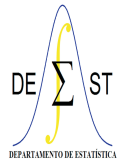




UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA



Uma Avaliação do Processo de Chegadas no Restaurante Universitário da UFOP

Juliana Gonçalves Soares

Ouro Preto-MG

Julho de 2019

Juliana Gonçalves Soares

Uma Avaliação do Processo de Chegadas no Restaurante Universitário da UFOP

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador

Dr. Anderson Ribeiro Duarte

Co-orientador

Me. Helgem de Souza Ribeiro Martins

UNIVERSIDADE FEDERAL DE OURO PRETO – UFOP
DEPARTAMENTO DE ESTATÍSTICA – DEEST

Ouro Preto-MG

Julho de 2019

Ata de Defesa

Ata da sessão pública para julgamento da monografia de Juliana Gonçalves Soares, aluna do curso de Bacharelado em Estatística, do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto.

Aos dez dias do mês de julho do ano de dois mil e dezenove, às quinze horas, no Auditório do Departamento de Estatística, reuniu-se a comissão julgadora composta pelos professores Anderson Ribeiro Duarte e Spencer Barbosa da Silva, e pelos servidores técnicos administrativos em educação Helgem de Souza Ribeiro Martins e Josino José Barbosa, para avaliar a monografia da aluna **Juliana Gonçalves Soares**, intitulada *Uma Avaliação do Processo de Chegadas no Restaurante Universitário da UFOP*. A sessão pública foi aberta pelo professor Anderson Ribeiro Duarte, presidente da comissão julgadora e orientador da pesquisa, que, após formalidades de praxe, passou a palavra a aluna para apresentação oral e, em seguida, arguição pelos membros da banca. Terminada a arguição, a comissão reuniu-se em sessão secreta para elaborar o relatório individual de apreciação da monografia e decidiu pela aprovação da mesma com a nota 10,0 (dez pontos e zero décimos). Nada havendo mais a tratar, foi encerrada a sessão da qual se lavrou a presente ata que vai assinada pela comissão julgadora.

Ouro Preto, 10 de julho de 2019.



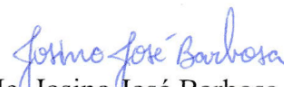
Prof. Dr. Anderson Ribeiro Duarte
Departamento de Estatística - DEEST
Universidade Federal de Ouro Preto - UFOP
Presidente



Me. Helgem de Souza Ribeiro Martins
Pró-Reitoria de Pesquisa e Pós-Graduação
PROPP
Universidade Federal de Ouro Preto - UFOP
Coorientador da pesquisa



Prof. Dr. Spencer Barbosa da Silva
Departamento de Estatística - DEEST
Universidade Federal de Ouro Preto - UFOP
Membro da banca examinadora



Me. Josino José Barbosa
Núcleo de Avaliação Institucional - NAI
Universidade Federal de Ouro Preto - UFOP
Membro da banca examinadora

Dedico esse trabalho a Deus, pela vida e força.

Aos meus pais João e Helena, pelo apoio e amor incondicional.

Dedico a minha irmã Janaína, pelo companheirismo.

As minhas amigas Carolina, Débora e Natália, pelo apoio, amizade e aprendizado.

Aos professores e ao departamento de estatística da

Universidade Federal de Ouro Preto, pelo conhecimento e auxílio.

A empresa júnior Estatis, pela oportunidade de aplicar os conhecimentos adquiridos.

Aos meus amigos, familiares, e a todos que contribuíram direta ou indiretamente para
a conclusão dessa etapa.

AGRADECIMENTOS

Ao professor Anderson, pela orientação, suporte e conhecimento transmitido, tanto durante o curso, como na execução desse trabalho.

Ao professor Helgem, pela co-orientação, pelos conhecimentos transmitidos e auxílio.

A Universidade Federal de Ouro Preto, pela oportunidade de participar do projeto de iniciação científica, que conseqüentemente é a base desse trabalho.

A gestão do Restaurante Universitário da UFOP, pela disponibilização dos dados e informações sobre a coleta deles.

*“Não é a força, mas a perseverança
que realiza grandes coisas.”*

Samuel Johnson

LISTA DE FIGURAS

1	Processo de Poisson não homogêneo com taxa $a(t)$	25
2	Processo de Poisson não homogêneo com taxa $a(t)$	26
3	Processo de filas básico. (Modificado de Hillier e Lieberman (2013)) ...	35
4	Tamanhos das sub-amostras - segunda-feira	49
5	Tamanhos das sub-amostras - terça-feira	51
6	Tamanhos das sub-amostras - quarta-feira	52
7	Tamanhos das sub-amostras - quinta-feira	53
8	Tamanhos das sub-amostras - sexta-feira	55
9	Boxplots dos tempos de serviço por dia e intervalo	56
10	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas	58
11	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Segunda-feira	59
12	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Terça-feira	60
13	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Quarta-feira	61
14	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Quinta-feira	62
15	Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Sexta-feira	63
16	FAC e FACP da série (22/01/2018)	65
17	FAC e FACP dos resíduos (22/01/2018)	66
18	Resíduos Padronizados (22/01/2018)	67
19	Histograma e QQplot dos resíduos (22/01/2018)	67
20	FAC e FACP da série (29/01/2018)	68
21	FAC e FACP dos resíduos (29/01/2018)	69
22	Resíduos Padronizados (29/01/2018)	70
23	Histograma e QQplot dos resíduos (29/01/2018)	70
24	FAC e FACP da série (05/03/2018)	71
25	FAC e FACP dos resíduos (05/03/2018)	72
26	Resíduos Padronizados (05/03/2018)	73

27	Histograma e QQplot dos resíduos (05/03/2018)	73
28	FAC e FACP da série (09/04/2018)	74
29	FAC e FACP dos resíduos (09/04/2018)	75
30	Resíduos Padronizados (09/04/2018)	76
31	Histograma e QQplot dos resíduos (09/04/2018)	76
32	FAC e FACP da série (16/04/2018)	77
33	FAC e FACP dos resíduos (16/04/2018)	78
34	Resíduos Padronizados (16/04/2018)	79
35	Histograma e QQplot dos resíduos (16/04/2018)	79

LISTA DE TABELAS

1	Estatísticas descritivas gerais	45
2	Estatísticas descritivas por dia da semana	46
3	Limites dos intervalos de tempo	47
4	Estatísticas descritivas por faixa	47
5	Estatísticas descritivas e tamanho da sub-amostra por faixa - segunda-feira	48
6	Estatísticas descritivas e tamanho da sub-amostra por faixa - terça-feira	50
7	Estatísticas descritivas e tamanho da sub-amostra por faixa - quarta-feira	51
8	Estatísticas descritivas e tamanho da sub-amostra por faixa - quinta-feira	53
9	Estatísticas descritivas e tamanho da sub-amostra por faixa - sexta-feira	54
10	Coeficiente e medida AIC (22/01/2018) - modelo ARIMA(1,1,0)	66
11	Medidas de acurácia (22/01/2018) - modelo ARIMA(1,1,0)	68
12	Coeficientes e medida AIC (29/01/2018) - modelo AR(2)	69
13	Medidas de acurácia (29/01/2018) - modelo AR(2)	71
14	Coeficiente e medida AIC (05/03/2018) - modelo ARIMA(1,1,0)	72
15	Medidas de acurácia (05/03/2018) - modelo ARIMA(1,1,0)	74
16	Coeficiente e medida AIC (09/04/2018) - modelo ARIMA(1,1,0)	75
17	Medidas de acurácia (09/04/2018) - modelo ARIMA(1,1,0)	77
18	Coeficientes e medida AIC (16/04/2018) - modelo ARIMA(1,0,0)(0,0,1) ₅ .	78
19	Medidas de acurácia (16/04/2018) - modelo ARIMA(1,0,0)(0,0,1) ₅	80
20	Coeficientes e medidas AIC (terça-feira)	81
21	Medidas de acurácia (terça-feira)	81
22	Coeficientes e medidas AIC (quarta-feira)	82
23	Medidas de acurácia (quarta-feira)	83
24	Coeficientes e medidas AIC (quinta-feira)	84
25	Medidas de acurácia (quinta-feira)	85
26	Coeficientes e medidas AIC (sexta-feira)	85
27	Medidas de acurácia (sexta-feira)	86

RESUMO

A grande maioria dos sistemas de filas, com interesse prático, possui limitação de recursos, isso vale tanto nos servidores quanto nas áreas de espera. Tais limitações geram impactos financeiros diretos nos custos do processo, e portanto, o uso desses recursos de maneira racional se torna preponderante. A solução de problemas dessa natureza se torna ainda mais complexa quando o objeto de interesse deixa de ser um sistema de filas simples e se torna uma rede de filas complexa. Com o intuito de aprofundar no tema, esta pesquisa discutiu alguns conceitos de Teorias de filas. Um objetivo de utilização de todo esse arcabouço teórico foi a tentativa de modelagem do processo de chegadas no sistema de atendimento do Restaurante Universitário da Universidade Federal de Ouro Preto (UFOP) por meio de Teoria das filas. Conhecimentos teóricos de modelos de séries temporais também ofereceram um valioso suporte nessa investigação. A metodologia de pesquisa adotada foi o estudo de caso considerando a disponibilidade de dados. Os dados de passagem de usuários do Restaurante Universitário pela catraca, com horário exato, estavam disponíveis. Esse estudo é útil em contribuir para o melhor gerenciamento das capacidades do Restaurante Universitário, além de fornecer informações que contribuam para o planejamento de expansões e melhorias no local.

Palavras-chave: Processos de Poisson; Restaurante Universitário; Teoria de Filas; Séries temporais.

ABSTRACT

Several queuing systems, with practical interest, have resource limitations, this holds true both in servers and in waiting areas. These limitations generate direct financial impacts on the process costs, and therefore, the use of these resources in a rational way becomes preponderant. The solution of problems becomes even more complex when the object of interest it is not a simple queuing system and becomes a complex queue network. In order to discuss it in depth into the topic, this research address to some issue of Queue Theories. An objective of using this theoretical framework was the attempt to model the arrival process in the attendance system of the University Restaurant of the Federal University of Ouro Preto (UFOP) by means of Queue Theory. Theoretical knowledge of time series models also offered valuable support in this investigation. The research methodology adopted was the case study considering the availability of data. The data of passage of users of the University Restaurant by ticket gate, with exact time, were available. This study is useful in contributing to the better management of the University Restaurant capabilities, as well as providing information that contributes to the planning of expansions and improvements.

Keywords: Poisson Processes; University Restaurant; Queue Theories; Time Series.

