430475
L2.	.1174538	.0286105	4.11	0.000	.0613783	.1735292
L3.	.1292596	.0285097	4.53	0.000	.0733816	.1851377
L4.	.1145495	.028627	4.00	0.000	.0584416	.1706573
enotcia						
L1.	-91.93954	215.8008	-0.43	0.670	-514.9014	331.0223
L2.	426.2624	215.7525	1.98	0.048	3.395273	849.1295
L3.	-191.9201	216.051	-0.89	0.374	-615.3724	231.5322
L4.	-37.05013	216.1118	-0.17	0.864	-460.6215	386.5213
_cons	3251.62	272.4973	11.93	0.000	2717.535	3785.705
enotcia						
quantidade						
L1.	3.75e-06	3.85e-06	0.98	0.329	-3.79e-06	.0000113
L2.	-6.52e-07	3.83e-06	-0.17	0.865	-8.16e-06	6.86e-06
L3.	3.24e-06	3.82e-06	0.85	0.396	-4.24e-06	.0000107
L4.	6.16e-06	3.83e-06	1.61	0.108	-1.35e-06	.0000137
enotcia						
L1.	-.0302712	.0288974	-1.05	0.295	-.086909	.0263666
L2.	.0145489	.0288909	0.50	0.615	-.0420762	.071174
L3.	.0288069	.0289309	1.00	0.319	-.0278966	.0855103
L4.	-.0091865	.028939	-0.32	0.751	-.0659059	.0475329
_cons	-.0311647	.0364895	-0.85	0.393	-.1026828	.0403533

A exemplo do que se verificou com a variável campanha, o modelo não apresentou significância a 5%, sendo até mais duvidoso na explicação da demanda. Importante comentar, novamente, que esse resultado não anula os resultados anteriores (AED, Correlação,

35

Regressão), apenas leva a novas perspectivas de análise, em que, primordialmente, não se pode considerar como primordial levar em conta a variável notícias para se entende o comportamento da demanda.

CONCLUSÕES

Com o propósito de se analisar a demanda, o presente trabalho buscou investigar, estatisticamente, como alguns fatores podem ser influenciadores ou não da quantidade de passageiros transportados. Fazendo uso de uma série temporal com 1218 observações, compreendendo o período de 2013 a 2016, estimou-se uma série de modelos e testou-se de forma razoável variáveis que, originalmente, esperava-se que pudessem ser capazes de explicar de alguma forma o crescimento ou declínio da demanda de passageiros.

Considerando as variáveis dia da semana (controle), eventos externos, campanhas, acidentes e notícias, buscou-se entender como a variável dependente quantidade de passageiros transportados - diariamente - poderia ser influenciada. A princípio, pela clássica análise exploratória, percebeu-se que as variáveis campanhas e acidentes seriam as mais relevantes para impactarem a demanda, no primeiro caso de forma positiva e no segundo de forma negativa. A partir das análises das correlações e regressões, por sua vez, verificou-se que não somente se tratam de variáveis relevantes, mas, no caso dos acidentes, tem-se uma variável que consegue impactar de forma bastante forte a demanda, levando, possivelmente aos picos de queda que claramente se verifica na série temporal

A fim de se verificar com maior detalhamento, dessa vez fazendo uso de um modelo de séries temporais, foi modelada cada variável, juntamente à quantidade de passageiros transportados, por meio de modelos VAR, a fim de se testar uma última vez quais fatores são mais influentes para demanda. Os resultados levaram à conclusão de que a variável acidentes é, de fato, a mais impactante para a demanda, sendo a principal responsável, aparentemente, pelas quedas significantes no volume transportado. De forma mais modesta, por outro lado, as campanhas e ações da empresa parecem ter um resultado claramente positivo na demanda, persistindo de forma mais modesta.

Diante dos resultados encontrados, por fim, percebe-se que a demanda do Sistema metroferroviário de João Pessoa/PB parece ser influenciado, principalmente pelos fatores "acidentes" e "campanhas internas" da CBTU, isto é, a elevação do número de acidentes ou situações que envolvem riscos pode levar a efeitos negativos relativamente persistentes na série, explicando, inclusive, a elevada variância no número de passageiros transportados. Por outro lado, ao se analisar o fator "campanhas internas" (tais como ações sociais, campanhas de saúde, campanhas de marketing etc.) percebe-se que o público do sistema parece demanda mais transporte, de forma modesta, mas ainda assim significativa, quando tais eventos acontecem. Importante levar em conta, ainda, que embora o fator acidentes persista por mais tempo na série, há de se comentar que o fator campanhas tem sua relevância bastante clara, inclusive como tática capaz de gerar resultados positivos para o transporte, no longo prazo (além de ações para redução de acidentes e outros riscos aos passageiros).

REFERÊNCIAS BIBLIOGRÁFICAS

BROWN, R. G. Statistical Forecasting for Inventory Control. New York: McGraw-Hill Book, 1959.

MURÇA, M. C. R; MULLER, C. Urban public transportation: a demand analysis for the city of Salvador. Journal of Transport Literature, vol. 8, n. 1, 2014.

BRASIL, Companhia Brasileira de Trens Urbanos (CBTU). O sistema metroferroviário de João Pessoa, 2017. Disponível em: < <http://www.cbtu.gov.br> >. Acesso em: 15 de junho de 2017.

PRODANOV, C. C; FREITAS, E. C. Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2ª Edição, Universidade Feevale: Novo Hamburgo, RS, 2013.

VIANNA, I. O. Metodologia do Trabalho Científico: um enfoque didático da produção científica. EPU, 2001.

LUSTOSA, L.; MESQUISTA, A. M.; QUELHAS, O.; OLIVEIRA, R. Planejamento e controle da produção. Rio de Janeiro: Elsevier, 2008.

MOREIRA, D. A. Administração da produção e operações. 2ª edição, São Paulo: Cengage Learning, 2011.

ANDERSON, D. R. Estatística aplicada à Administração e Economia. 2ª Edição. São Paulo: Cengage Learning, 2011.

FIELD, A. Descobrendo a estatística usando SPSS. 2ª edição. Porto Alegre: Artmed, 2009.

BRUNI, A. L. PASW aplicado à pesquisa acadêmica. 2ª edição. São Paulo: Atlas, 2011.

FREUND, J. E. Estatística aplicada: economia, administração e contabilidade. 11ª Edição. Porto Alegre: Bookman, 2006.

SERRA, C. M. V.; TAVARES, H. R.; SANTOS, J. C. C. Aplicação de séries temporais na análise de demanda turística no estado do Pará usando modelos de Holt-Winters. XX Encontro Nac. de Eng. De Produção. Brasil: Porto Alegre, RS, 29 de out a 01 de novembro, 2005.

LATORRE, M. R. D. O.; CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre aspectos metodológicos. Revista brasileira de epidemiologia, vol. 4, n. 3, São Paulo, nov., 2011.

MORETTIN, P. A.; TOLOI, C. M. Análise de Séries Temporais. São Paulo: Edgar Blucher, 2004.

CAVALCANTI, M. A. F. H. Identificação de modelos VAR e causalidade de Granger: uma nota de advertência. Economia Aplicada, vol. 14, n. 2, Ribeirão Preto, abril/junho, 2010.