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	15
2	ASPECTOS METODOLÓGICOS	16
2.1	PROCESSOS DE CONTAGEM	16
2.1.1	Processos de Poisson Homogêneos	17
2.1.2	Processos de Nascimento	23
2.1.3	Processos de Poisson Não Homogêneos	25
2.2	MODELOS DE SÉRIES TEMPORAIS	26
2.2.1	Modelos de Erro ou de Regressão	29
2.2.2	Modelos AR	29
2.2.3	Modelos MA	31
2.2.4	Modelos ARMA	32
2.2.5	Modelos ARIMA	33
2.3	TEORIA DAS FILAS	34
2.3.1	Formas de Chegadas	36
2.3.2	Formas de Atendimento	36
2.3.3	Disciplina da Fila	37
2.3.4	Medidas de Desempenho no Sistema de Filas	38
2.3.5	Distribuição Exponencial	39
2.3.6	Distribuição Poisson	40
2.3.7	Processo de Nascimento e Morte	40
2.3.8	Modelos de Filas	42
2.3.8.1	Modelo $M/M/1$	43
2.3.8.2	Modelo $M/M/c$	43
3	RESULTADOS E DISCUSSÕES	45
3.1	ANÁLISES DESCRITIVAS DOS TEMPOS DE SERVIÇO	45
3.1.1	Divisão do Horário de Funcionamento em Intervalos	46
3.1.2	Estatísticas Descritivas por Dia da Semana	48
3.1.2.1	Segunda–Feira	48
3.1.2.2	Terça–Feira	50

3.1.2.3	Quarta–Feira	51
3.1.2.4	Quinta–Feira	52
3.1.2.5	Sexta–Feira	54
3.1.2.6	Representações gráficas	55
3.2	AJUSTE DO MODELO DE SÉRIES TEMPORAIS	64
4	CONSIDERAÇÕES FINAIS.....	88
	REFERÊNCIAS.....	91

1 INTRODUÇÃO

Conforme ressaltado no livro, *Engineering Optimization: An Introduction with Metaheuristic Applications* (YANG, 2010), embora os problemas de otimização estejam presentes em todo lugar, da engenharia à ciência da computação e do sequenciamento de tarefas à economia, constatar tal fato não torna a resolução de tais problemas mais fácil. Na verdade, problemas de descrição bastante simples podem ser muito difíceis de resolver. Tome-se por exemplo o problema do caixeiro viajante, no qual um vendedor precisa visitar, digamos, 50 cidades, exatamente uma única vez, em uma sequência tal que a distância total percorrida seja minimizada. Apesar da facilidade de definição e compreensão deste problema e da simplicidade do objetivo a ser minimizado, é de certa forma surpreendente que não se conheça ainda um algoritmo eficiente para ele.

Os desenvolvimentos mais recentes que foram criados ao longo das últimas duas décadas para o problema do caixeiro viajante tendem a usar algoritmos metaheurísticos. Na verdade, as mais modernas técnicas de otimização de maneira geral são usualmente heurísticas ou metaheurísticas, tais como o *simulated annealing*, a otimização por enxame de partículas, a busca harmônica e os algoritmos genéticos. Estes algoritmos vêm se tornando bastante poderosos na resolução de difíceis problemas de otimização, em todas as principais áreas da ciência e da engenharia, bem como nas aplicações industriais.

Sempre que há incerteza sobre o fluxo de produtos, usuários, mensagens, e assim por diante, com uma taxa de chegada, e seu processamento, com uma taxa de serviço, tem-se como resultado um sistema de filas. O seu arranjo em uma configuração em rede, é uma generalização natural e relevante, pelos diversos sistemas reais que pode modelar.

A solução de problemas dessa natureza está diretamente vinculada ao desempenho da rede de filas para a qual deseja-se obter a solução ótima e a configu-

ração de número de servidores e tamanho das áreas de espera em cada nó de uma rede de filas vista como um grafo $G = (N, A, P)$, em que N é o conjunto de todos os nós que compõe a rede, A é o conjunto de arcos que interconectam pares de nós e P são as respectivas probabilidades de roteamento entre os arcos.

O clássico problema BCAP (*Buffer and server optimization problem*) busca otimizar o número de áreas de espera e servidores de forma que a taxa de atendimento resultante seja maior que um limiar predefinido de taxa de atendimento. Para tal, serão necessários resultados sobre obtenção de medidas de desempenho em filas e redes de filas finitas. Diversas são as ferramentas existentes para a análise de desempenho em redes de filas. Dentre elas, destacam-se os métodos forma-produto, métodos numéricos, métodos aproximados e simulações.

Uma metodologia amplamente empregada na obtenção de medidas de desempenho para redes de filas são os métodos de simulação. Diversos são os estudos bem sucedidos baseados em tais métodos. Uma ferramenta bastante empregada conjuntamente com as metodologias para análise de desempenho são os métodos de otimização numérica. A otimização é aplicada aos modelos de programação matemática para resolução de problemas nos quais deseja-se minimizar ou maximizar uma função objetivo através de uma escolha sistemática de valores para suas variáveis dentre os valores pertencentes ao seu conjunto de possíveis valores. Dentre os métodos de otimização, os não-lineares destacam-se na solução de problemas de otimização em redes de filas, devido à natureza da função objetivo, ou das restrições existentes nos modelos de otimização, geralmente não-lineares.

A questão colocada neste estudo é de como é possível modelar sistemas reais envolvendo incerteza no fluxo de produtos e no seu processamento, prever com precisão medidas de desempenho dos modelos e planejar adequadamente tais sistemas para que tenham desempenho ótimo.

O presente estudo teve em sua origem o tratamento de filas com auxílio de

resultados de otimização e estudos de teoria de filas, o objetivo inicial era completamente teórico. Um contato realizado entre os orientadores e a gestão do Restaurante Universitário da UFOP levou para uma bifurcação entre o objetivo previamente teórico e um novo objetivo completamente aplicado. Um vasto conjunto de dados reais a respeito da fila de atendimento do restaurante foi disponibilizado e não poderia ser ignorado.

Tratou-se de um objeto de estudo inédito e interessantíssimo para a universidade. Acessar esse conjunto de dados foi um convite que não poderia ser negado pelos envolvidos neste projeto. Desta forma, a parte inicial desse estudo foi consumida por leitura e análise de dados teóricos, e por um conjunto de reuniões entre os envolvidos e o restaurante universitário para de fato separar e disponibilizar os itens relevantes desse vasto banco de dados.

O procedimento seguiu inicialmente de um estudo da estrutura de dados e informações que poderiam ser disponibilizadas pelo Restaurante Universitário da UFOP. Esse estudo passou inicialmente por reuniões com os gestores do restaurante, isso para que se pudesse conhecer os dados que realmente são armazenados e os dados que se perdiam por não serem adequadamente coletados.

Após o trâmite com os gestores do restaurante universitário, a base de dados foi disponibilizada. Em suma, apenas existia uma marcação de horário de cada entrada individual por cada uma das roletas no restaurante universitário. Obviamente, muito de informação estava portanto indisponível. Por outro lado, esse conjunto de informações já era bastante vasto e serviu para um estudo bastante relevante.

O passo inicial foi a execução de um detalhado estudo descritivo dos dados, esse estudo se encontra descrito na seção 3.1, e posteriormente estudos mais direcionados para análises de teoria de filas foram encaminhados como poderá ser visto ao longo desse texto.

1.1 OBJETIVOS

Com base nesse propósito de pesquisa, os objetivos do presente estudo são:

- apresentação de uma revisão bibliográfica na área de teoria de filas e séries temporais;
- apresentação de discussões teóricas sobre processos de filas com definições de processos de Poisson homogêneos e não homogêneos;
- discussão sobre modelos de séries temporais para ajustar as taxas associadas ao processo de filas na entrada do Restaurante Universitário da UFOP;
- análise e discussão das constatações baseadas nas investigações da disponibilidade de dados.

O presente relato é iniciado com um capítulo introdutório que discute os aspectos de pesquisa abordados durante a concepção desse projeto de estudo, o delineamento prévio de objetivos a serem trabalhados. Posteriormente, o capítulo de Aspectos Metodológicos detalha o vasto ferramental estatístico que foi discutido e utilizado ao longo da concepção dessa monografia, essa apresentação é uma mescla de relatório didático de conceitos associada com um pequeno referencial teórico acerca dos assuntos discutidos. O terceiro capítulo apresenta de forma mais definitiva todo o conjunto de resultados alcançados, detalhando a análise estatística descritiva do conjunto de dados sob investigação e também o ajuste de séries temporais adotado para descrever a função taxa referente ao processo de Poisson não homogêneo que descreve o fluxo de chegadas de usuários no restaurante universitário da Universidade Federal de Ouro Preto. Por fim, o último capítulo apresenta as conclusões alcançadas através dessa investigação e também propostas de continuidade desse estudo.

2 ASPECTOS METODOLÓGICOS

Conhecimentos de definições básicas em processos de contagem e teoria de filas são preponderantes para uma melhor compreensão dos assuntos abordados nesse texto, uma pequena revisão será apresentada nesse capítulo, de uma forma geral os conceitos aqui mencionados formam uma coletânea de estudos e resultados que se encontram em diversas referências clássicas, em particular, foram utilizadas informações de Bass (2011), Bhat (1984), Gross e Harris (1985) e Ross (1992).

Da mesma forma, um resumo informativo com conceitos introdutórios em modelagem de séries temporais é relevante para o presente estudo (MORETTIN, 2008; MORETTIN; TOLOI, 2006). Além disso, uma abordagem acerca de estudos de teoria de filas e medidas de desempenho nos sistemas de filas (ARENALES; ARMEN-TANO; MORABITO, 2007; BASS, 2011; GROSS; HARRIS, 1985; HILLIER; LIEBER-MAN, 2013; ROSS, 1992). Diversas referências confirmam a relevância e importância de investigações nessa área (CRUZ; DUARTE; WOENSEL, 2008; CRUZ et al., 2012; CRUZ; DUARTE; SOUZA, 2018; MARTINS et al., 2019).

2.1 PROCESSOS DE CONTAGEM

Definição: Um processo $X = \{X_t; t \geq 0\}$ assumindo valores no espaço de estados $E = \{0, 1, 2, \dots\}$ é dito ser um processo de contagem se:

- i) $X_0 = 0$.
- ii) X_t cresce somente em saltos.
- iii) Se $s < t$ então $X_s \leq X_t$.
- iv) É uma função contínua à direita.

Exemplos tradicionais dos processos de contagem são aqueles que registram o número de ocorrências de determinados eventos.

- Processos que registram o número de acidentes que ocorrem em um trecho de estrada;
- Número de itens defeituosos produzidos por uma linha de produção;
- Número de chegadas a uma estação de serviço.

Esses são todos exemplos desses processos.

2.1.1 Processos de Poisson Homogêneos

Definição: Um processo de contagem $X = \{X_t; t \geq 0\}$ é um processo de Poisson homogêneo se possui incrementos independentes e estacionários e, além disso, satisfaz as seguintes condições adicionais:

- $P(X_{t+h} - X_t = 1 | X_t = i) = \lambda h + o(h)$ quando $h \downarrow 0$;
- $P(X_{t+h} - X_t = 0 | X_t = i) = 1 - \lambda h + o(h)$ quando $h \downarrow 0$;
- $X_0 = 0$.

Uma função f é dita ser uma função $o(h)$ quando $\lim_{h \downarrow 0} \frac{f(h)}{h} = 0$.

Exemplo: Se X é uma variável aleatória com distribuição Exponencial de parâmetro λ , então têm-se que $P(X \leq h) = \lambda h + o(h)$ quando $h \downarrow 0$.

$$P(X \leq h) = 1 - e^{-\lambda h}$$

expandindo em série de Taylor tem-se:

$$\begin{aligned} P(X \leq h) &= 1 - \left[\sum_{n=0}^{\infty} \frac{(-\lambda h)^n}{n!} \right] \\ &= 1 - \left[1 - \lambda h + \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} \right] \\ &= \lambda h - \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} \end{aligned}$$

Note que como $h \downarrow 0$, então é razoável observar a situação com $0 < h < \frac{1}{\lambda}$, portanto:

$$\begin{aligned} \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} &= (-\lambda h)^2 \sum_{n=2}^{\infty} \frac{(-\lambda h)^{n-2}}{n!} \\ &\leq (-\lambda h)^2 \sum_{n=2}^{\infty} \frac{(-\lambda h)^{n-2}}{(n-2)!} \\ &\leq (\lambda h)^2 e^{-1} = o(h) \end{aligned}$$

Portanto:

$$\begin{aligned} P(X \leq h) &= \lambda h - \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} \\ &= \lambda h - o(h) \end{aligned}$$

O parâmetro λ envolvido na definição do processo de Poisson é a taxa de ocorrência ou de nascimento do processo.

Considere:

- $P(X_{t+h} - X_t = 1 | X_t = i) = \lambda h + o(h)$ quando $h \downarrow 0$;
- $P(X_{t+h} - X_t = 0 | X_t = i) = 1 - \lambda h + o(h)$ quando $h \downarrow 0$.

Portanto $P(X_{t+h} - X_t = 2 | X_t = i) = o(h)$ quando $h \downarrow 0$, ou seja, não acontecem ocorrências em blocos. A probabilidade de acontecer mais de um evento em um pequeno intervalo de tempo é desprezível quando dividido pelo comprimento do intervalo. A interpretação disso é que dois ou mais eventos não ocorrem simultaneamente, ou seja os saltos do processo são unitários.

De interesse em um processo de Poisson, é calcular a função de probabilidade de X_t , isto é $P(X_t = j)$ para $t > 0$ e para $j \in \mathbb{N}$. Para $t > 0$ e $j \in \mathbb{N}$, defina $P_j(t) = P(X_t = j)$. Para obter tais probabilidades, considere a decomposição do intervalo $[0, t+h] = [0, t] \cup (t, t+h]$.

Considere $j = 0$

$$\begin{aligned}
 P_0(t+h) &= P(X_{t+h} = 0) \\
 &= P(X_t = 0)P(X_{t+h} - X_t = 0 | X_t = 0) \\
 &= P_0(t)(1 - \lambda h + o(h)) \\
 &= P_0(t) - \lambda h P_0(t) + o(h)P_0(t) \\
 P_0(t+h) - P_0(t) &= -\lambda h P_0(t) + o(h)
 \end{aligned}$$

dividindo os dois lados da igualdade por h e tomando o limite com $h \downarrow 0$, tem-se:

$$\begin{aligned}
 \frac{P_0(t+h) - P_0(t)}{h} &= \frac{-\lambda h P_0(t)}{h} + \frac{o(h)}{h} \\
 \lim_{h \downarrow 0} \frac{P_0(t+h) - P_0(t)}{h} &= \lim_{h \downarrow 0} \frac{-\lambda h P_0(t)}{h} + \lim_{h \downarrow 0} \frac{o(h)}{h} \\
 P_0'(t) &= -\lambda P_0(t)
 \end{aligned}$$

Dessa forma, para obter $P(X_t = 0) = P_0(t)$ é necessário obter a resolução da equação diferencial $P_0'(t) = -\lambda P_0(t)$. Não é objetivo desse texto profundidade em técnicas para solucionar equações diferenciais, mas é possível afirmar que para qualquer equação do tipo: $f'(t) + c_1 f(t) = c_2$, a solução é dada por:

$$f(t) = \frac{c_2}{c_1} + k e^{-c_1 t}$$

$P_0'(t) = -\lambda P_0(t) \implies P_0'(t) + \lambda P_0(t) = 0 \implies P_0(t) = k e^{-\lambda t}$. Observe que $P_0(0) = P(X_0 = 0) = 1$, portanto $P_0(0) = k e^{-\lambda 0} = k = 1$ e então, $P_0(t) = e^{-\lambda t}$.

Considere agora $j = 1$

$$\begin{aligned}
 P_1(t+h) &= P(X_{t+h} = 1) \\
 P_1(t+h) &= P(X_t = 0)P(X_{t+h} - X_t = 1 | X_t = 0) \\
 &\quad + P(X_t = 1)P(X_{t+h} - X_t = 0 | X_t = 1) \\
 P_1(t+h) &= P_0(t)(\lambda h - o(h)) + P_1(t)(1 - \lambda h - o(h)) \\
 P_1(t+h) &= e^{-\lambda t}(\lambda h - o(h)) + P_1(t)(1 - \lambda h - o(h)) \\
 P_1(t+h) - P_1(t) &= -\lambda h P_1(t) + \lambda h e^{-\lambda t} - o(h)(e^{-\lambda t} + P_1(t)) \\
 P_1(t+h) - P_1(t) &= -\lambda h P_1(t) + \lambda h e^{-\lambda t} + o(h)
 \end{aligned}$$

dividindo os dois lados por h e tomando o limite com $h \downarrow 0$, tem-se:

$$\lim_{h \downarrow 0} \frac{P_1(t+h) - P_1(t)}{h} = \lim_{h \downarrow 0} \frac{-\lambda h P_1(t)}{h} + \lim_{h \downarrow 0} \frac{\lambda h e^{-\lambda t}}{h} + \lim_{h \downarrow 0} \frac{o(h)}{h}$$

$$\lim_{h \downarrow 0} \frac{P_1(t+h) - P_1(t)}{h} = \lim_{h \downarrow 0} \frac{-\lambda h P_1(t)}{h} + \lim_{h \downarrow 0} \frac{\lambda h e^{-\lambda t}}{h} + \lim_{h \downarrow 0} \frac{o(h)}{h}$$

$$P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}$$

Qualquer equação do tipo: $f'(t) + p(t)f(t) = q(t)$, tem solução dada por:

$$f(t) = \exp\left(-\int_0^t p(x)dx\right) \int_0^t q(x) \exp\left(\int_0^x p(y)dy\right) dx + k$$

$$\lim_{h \downarrow 0} \frac{P_1(t+h) - P_1(t)}{h} = \lim_{h \downarrow 0} \frac{-\lambda h P_1(t)}{h} + \lim_{h \downarrow 0} \frac{\lambda h e^{-\lambda t}}{h} + \lim_{h \downarrow 0} \frac{o(h)}{h}$$

$$P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}$$

Resolvendo a equação diferencial têm-se:

$$f(t) = \exp\left(-\int_0^t p(x)dx\right) \int_0^t q(x) \exp\left(\int_0^x p(y)dy\right) dx + k$$

$$P_1(t) = \exp\left(-\int_0^t \lambda dx\right) \int_0^t \lambda e^{-\lambda x} \exp\left(\int_0^x \lambda dy\right) dx + k$$

$$P_1(t) = \exp(-\lambda t) \int_0^t \lambda e^{-\lambda x} \exp(\lambda x) dx + k$$

$$P_1(t) = e^{-\lambda t} \int_0^t \lambda e^{-\lambda x + \lambda x} dx + k \implies P_1(t) = e^{-\lambda t} \int_0^t \lambda dx + k$$

$$P_1(t) = e^{-\lambda t} (\lambda t + k)$$

Note que $P_1(0) = P(X_0 = 1) = e^{-\lambda 0} (\lambda 0 + k) = k = 0$

$$P_1(t) = \lambda t e^{-\lambda t}$$

Procedendo dessa maneira, é possível encontrar $P_2(t), P_3(t), \dots$ mas usando o Princípio de Indução Matemática isso pode ser garantido de forma mais eficiente. Admita que $P_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!}$. Já foi verificado que essa fórmula é válida para $j = 0$ e $j = 1$, admitindo validade para j , garantir validade para $j + 1$ é suficiente para validar através da Indução Matemática para todo $j \in \mathbb{N}$.

$$P_{j+1}(t+h) = P(X_{t+h} = j+1)$$

$$P_{j+1}(t+h) = P(X_t = j)P(X_{t+h} - X_t = 1 | X_t = j) \\ + P(X_t = j+1)P(X_{t+h} - X_t = 0 | X_t = j+1)$$

$$P_{j+1}(t+h) = P_j(t) (\lambda h - o(h)) + P_{j+1}(t) (1 - \lambda h - o(h))$$

$$P_{j+1}(t+h) = \frac{e^{-\lambda t} (\lambda t)^j}{j!} (\lambda h - o(h)) + P_{j+1}(t) (1 - \lambda h - o(h))$$

$$P_{j+1}(t+h) - P_{j+1}(t) = -\lambda h P_{j+1}(t) + \lambda h \frac{e^{-\lambda t} (\lambda t)^j}{j!} \\ - o(h) \left(\frac{e^{-\lambda t} (\lambda t)^j}{j!} + P_{j+1}(t) \right)$$

dividindo os dois lados por h e tomando o limite com $h \downarrow 0$, tem-se:

$$\lim_{h \downarrow 0} \frac{P_{j+1}(t+h) - P_{j+1}(t)}{h} = \lim_{h \downarrow 0} \frac{-\lambda h P_{j+1}(t)}{h} \\ + \lim_{h \downarrow 0} \frac{\lambda h \frac{e^{-\lambda t} (\lambda t)^j}{j!}}{h} + \lim_{h \downarrow 0} \frac{o(h)}{h}$$

Obtendo a seguinte equação diferencial:

$$P'_{j+1}(t+h) + \lambda P_{j+1}(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^j}{j!}$$

A solução dessa equação diferencial é dada por:

$$P_{j+1}(t) = \exp\left(-\int_0^t \lambda dx\right) \left[\int_0^t \frac{\lambda e^{-\lambda x} (\lambda x)^j}{j!} \exp\left(\int_0^x \lambda dy\right) dx + k \right]$$

$$P_{j+1}(t) = e^{-\lambda t} \left[\int_0^t \frac{\lambda (\lambda x)^j}{j!} dx + k \right] = e^{-\lambda t} \left[\frac{\lambda^{j+1}}{j!} \left(\frac{x^{j+1}}{j+1} \right) \Big|_0^t + k \right]$$

$$P_{j+1}(t) = e^{-\lambda t} \left[\left(\frac{\lambda^{j+1} t^{j+1}}{j!(j+1)} \right) + k \right]$$

$$P_{j+1}(0) = e^{-\lambda \cdot 0} \left[\frac{\lambda^{j+1}}{j!} \left(\frac{0^{j+1}}{j+1} \right) + k \right] = k = 0$$

$$P_{j+1}(t) = e^{-\lambda t} \left(\frac{\lambda^{j+1} t^{j+1}}{j!(j+1)} \right) = \frac{e^{-\lambda t} (\lambda t)^{j+1}}{(j+1)!}$$

Dessa forma, pelo princípio da indução matemática, $P_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!}$ para todo $j \in \mathbb{N}$.

De grande interesse em processos de Poisson é a distribuição do tempo até a primeira ocorrência. Seja T_0 esse tempo. Observe que o evento $[T_0 > t]$ é equivalente ao evento $[X_t = 0]$. Então $P(T_0 > t) = P(X_t = 0) = e^{-\lambda t}$ para $t > 0$. Note que $P(T_0 \leq t) = 1 - e^{-\lambda t}$, mas é a função de distribuição acumulada de T_0 , e também a acumulada de uma exponencial de parâmetro λ . Portanto, $T_0 \sim \text{exponencial}(\lambda)$.

Pode-se provar que se T_i é o tempo entre a i -ésima e a $(i+1)$ -ésima ocorrência, tem-se que T_1, T_2, \dots é uma sequência de variáveis aleatórias independentes e identicamente distribuídas de acordo a uma exponencial de parâmetro λ . Como consequência disso, $W_n = T_1 + T_2 + \dots + T_n$, tempo da n -ésima ocorrência, tem distribuição Gama com parâmetros n e λ . Observe que $\mathbb{E}(T_i) = \frac{1}{\lambda}$, o recíproco da taxa de ocorrência. Essa relação é bastante intuitiva, assumamos por exemplo que $\lambda = 10$ ocorrências por hora, então $\mathbb{E}(T_i) = \frac{1}{\lambda}$ horas.

2.1.2 Processos de Nascimento

Uma generalização do processo de Poisson descrito anteriormente, é permitir que o parâmetro, λ , dependa do estado em que o processo se encontra. A definição do processo de Poisson satisfaz:

- $P(X_{t+h} - X_t = 1 | X_t = i) = \lambda h + o(h)$ quando $h \downarrow 0$;
- $P(X_{t+h} - X_t = 0 | X_t = i) = 1 - \lambda h + o(h)$ quando $h \downarrow 0$;
- $X_0 = 0$.

A nova definição para o caso do processo de nascimento satisfaz:

- $P(X_{t+h} - X_t = 1 | X_t = j) = \lambda_j h + o(h)$ quando $h \downarrow 0$;
- $P(X_{t+h} - X_t = 0 | X_t = j) = 1 - \lambda_j h + o(h)$ quando $h \downarrow 0$;
- $X_0 = 0$.

Os parâmetros λ_j são chamados taxas de ocorrência quando o processo está no estado j . Procedendo de forma análoga ao caso em que λ não depende do estado, para $j = 0$:

$$P_0(t+h) = P_0(t)[P(X_{t+h} - X_t = 0 | X_t = 0) + o(h)]$$

$$P_0(t+h) = P_0(t)[1 - \lambda_0 h + o(h)]$$

A partir daí, $P_0(t+h) - P_0(t) + \lambda_0 h P_0(t) = 0$. Lembrando que $P_0(0) = 1$, obtem-se:

$$P_0(t) = e^{-\lambda_0 t}$$

Para $j = 1$:

$$P_1(t+h) = P_1(t)P(X_{t+h} - X_t = 0 | X_t = 1) + P_0(t)P(X_{t+h} - X_t = 1 | X_t = 0) + o(h)$$

$$P_1(t+h) = P_1(t)[1 - \lambda_1 h + o(h)] + P_0(t)[1 - \lambda_0 h + o(h)] + o(h)$$

Daí, obtem-se que $P_1(t+h) - P_1(t) = \lambda_1 h P_1(t) + \lambda_0 h P_0(t) + o(h)$.

Dividindo por h e fazendo $h \downarrow 0$:

$$\lim_{h \downarrow 0} \frac{P_1(t+h) - P_1(t)}{h} = \lim_{h \downarrow 0} [\lambda_1 P_1(t) + \lambda_0 P_0(t) + o(h)]$$

$$P_1'(t) + \lambda_1 P_1(t) = \lambda_0 e^{-\lambda_0 t}$$

Desde que $P_1(0) = 0$, $P_1(t) = e^{-\lambda_1 t} \int_0^t \lambda_0 e^{-(\lambda_1 - \lambda_0)s} ds$. Suponha conhecer $P_j(t)$, para obter $P_{j+1}(t)$:

$$P_{j+1}(t+h) = P_{j+1}(t)P(X_{t+h} - X_t = 0 | X_t = j+1) + P_j(t)P(X_{t+h} - X_t = 1 | X_t = j) + o(h)$$

$$P_{j+1}(t+h) = P_{j+1}(t)[1 - \lambda_{j+1} h + o(h)] + P_j(t)[1 - \lambda_j h + o(h)] + o(h)$$

Daí, obtem-se que $P_{j+1}(t+h) - P_{j+1}(t) + \lambda_{j+1} h P_{j+1}(t) = \lambda_j h P_j(t) + o(h)$.

Partindo da equação anterior: $P_{j+1}(t+h) - P_{j+1}(t) + \lambda_{j+1} h P_{j+1}(t) = \lambda_j h P_j(t) + o(h)$. Tem-se:

$$P_{j+1}'(t) + \lambda_{j+1} P_{j+1}(t) = \lambda_j P_j(t)$$

Resolvendo a equação diferencial:

$$P_{j+1}(t) = e^{-\lambda_{j+1}t} \int_0^t \lambda_j e^{-(\lambda_j-1)s} P_j(s) ds$$

Observa-se que um processo de nascimento é não decrescente. Neste caso, então não existe distribuição invariante.

2.1.3 Processos de Poisson Não Homogêneos

O processo X_t com $t \geq 0$ é um processo de Poisson não homogêneo se:

- i) X_t tem saltos de tamanho um;
- ii) a distribuição de $X_{t+s} - X_t$ é independente de t .

Seja $a(t) = \mathbb{E}(X_t)$ e observe que $a(t)$ é crescente pois:

$$X_{t+s} \geq X_t \implies \mathbb{E}(X_{t+s}) \geq \mathbb{E}(X_t) \implies a(t+s) \geq a(t)$$

A função $a(t)$ é chamada de taxa do processo. Suponha que $a(t)$ é contínua e defina: $\tau(u) = a^{-1}(u)$, a inversa de a .

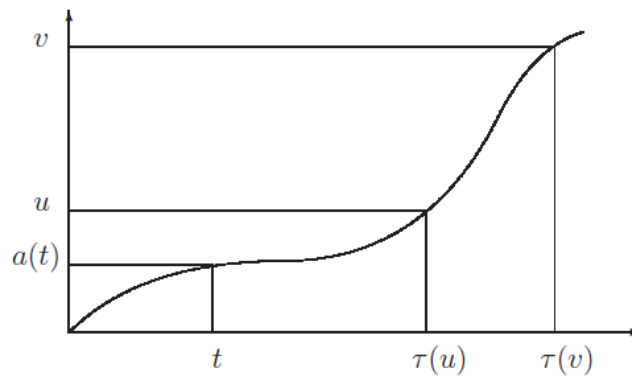


Figura 1: Processo de Poisson não homogêneo com taxa $a(t)$

Seja X_t com $t \geq 0$ é um processo de Poisson não homogêneo com taxa $a(t)$.

Defina o processo $M_t = X_\tau(t)$. Então M_t é um processo de Poisson com taxa 1. Observe que $\mathbb{E}(M_t) = \mathbb{E}(X_\tau(t)) = a[\tau(t)] = t$.

Seja X_t com $t \geq 0$ é um processo de Poisson não homogêneo com taxa $a(t)$ e seja $b(t, s) = a(t+s)a(t)$, então:

- i) os incrementos têm distribuição $P(X_{t+s} - X_t = k) = \frac{e^{-b(t,s)} b(t,s)^k}{k!}$;
- ii) se T_n é o tempo do n -ésimo evento do processo, então $P(T_{n+1} - T_n > t | T_1, T_2, \dots, T_n) = e^{-\int_{T_n}^{T_n+t} a(t) dt}$,

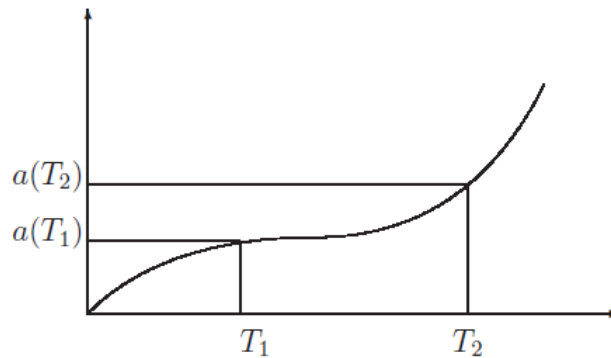


Figura 2: Processo de Poisson não homogêneo com taxa $a(t)$

2.2 MODELOS DE SÉRIES TEMPORAIS

Os modelos utilizados para descrever séries temporais são processos estocásticos. Existe um vasto número de modelos distintos para descrever o comportamento de uma série particular. Uma série temporal é dita estacionária quando ela se desenvolve no tempo aleatoriamente em torno de uma média constante. Na prática, a maioria das séries apresentam algum tipo de não estacionariedade, por exemplo, tendência.

Uma série pode ser estacionária por períodos curtos ou longos. A classe de modelos ARIMA, em geral, é sempre capaz de descrever de maneira satisfatória

séries estacionárias e séries não estacionárias sem um comportamento explosivo. Este formato de não estacionariedade é usualmente denominado homogêneo, quando a série pode ser estacionária, transitando ao redor de um nível, por um determinado período, posteriormente mudar de nível e flutuar em torno de um novo nível e assim por diante.

A maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, portanto, muitas vezes é necessário transformar os dados originais se estes não formam uma série estacionária. As transformações mais usuais consistem em tomar diferenças sucessivas da série original, até se obter uma série estacionária. A primeira diferença $Z(t)$ é definida por:

$$\Delta Z(t) = Z(t) - Z(t - 1)$$

então, a segunda diferença é:

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t - 1)]$$

$$\Delta^2 Z(t) = Z(t) - 2Z(t - 1) + Z(t - 2).$$

De modo geral, a n -ésima diferença de $Z(t)$ é:

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)].$$

Na maioria dos cenários, uma ou duas diferenças são suficientes para que a série se torne estacionária. Estudaremos modelos paramétricos para séries temporais, ou seja, modelos que o número de parâmetros é finito. Nos modelos paramétricos para séries temporais, a análise é feita no domínio do tempo. Os mais frequentemente usados são os modelos de regressão, os modelos autorregressivos, de médias móveis (ARMA) e os modelos autorregressivos integrados e de médias móveis (ARIMA).

Para descrever o comportamento de séries com erros autocorrelacionados utiliza-se modelos ARMA. Esses modelos são úteis nesses casos, uma vez que o fato de serem autocorrelacionados influencia na evolução do processo. Há três casos particulares de modelos ARIMA que serão abordados:

- i) processo autorregressivo de ordem p : AR(p);
- ii) processo de médias móveis de ordem q : MA(q)
- iii) processo autorregressivo e de médias móveis de ordens p e q : ARMA(p,q).

Tais processos podem ser generalizados de maneira adequada pelos chamados modelos autorregressivo integrados e de médias móveis de ordem p , d e q : ARIMA(p,d,q), em que d representa a inclusão de um operador de diferenças. É possível descrever os três processos através do ARIMA, por exemplo, um processo autorregressivo de ordem p , AR(p), pode ser escrito como ARIMA($p,0,0$).

Esta metodologia consiste em ajustar modelos autorregressivos integrados de médias móveis, ARIMA(p,d,q), a um conjunto de dados. A concepção do modelo segue um algoritmo em que a escolha da estrutura do modelo é baseada nos próprios dados, seguindo os seguintes passos:

- i) Considera-se uma classe geral de modelos para a análise;
- ii) identifica-se um modelo com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- iii) estima-se os parâmetros do modelo identificado;
- iv) executa-se a verificação da adequação do modelo ajustado aos dados através de uma análise de resíduos.
- v) Caso o modelo não seja adequado o algoritmo é repetido, voltando à fase de identificação.

Existem vários critérios para identificação de um modelo, por isso, é possível identificar modelos diferentes dependendo do critério que foi escolhido para identificação.

2.2.1 Modelos de Erro ou de Regressão

Uma série temporal observada pode ser escrita da seguinte forma:

$$Z_t = f(t) + a_t, \quad t = 1, 2, \dots, N, \text{ em que } f(t) \text{ é chamado sinal e } a_t \text{ ruído.}$$

Aqui, o sinal $f(t)$ é uma função do tempo completamente determinada e a_t é uma sequência aleatória, independente de $f(t)$. Supõe-se que as variáveis aleatórias a_t sejam não correlacionadas, tenham média nula e variância constante, isto é,

$$\mathbb{E}[a_t] = 0, \forall t, \quad \mathbb{E}[a_t^2] = \sigma_a^2, \forall t, \quad \mathbb{E}[a_t a_s] = 0, s \neq t.$$

A série a_t com as características acima é conhecida como ruído branco.

2.2.2 Modelos AR

Afirma-se que $X_t, t \in Z$ é um processo autorregressivo de ordem p $X_t \sim AR(p)$ se é possível escrever o processo na seguinte forma:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

em que $\phi_0, \phi_1, \dots, \phi_p$ são parâmetros reais e ε_t é i.i.d. com $\mathbb{E}(\varepsilon_t) = 0$ e $Var(\varepsilon_t) = \sigma^2$.

Um caso particular simples de grande importância é o processo autorregressivo de ordem $p = 1$, AR(1)

$$X_t = \phi X_{t-1} + \varepsilon_t$$

Fazendo substituições sucessivas, X_{t-1}, X_{t-2} , etc, na equação anterior obtem-

se:

$$X_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \quad (1)$$

em que a convergência é em média quadrática. Portanto, a condição $|\phi| < 1$ é suficiente para que X_t seja estacionário.

É possível calcular a função de autocovariância (f.a.c.v.) multiplicando ambos os lados da equação (1) por $X_{t-\tau}$ e tomando a esperança

$$E(X_t X_{t-\tau}) = \phi E(X_{t-1} X_{t-\tau}) + \phi^2 E(X_{t-2} X_{t-\tau}) + \dots + \phi^p E(X_{t-p} X_{t-\tau}) + E(\varepsilon_t X_{t-\tau})$$

$$\gamma_\tau = \phi\gamma_{\tau-1} = \dots = \phi^\tau \gamma_0$$

Considerando que $\gamma_0 = \sigma_X^2 = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1 - \phi^2}$, segue que

$$\gamma_\tau = \frac{\sigma^2}{1 - \phi^2} \phi^{|\tau|}, \quad \tau \in Z$$

Dessa forma, da função de autocovariância obtêm-se a função de autocorrelação (f.a.c.)

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \phi^{|\tau|}, \quad \tau \in Z$$

No caso geral, calcula-se as funções de autocovariância e autocorrelação por:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, \quad j > 0$$

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \quad j > 0, \quad (2)$$

respectivamente.

Para $j = 1, 2, \dots, p$, na equação (2) obtém-se:

$$\rho_1 = \phi_1 + \phi_2\rho_2 + \dots + \phi_p\rho_{p-1}$$

$$\rho_2 = \phi_1\rho_1 + \rho_2 + \dots + \phi_p\rho_{p-2}$$

$$\vdots$$

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p$$

que é um método recursivo para calcular as autocorrelações, conhecidas como equações de Yule-Walker.

2.2.3 Modelos MA

Considere um processo linear $\{X_t, t \in Z\}$, este processo será dito ser de médias móveis de ordem q , denotado por MA(q), se satisfizer a equação de diferenças:

$$X_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

em que $\mu, \theta_1, \dots, \theta_q$ são constantes reais, a_t é i.i.d. e $a_t \sim N(0, \sigma^2)$.

Segue que X_t é estacionária, de média μ . Suponha que $\mu = 0$, calculando a função de autocovariância(f.a.c.v.) do processo obtém-se:

$$\gamma_j = \mathbb{E}(X_t X_{t-j}) = \mathbb{E} \left\{ \left[a_t - \sum_{k=1}^q \theta_k a_{t-k} \right] \left[a_{t-j} - \sum_{l=1}^q \theta_l a_{t-l-j} \right] \right\}$$

$$\gamma_j = \mathbb{E}(a_t a_{t-j}) - \sum_{k=1}^q \mathbb{E}(a_{t-j} a_{t-k}) - \sum_{l=1}^q \mathbb{E}(a_t a_{t-j-l}) + \sum_{k=1}^q \sum_{l=1}^q \theta_k \theta_l \mathbb{E}(a_{t-l} a_{t-j-l})$$

Como os a_t são não correlacionados:

$$\gamma_a(j) = \mathbb{E}(a_t a_{t-j}) = \begin{cases} \sigma_a^2, & j = 0, \\ 0, & j \neq 0 \end{cases}$$

$$\gamma_0 = \text{Var}(X_t) = \sigma_X^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2)$$

Logo:

$$\gamma_j = \gamma_a(j) - \sum_{k=1}^q \gamma_a(k-j) - \sum_{l=1}^q \gamma_a(j+l) + \sum_{k=1}^q \sum_{l=1}^q \theta_k \theta_l \mathbb{E}(a_{t-l} a_{t-j-l}) \gamma_a(j+l+k)$$

Portanto, a f.a.c.v. pode ser definida como:

$$\gamma_j = \begin{cases} \sigma_a^2 \left(\theta_j + \sum_{l=1}^{q-l} \theta_l \theta_{j+l} \right) & , j = 0, \\ \sigma_a^2 (\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots + \theta_q \theta_{q-j}) & , j = 1, \dots, q, \\ 0 & , j > q. \end{cases}$$

Assim, da f.a.c.v. obtem-se a f.a.c

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \begin{cases} \frac{\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots + \theta_q \theta_{q-j}}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & , j = 1, \dots, q, \\ 0 & , j > q. \end{cases}$$

2.2.4 Modelos ARMA

Modelos autorregressivos e de médias móveis é uma junção de modelos AR e MA. Denotados por ARMA(p,q) um processo autorregressivo e de médias móveis de ordem (p, q) pode ser representado por

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

No caso geral, calcula-se a f.a.c.v. por:

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p} + \gamma_{Xa}(j) - \theta_1\gamma_{Xa}(j-1) - \cdots - \theta_q\gamma_{Xa}(j-q)$$

em que

$$\gamma_{Xa}(j)$$

é a covariância entre X_t e a_t , definida por:

$$\gamma_{Xa}(j) = \begin{cases} = 0, & j > 0, \\ \neq 0, & j \leq 0. \end{cases}$$

Assim, a f.a.c.v. fica:

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p}, \quad j > q$$

e a f.a.c. é obtida por:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \cdots + \phi_p\rho_{j-p}, \quad j > q.$$

2.2.5 Modelos ARIMA

Os modelos AR, MA e ARMA, são apropriados para descrever séries temporais estacionárias, isto é, séries que se desenvolvem no tempo ao redor de uma média constante. Modelos autorregressivos, integrados e de médias móveis (ARIMA) representam séries diferenciadas por um modelo ARMA, aplicados em situações iniciais de não estacionariedade.

Seja Z_t uma série temporal não estacionária. Seja $W_t = \Delta Z_t = Z_t - Z_{t-1}$ sendo a série diferenciada uma vez de Z_t , denota-se por $W_t^d = \Delta^d Z_t$ a série temporal diferenciada d vezes de Z_t .

Pode-se representar W_t por um modelo ARMA, como W_t é uma diferença de Z_t então, Z_t é uma integral de W_t , assim se diz que Z_t segue um modelo autorregressivo, integrado e de médias móveis, isto é, um modelo ARIMA de ordem (p,d,q) e escreve-se ARIMA (p,d,q) em que p é a ordem da componente autorregressiva, d é o número de diferenças tomadas na série e q é a ordem da componente de médias móveis. Dessa forma, os modelos anteriores podem ser descritos como segue:

i) ARIMA $(p,0,0) = AR(p)$;

ii) ARIMA $(0,0,q) = MA(q)$;

iii) ARIMA $(p,0,q) = ARMA(p,q)$.

e no caso de uma série não estacionária utilizamos o modelo completo, ARIMA (p,d,q) com d diferenças na série original.

2.3 TEORIA DAS FILAS

A Teoria das Filas estuda as relações entre as demandas e atrasos dos usuários do sistema, por meio de modelos matemáticos. A formação de filas ocorre quando a demanda é maior que a capacidade de oferecer determinado serviço em um período de tempo. Assim, o objetivo da Teoria das Filas é encontrar um balanceamento adequado entre os custos de oferecer serviços no sistema e os custos dos atrasos dos usuários sofridos pelo sistema (ARENALES; ARMENTANO; MORABITO, 2007).

Os modelos de filas utilizados, possuem fórmulas que indicam como o sistema funcionará, ou seja, a distribuição dos tempos entre chegadas, tempos de atendimento, o número de servidores e demais informações sobre a fila. Com essas informações, podem ser calculadas as medidas de desempenho, por exemplo, tempo médio de espera na fila, tempo médio de espera no sistema, entre outras (HILLIER; LIEBERMAN, 2013).

Através dos resultados das medidas de desempenho obtidas pela utilização da teoria das filas, adquire-se o conhecimento do sistema, e por meio dele, é possível buscar um equilíbrio apropriado entre custo de serviço e tempo de espera.

Um sistema de filas é composto pelos clientes, pelo local de espera, ou seja, em que as filas são formadas e pelos canais de atendimento.

O processo básico de filas é: clientes chegam a um sistema para solicitar um determinado tipo de serviço, entram na fila caso os canais de atendimentos estejam cheios. Em seguida, à medida que os atendentes finalizam os serviços dos que já estavam em atendimento, novos clientes são selecionados de acordo com algum critério, e, por fim, saem do sistema, quando o atendimento é finalizado. Esse processo está representado na figura 3.

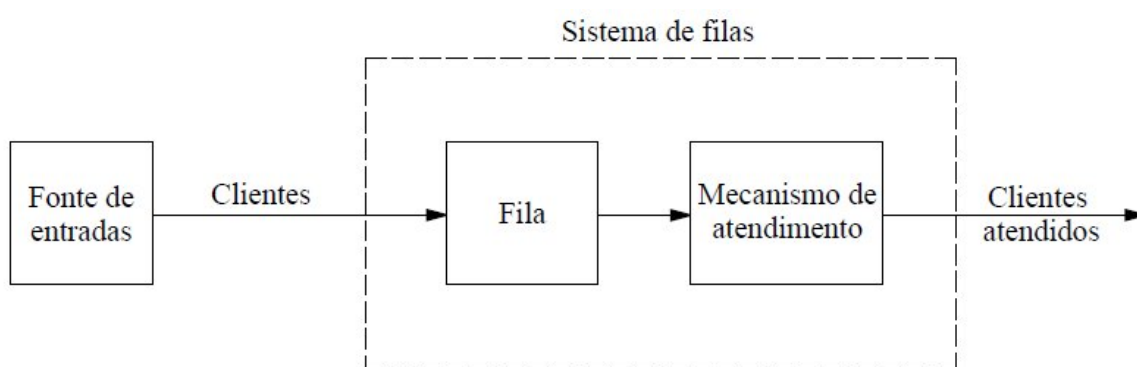


Figura 3: Processo de filas básico. (Modificado de Hillier e Lieberman (2013))

Segundo Arenales, Armentano e Morabito (2007), o sistema de filas é descrito com base em três informações fundamentais. São elas: a entrada ou processo de chegada de usuários no sistema, disciplina da fila (ou ordem de atendimento dos usuários em fila) e a saída ou processo de serviço (atendimento).

Geralmente, o número de chegadas e o instante de tempo que elas ocorrem são desconhecidos. Quando isso ocorre, eles são ditos aleatórios e são modelados por distribuições de probabilidade.

Nas próximas subseções, serão discutidos os processos de chegada e de

atendimento, bem como as distribuições de probabilidades mais utilizadas para modelagem deles.

2.3.1 Formas de Chegadas

O conhecimento das formas de chegadas no sistema é de grande importância para o estudo das filas, pois é um dos fatores que influencia na maneira que a fila é formada.

Segundo Hillier e Lieberman (2013), uma característica da fonte de entradas é o seu tamanho, isto é, o número total de possíveis clientes distintos. Essa população de onde provêm às chegadas pode ser de tamanho finito ou infinito (e a fonte de chegadas é dita limitada ou ilimitada). Normalmente, quando não especificada, admite-se que a população é infinita, mesmo que o tamanho dela seja finito relativamente grande, para facilitar os cálculos. A hipótese finita só é elaborada quando a taxa de chegadas seja significativamente afetada pelo número de clientes no sistema.

O processo de chegadas é explicado como o intervalo de tempo entre chegadas sucessivas de usuários. Geralmente, admite-se que esse processo não se altera ao longo do tempo, e que não é afetado pelo número de usuários presentes no sistema (ARENALES; ARMENTANO; MORABITO, 2007).

Como já mencionado, quando o processo de chegadas ocorre de forma aleatória, é modelado por uma distribuição de probabilidade. Geralmente, o número de clientes gerado até um determinado momento tem distribuição de Poisson, ou seja, as chegadas ocorrem a uma taxa média fixa, independentemente de quantos clientes já se encontrarem no sistema. Equivalentemente, pode-se dizer que os tempos entre chegadas seguem uma distribuição Exponencial.

2.3.2 Formas de Atendimento

Em um sistema de filas básico, existe um ou mais canais de atendimento paralelos, onde os clientes são atendidos completamente pelos servidores. De acordo

com Hillier e Lieberman (2013), a maioria dos modelos de filas partem do pressuposto de uma instalação de atendimento com um atendente ou com um número finito de atendentes.

O tempo de atendimento é descrito como o tempo de serviço por cliente, em que o servidor pode ser um ou mais indivíduos (ou até mesmo máquinas) que realize um determinado tipo de serviço simultaneamente (ARENALES; ARMENTANO; MORABITO, 2007).

Considera-se que somente um cliente pode ser atendido por um servidor no mesmo instante. Também, admite-se que o processo de atendimento não varia com o tempo, e nem é afetado pelo número de clientes presentes no sistema (ARENALES; ARMENTANO; MORABITO, 2007).

Devido à aleatoriedade dos tempos de serviço, eles são modelados por uma distribuição de probabilidade. Usualmente, é assumida a mesma distribuição para todos os atendentes. Segundo Hillier e Lieberman (2013), a distribuição dos tempos de serviço mais utilizada é a distribuição Exponencial, em grande parte, por ser bem mais tratável.

2.3.3 Disciplina da Fila

Segundo Arenales, Armentano e Morabito (2007), a disciplina da fila é definida como a ordem que os usuários na fila são selecionados para atendimento.

As disciplinas mais comuns são: FCFS (*First Come, First Served*), ou seja, o primeiro a chegar é o primeiro a ser servido, e LCFS (*Last Come, First Served*), ou seja, o último a chegar é o primeiro a ser servido. Há disciplinas que levam em consideração algum tipo de prioridade (ARENALES; ARMENTANO; MORABITO, 2007). Geralmente, quando não é especificada, assume-se que a disciplina da fila é FCFS (ABENSUR, 2011).

2.3.4 Medidas de Desempenho no Sistema de Filas

Através da teoria das filas, é possível calcular as medidas de desempenho, isto é, índices que permitem a avaliação da eficiência de um modelo de filas adotado. São apresentadas a seguir, as variáveis que irão denotar essas medidas, as probabilidades e taxa relacionadas ao sistema:

- L : Número esperado de indivíduos no sistema;
- L_q : Número esperado de indivíduos na fila (desconsidera os que estão sendo atendidos);
- W : Tempo médio de espera no sistema;
- W_q : Tempo médio de espera na fila (desconsidera o tempo de atendimento);
- $P(0)$: Probabilidade do sistema estar ocioso;
- $P(n)$: Probabilidade do sistema ter n indivíduos;
- ρ : Taxa de utilização do servidor (medida de congestionamento do servidor).

De acordo com Hillier e Lieberman (2013), as medidas de desempenho são válidas quando o sistema atingir uma condição de estado estável, isto é, decorrido determinado tempo, o estado do sistema torna-se independente do estado inicial. Em outras palavras, quando o estado do sistema é estável, a distribuição probabilística do estado do sistema permanece a mesma ao longo do tempo.

As fórmulas relacionadas às notações apresentadas, serão exibidas para os modelos de filas nas subseções referentes a eles, com exceção das fórmulas de probabilidades. Elas serão omitidas porque não serão utilizadas nesse trabalho, visto que o interesse é estimar a taxa do tempo de serviço no restaurante universitário.

2.3.5 Distribuição Exponencial

É uma distribuição de probabilidades de variáveis contínuas, muito utilizada para analisar experimentos com tempo de duração variável e sem interrupções, por exemplo, o tempo de vida útil de lâmpadas, entre outras aplicações.

Em Teoria das Filas, é a distribuição mais importante, utilizada para modelar os tempos entre chegadas dos usuários ao sistema, assim como os tempos de atendimentos (HILLIER; LIEBERMAN, 2013).

Uma variável aleatória X tem distribuição Exponencial com o parâmetro λ , $\lambda > 0$, se tiver função densidade de probabilidade dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{c.c.} \end{cases},$$

em que x é o valor que a variável aleatória assume e λ é a taxa de ocorrências dessa variável. A média e a variância de X são, respectivamente,

$$\mathbb{E}(X) = \frac{1}{\lambda} \text{ e } \text{Var}(X) = \frac{1}{\lambda^2},$$

A distribuição Exponencial possui a propriedade de perda de memória, ou seja, o futuro é independente do passado. Essa propriedade pode ser ilustrada por:

$$P(X > t + \Delta t \mid X > \Delta t) = P(X > t),$$

em que, segundo Hillier e Lieberman (2013), para quaisquer valores positivos de t e Δt , “a distribuição probabilística do tempo remanescente até o evento (chegada ou término do atendimento) ocorrer é sempre a mesma, independente de quanto tempo (Δt) já tiver passado”.

2.3.6 Distribuição Poisson

É uma função de probabilidade muito utilizada para modelar número de ocorrências de evento por um certo período de tempo. A aplicação dessa função de probabilidade em teoria das filas é relacionada às formas de chegada dos indivíduos ao sistema de filas, ou seja, através dela pode-se obter a probabilidade de ocorrências de n chegadas, por exemplo, em um período de tempo especificado.

Uma variável aleatória X tem distribuição Poisson com parâmetro λ , $\lambda > 0$, se tiver função de probabilidade dada por:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots,$$

em que, λ é a taxa média de ocorrência do evento e x é o número de ocorrências. Para utilizar essa distribuição, é necessário atender às seguintes suposições:

- a taxa média de ocorrência λ deve ser constante ao longo do tempo;
- a informação sobre o número de ocorrências em um período nada revela sobre o número de ocorrências em outro período.

De acordo com Hillier e Lieberman (2013), se as chegadas são ditas aleatórias, significa que elas ocorrem de acordo com um processo de entrada de Poisson, isto é, independentemente de quando ocorreu a última chegada, a chance de uma nova chegada é a mesma durante um determinado período de tempo.

2.3.7 Processo de Nascimento e Morte

No contexto da teoria das filas, nascimento refere-se à chegada de um novo cliente no sistema e morte refere-se à saída do cliente atendido. Portanto, no modelo de filas, tem-se um processo de nascimento e morte quando a chegada de indivíduos no sistema ocorre a uma certa taxa (genericamente, λ), e a saída após o atendimento deles ocorre a uma taxa (genericamente, μ).

Considere $N(t)$ o número de clientes no sistema de filas no instante t . Segundo Hillier e Lieberman (2013), o processo de nascimento e morte descreve probabilisticamente como $N(t)$ muda à medida que t aumenta.

Geralmente, considera-se que os intervalos de tempos entre chegadas e os tempos de serviço são distribuídos exponencialmente.

Diante disso, com base nos valores obtidos para as taxas λ e μ , podem ser calculados:

- tempo médio entre as chegadas dos clientes: $\frac{1}{\lambda}$;
- tempo médio de atendimento dos clientes: $\frac{1}{\mu}$.

Além disso, pode-se calcular a intensidade do tráfego de usuários no sistema, dada por $\rho = \frac{\lambda}{\mu}$. De acordo com Soares (2016), a partir do resultado obtido para ρ , pode-se fazer as seguintes observações:

- se $\rho > 1$: há mais chegadas de clientes do que saídas. Logo, o número de clientes é ilimitado e o estado é instável;
- se $\rho < 1$: há mais saídas de clientes do que chegadas. Diante disso, existe uma solução estacionária para o sistema;
- se $\rho = 1$: chega, em média, o mesmo número de clientes que saem do sistema. O número de clientes no sistema é equiprovável e o sistema é instável.

Segundo Hillier e Lieberman (2013), quando os tempos entre chegadas e os tempos de serviço são modelados pela distribuição e se forem mutuamente independentes, o processo de Nascimento e Morte é considerado um tipo especial de cadeia de Markov de tempo contínuo.

2.3.8 Modelos de Filas

Para simplificar a análise dos sistemas de filas, os modelos de filas mais simples são nomeados, utilizando a notação de Kendall (1953). Segundo Arenales, Armentano e Morabito (2007), “esta notação considera sistemas de fila única com um ou mais servidores idênticos em paralelo”. A notação é apresentada abaixo:

$$A/B/c/K/N/Z,$$

em que,

A: distribuição de probabilidade de intervalos de tempo entre chegadas sucessivas;

B: distribuição de probabilidade dos tempos de atendimento;

c: capacidade de servidores;

K: indica o número máximo de usuários no sistema (usuários em serviço e fila);

N: tamanho da população de usuários;

Z: disciplina da fila.

Os três últimos termos podem ser omitidos, ao utilizar essa notação. Quando isso acontece, significa dizer *K* é ilimitado, *N* é infinito e *Z* corresponde à disciplina FCFS.

A partir da classificação do modelo de fila, há o interesse de calcular as medidas de desempenho: número esperado de usuários no sistema e na fila (L e L_q , respectivamente), tempo médio de espera no sistema e na fila (W e W_q , respectivamente).

São apresentados nas subseções seguintes, os modelos de filas mais básicos. O termo *M* presente na notação, significa distribuição Exponencial (para tempos entre chegadas e/ou atendimento).

2.3.8.1 Modelo $M/M/1$

É o modelo de filas mais simples, que possui tempos de chegada e de serviços Markovianos (ou seja, têm distribuição Poisson e Exponencial, respectivamente), apenas um servidor e capacidade infinita.

Para esse modelo de filas, as medidas de desempenho: número esperado de indivíduos no sistema (3), número esperado de indivíduos na fila (4), tempo médio de espera no sistema (5) e tempo médio de espera na fila (6) são calculadas utilizando as seguintes equações:

$$L = \frac{\lambda}{\mu - \lambda} \quad (3)$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (4)$$

$$W = \frac{1}{\mu - \lambda} \quad (5)$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}, \quad (6)$$

em que, para todas as equações, λ é a taxa média de chegadas de clientes e μ a taxa média de atendimento.

2.3.8.2 Modelo $M/M/c$

É o modelo de filas, que possui tempos de chegada e de serviços Markovianos (ou seja, têm distribuição Poisson e Exponencial, respectivamente), assim como no modelo $M/M/1$, porém possui c servidores, em que $c > 1$, e capacidade infinita.

As medidas de desempenho: número esperado de indivíduos no sistema (7), número esperado de indivíduos na fila (8), tempo médio de espera no sistema (9)

e tempo médio de espera na fila (10) para filas $M/M/c$ são calculadas através das equações:

$$L = L_q + \frac{\lambda}{\mu} \quad (7)$$

$$L_q = \frac{P_0(\lambda/\mu)^2 \rho}{s!(1-\rho)^2} \quad (8)$$

$$W = W_q + \frac{1}{\mu} \quad (9)$$

$$W_q = \frac{L_q}{\lambda}, \quad (10)$$

em que, para todas as equações, λ é a taxa média de chegadas de clientes, μ a taxa média de atendimento, o c refere-se ao número de servidores. O fator de utilização do sistema (ρ) e a probabilidade do sistema ficar ocioso (P_0), são dados pelas expressões 11 e 12:

$$\rho = \frac{\lambda}{c\mu} \quad (11)$$

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^c}{c!} \left(\frac{1}{1 - \lambda/(c\mu)} \right) \right]^{-1}. \quad (12)$$

Com base nessas medidas, pode-se obter informações sobre o desempenho do sistema em estudo (HILLIER; LIEBERMAN, 2013).

3 RESULTADOS E DISCUSSÕES

A partir da disponibilização do banco de dados do Restaurante Universitário da UFOP, foi feita a amostragem. Assim, o banco de dados resultante é composto por 5 amostras por dia da semana, em que essas informações referem-se ao período de janeiro a abril do ano de 2018. Ressalta-se que será estudado somente o período do almoço.

Posteriormente às alterações no banco de dados, foi possível a utilização de ferramentas estatísticas, tais como estudo descritivo, para o conhecimento do Restaurante Universitário da UFOP e o ajuste de séries temporais, para a estimação das taxas associadas ao processo de filas na entrada do restaurante. Os resultados foram apresentados em tabelas e gráficos.

O termo “serviço” frequentemente citado nessa seção, refere-se à atividade passar pela catraca. Portanto, o “tempo de serviço” foi calculado pela diferença dos horários de duas passagens pela catraca de entrada no Restaurante Universitário da UFOP. Os resultados dessas diferenças são dados em segundos.

Neste capítulo serão apresentados o estudo descritivo dos dados, o ajuste de séries temporais e a discussão dos resultados obtidos.

3.1 ANÁLISES DESCRITIVAS DOS TEMPOS DE SERVIÇO

Foram calculadas as seguintes estatísticas: média, variância e coeficiente de variação, e os resultados dispostos em tabelas. Posteriormente, para a visualização da distribuição dos dados, foram apresentados gráficos dos tipos *boxplots* e histogramas.

Tabela 1: Estatísticas descritivas gerais

Média	Variância	Coeficiente de Variação
8,8137	204,0124	162,0577

A tabela 1 contém a média, variância e coeficiente de variação dos tempos de

serviço, independentemente do dia da semana e do horário.

Por meio da tabela 1, nota-se que o tempo médio de serviço é de aproximadamente 9 segundos e a variância é 204. O coeficiente de variação é 162%, o que indica alta dispersão dos tempos de serviço em relação à média.

Posteriormente, essas estatísticas foram calculadas por dia da semana. Os resultados e os tamanhos amostrais referentes à cada dia estão apresentados na tabela 2.

Tabela 2: Estatísticas descritivas por dia da semana

Dia da Semana	Média	Variância	Coeficiente de Variação	Tamanho da Amostra
Segunda	8,5926	252,9871	185,1084	4467
Terça	9,7951	229,9318	154,8070	4041
Quarta	8,9981	215,3026	163,0692	4297
Quinta	7,8059	116,4236	138,2285	5286
Sexta	9,1714	222,7672	162,7381	4352

Observa-se na tabela 2, que os tempos médios de serviço são aproximadamente 9 segundos, com exceção da quinta-feira, que possui menor média.

Em relação à variação do tempo de serviço, observa-se valores entre 215 e 250, exceto na quinta-feira, onde se percebe uma menor variabilidade.

Os coeficientes de variação de quarta e sexta são próximos do coeficiente geral apresentado na tabela 1. O menor valor é o de quinta, o que pode indicar menor dispersão dos tempos de serviço desse dia em relação aos demais dias da semana.

3.1.1 Divisão do Horário de Funcionamento em Intervalos

Para compreensão do comportamento dos tempos de serviço, o período de funcionamento foi dividido em intervalos de 15 minutos. Na tabela 3, são exibidos os limites de cada intervalo, denominado faixa.

Tabela 3: Limites dos intervalos de tempo

Faixa	Limite Inferior	Limite Superior
1	10: 45: 00	10: 59: 59
2	11: 00: 00	11: 14: 59
3	11: 15: 00	11: 29: 59
4	11: 30: 00	11: 44: 59
5	11: 45: 00	11: 59: 59
6	12: 00: 00	12: 14: 59
7	12: 15: 00	12: 29: 59
8	12: 30: 00	12: 44: 59
9	12: 45: 00	12: 59: 59
10	13: 00: 00	13: 14: 59
11	13: 15: 00	13: 29: 59

Em seguida, foram obtidas as estatísticas descritivas por faixa. Os resultados são apresentados na tabela 4.

Tabela 4: Estatísticas descritivas por faixa

Faixa	Média	Variância	Coefficiente de Variação	Tamanho da Amostra
1	3,5500	4,5103	59,8236	40
2	9,5139	296,4954	180,9879	1940
3	10,3792	262,4949	156,0980	2152
4	8,2808	164,1705	154,7296	2724
5	7,7065	154,1426	161,1023	2934
6	7,2072	130,2030	158,3240	3104
7	7,9489	154,8289	156,5370	2820
8	8,4972	203,6948	167,9640	2641
9	9,9584	242,7500	156,4548	2285
10	11,7982	282,4640	142,4510	1566
11	13,6878	509,3512	164,8831	237

Observa-se na tabela 4 que, com exceção da faixa 1, a maioria das faixas possuem média entre 8 e 9 segundos, isto é, aproximam-se da média geral, apresentada na tabela 1.

Em relação à variância, percebe-se que a maioria dos valores encontrados estão entre 130 e 300. As faixas 4 a 7 possuem valores menores que o geral. A maior variância é 509, referente à faixa 11. Nesse intervalo de tempo, ocorre o fechamento do Restaurante Universitário da UFOP, conseqüentemente, o número de pessoas é menor nessa faixa.

Quanto ao coeficiente de variação, percebe-se que o menor valor é da primeira faixa, ou seja, há menor dispersão dos tempos de serviço em relação à média para esse intervalo, ao contrário da faixa 2. A partir da faixa 3, os coeficientes são próximos, o que indica dispersão semelhante dos tempos de serviço em relação à média, independentemente da faixa.

Nota-se que as faixas 4 a 8 possuem maiores tamanhos de sub-amostras, isto é, o Restaurante Universitário da UFOP é mais frequentado no período de 11 : 30 à 12 : 45.

3.1.2 Estatísticas Descritivas por Dia da Semana

São apresentados através de tabelas, as estatísticas descritivas dos tempos de serviço de todos os intervalos de tempo por dia da semana e os tamanhos das sub-amostras por faixa. Esse último também é ilustrado por gráficos de barras.

3.1.2.1 Segunda-Feira

A tabela 5 contém as estatísticas descritivas e os tamanhos das sub-amostras referentes à segunda-feira.

Tabela 5: Estatísticas descritivas e tamanho da sub-amostra por faixa - segunda-feira

Faixa	Média	Variância	Coeficiente de Variação	Tamanho da sub-amostra
1	-	-	-	-
2	9,0533	707,3325	293,7697	338
3	11,0247	334,9004	165,9936	405
4	7,4825	173,3634	175,9666	601
5	7,1055	165,8926	181,2668	635
6	6,5695	179,5358	203,9578	683
7	8,4636	205,3317	169,3062	522
8	7,6252	154,6522	163,0896	595
9	11,6443	298,0696	148,2669	388
10	12,9048	315,7630	137,6989	273
11	13,5556	273,9487	122,1005	27

Observando-se tabela 5, se nota que os tempos médios de serviço das faixas 4 a 6 (que correspondem aos horários entre 11 : 30 e 12 : 15) são próximos de 7 segun-

dos. Para essas faixas, observam-se maiores tamanhos de sub-amostras, sobretudo para a faixa 6. O maior tempo médio é o da última faixa, que pode ter como causa os intervalos maiores entre os tempos que as pessoas chegam no Restaurante Universitário da UFOP, já que nesse horário o restaurante apresenta um menor número de usuários.

A maior variância é referente à faixa 2, em seguida, às faixas 3, 9 e 10. Nota-se que no horário de 11 : 30 a 12 : 15 (faixas 4 a 6) a variabilidade é semelhante.

O maior coeficiente de variação refere-se à faixa 2, o que indica que os tempos de serviço nesse intervalo são mais dispersos da média, do que os das demais faixas. Os demais coeficientes, exceto os das faixas 6 e 11, são próximos.

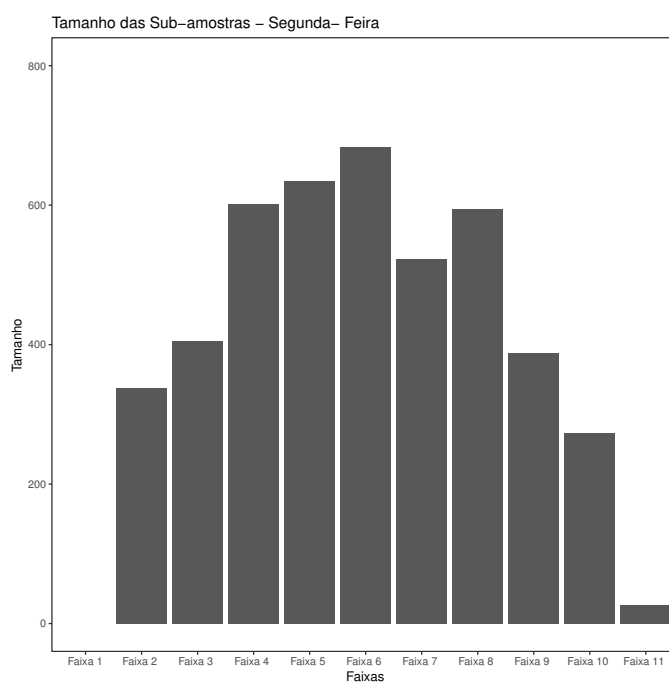


Figura 4: Tamanhos das sub-amostras - segunda-feira

A figura 4 mostra os tamanhos das sub-amostras referentes à segunda-feira. Percebe-se que a faixa 1 não possui registros. Nota-se também que as faixas mais frequentes são 4, 5, 6 e 8. Já a última faixa é menos frequente, pois esse intervalo é próximo do fechamento do restaurante.

3.1.2.2 Terça-Feira

A tabela 6 contém as estatísticas descritivas e os tamanhos das sub-amostras para terça-feira. Observa-se na tabela 6 que, ao contrário do que ocorre na segunda-feira, as faixas 3 a 9 têm tempos médios de serviço semelhantes, aproximadamente 9 segundos. Já o maior tempo médio é o da faixa 11, assim como ocorre na segunda.

Tabela 6: Estatísticas descritivas e tamanho da sub-amostra por faixa - terça-feira

Faixa	Média	Variância	Coeficiente de Variação	Tamanho da sub-amostra
1	3,4444	2,9673	50,0107	18
2	10,0296	210,9275	144,8054	406
3	9,3251	192,1576	148,6543	483
4	9,6098	178,3410	138,9668	469
5	9,5574	232,8272	159,6522	470
6	9,5608	173,1144	137,6175	469
7	9,4119	196,7959	149,0498	471
8	9,2635	204,5604	154,3961	482
9	9,6843	296,5477	177,8190	472
10	12,8939	430,5819	160,9319	264
11	15,8919	786,1547	176,4324	37

Em relação à variância, percebe-se que, das faixas 2 a 8, os valores são aproximadamente 200. A maior variabilidade refere-se à faixa 11, próximo de 790, que pode ocorrer devido à grande diferença de tempo entre as chegadas consecutivas ao Restaurante Universitário da UFOP nesse período.

Os maiores coeficientes de variação referem-se às faixas 9 e 11, ou seja, os tempos de serviço dessas faixas são mais dispersos das médias. Os demais, exceto o da faixa 1, são próximos, isso indica que a dispersão dos dados em torno da média de cada faixa é semelhante.

Através do gráfico na figura 5, nota-se que os tamanhos das sub-amostras das faixas 3 a 9 são semelhantes. Essas faixas correspondem ao intervalo de 11 : 15 a 13 : 00, ou seja, o restaurante é bem frequentado praticamente durante todo o horário

de funcionamento nesse dia da semana.

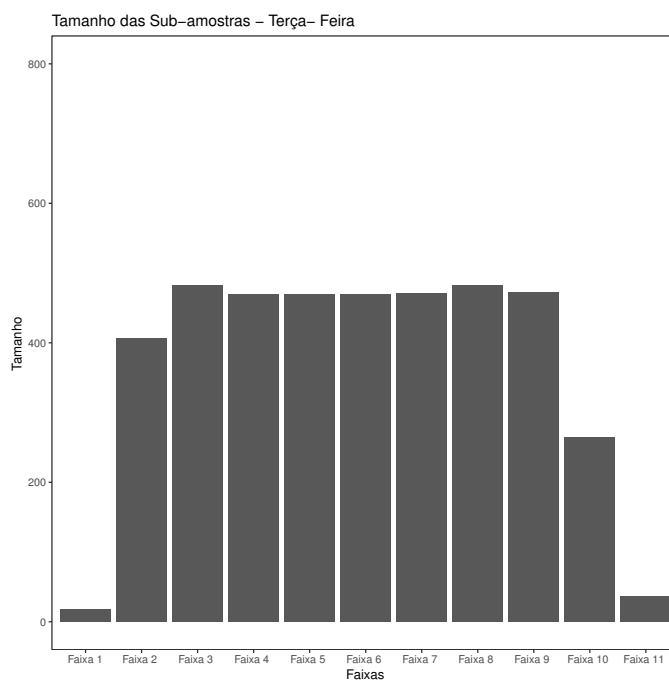


Figura 5: Tamanhos das sub-amostras - terça-feira

3.1.2.3 Quarta-Feira

A tabela 7 contém as estatísticas descritivas e os tamanhos das sub-amostras referentes à quarta-feira.

Tabela 7: Estatísticas descritivas e tamanho da sub-amostra por faixa - quarta-feira

Faixa	Média	Variância	Coefficiente de Variação	Tamanho da sub-amostra
1	-	-	-	-
2	10,5349	226,1912	142,7605	344
3	11,2537	375,0028	172,0762	402
4	8,7628	174,5100	150,7527	506
5	7,7729	144,3422	154,5662	590
6	7,7346	112,5016	137,1325	569
7	9,0886	231,6233	167,4538	508
8	7,3002	184,5536	186,0924	613
9	9,8261	276,1483	169,1183	460
10	11,7454	336,0868	156,0840	271
11	8,8529	98,3717	112,0334	34

Através da tabela 7, nota-se que os menores tempos médios de serviço são das faixas 5, 6 e 8. Já os maiores, referem-se às faixas 3 e 10, de aproximadamente 11 segundos.

A menor variação ocorre na faixa 11, aproximadamente 99. As maiores variabilidades são das faixas 3 e 10. Percebe-se que as faixas 4 e 7 possuem variâncias muito diferentes apesar da semelhança dos tamanhos das sub-amostras.

O maior coeficiente de variação refere-se à faixa 8, ou seja, os dados dessa faixa têm maior dispersão em relação à média. Os coeficientes das faixas 4, 5 e 10 são próximos, o que indica dispersão semelhante dos dados em relação à média delas.

Através do gráfico na figura 6, observa-se que as maiores frequências são das faixas 4 a 8. Ao contrário do que ocorre na segunda, a maior frequência refere-se à faixa 8, equivalente ao período de 12 : 30 a 12 : 45.

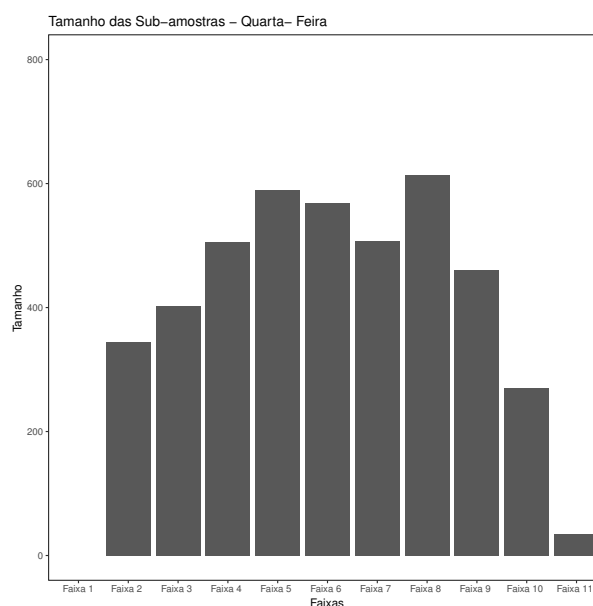


Figura 6: Tamanhos das sub-amostras - quarta-feira

3.1.2.4 Quinta–Feira

Observa-se na tabela 8 que os tempos médios de serviço das faixas 4 a 7 estão entre 6 e 7 segundos. São menores quando comparados às mesmas faixas nos

demais dias da semana. A maior média refere-se à faixa 11, aproximadamente 29 segundos.

Tabela 8: Estatísticas descritivas e tamanho da sub-amostra por faixa - quinta-feira

Faixa	Média	Variância	Coefficiente de Variação	Tamanho da sub-amostra
1	3,6364	5,9567	67,1175	22
2	8,9526	134,4626	129,5247	485
3	9,0542	109,3230	115,4795	498
4	6,6259	54,0841	110,9912	679
5	6,6588	64,3908	120,5082	677
6	5,6220	62,5662	140,6945	799
7	6,8647	65,5662	117,9548	658
8	8,7758	146,8227	138,0729	513
9	8,4737	154,2761	146,5808	532
10	11,3376	209,8781	127,7803	394
11	28,8966	2015,0961	155,3466	29

As maiores variabilidades são das faixas 9 a 11, essa última chega a 2015. Isso ocorre devido a uma grande diferença entre os tempos que os usuários chegam no Restaurante Universitário da UFOP, pois essa última faixa tem menor frequência.

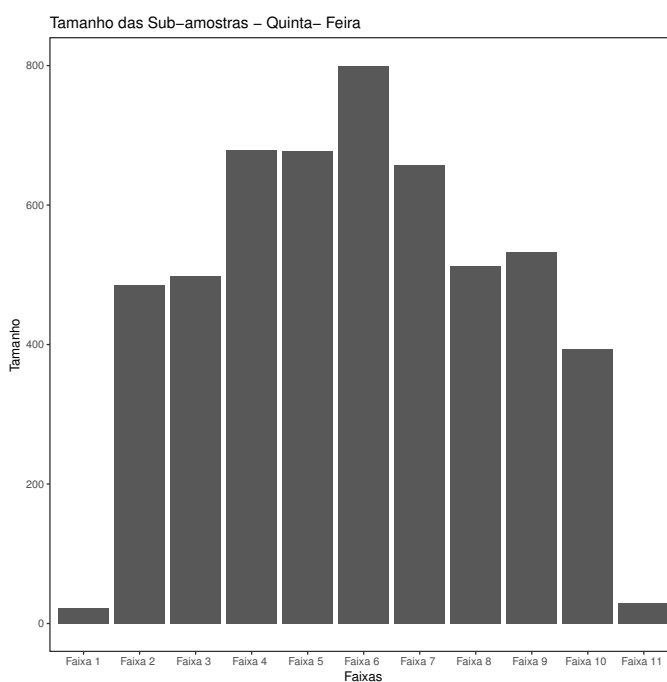


Figura 7: Tamanhos das sub-amostras - quinta-feira

Os coeficientes de variação das faixas 2 a 10 são próximos, o que pode indicar uma dispersão semelhante dos dados em relação à média delas. Nota-se que os

valores referentes às faixas 2, 3, 4, 5 e 7 são menores se comparados aos demais dias da semana.

Através da figura 7, observa-se que as faixas 4 a 7 possuem as maiores frequências. Esses intervalos também são mais frequentes quando comparados às mesmas faixas dos demais dias da semana.

3.1.2.5 Sexta-Feira

A tabela 9 contém as estatísticas descritivas e os tamanhos das sub-amostras referentes à sexta-feira.

Tabela 9: Estatísticas descritivas e tamanho da sub-amostra por faixa - sexta-feira

Faixa	Média	Variância	Coeficiente de Variação	Tamanho da sub-amostra
1	-	-	-	-
2	9,1526	294,2772	187,4280	367
3	11,9066	356,0904	158,4866	364
4	9,8507	278,6614	169,4608	469
5	8,0302	189,9082	171,6101	562
6	7,7175	139,8566	153,2381	584
7	6,7035	110,4180	156,7544	661
8	10,1872	358,0610	185,7475	438
9	10,7113	204,1132	133,3806	433
10	10,7115	188,8009	128,2774	364
11	10,4636	149,4253	116,8233	110

Através da tabela 9, nota-se que a maioria dos tempos médios de serviço são próximos de 10 segundos, o que não ocorre nos demais dias da semana. Observa-se também que o menor tempo médio refere-se à faixa 7, diferentemente dos demais dias da semana, que os menores valores observados eram referentes à faixa 1.

A menor variação é da faixa 7, de aproximadamente 110. Já as maiores variabilidades referem-se às faixas 3 e 8, aproximadamente 358. As demais, são próximas de 200.

A faixa 11 tem menor coeficiente de variação, aproximadamente 117%, o que

indica menor dispersão dos tempos em relação à média dessa faixa. Os maiores coeficientes referem-se às faixas 2 e 8, próximos de 187%.

O gráfico na figura 8 mostra os tamanhos das sub-amostras referentes à sexta-feira.

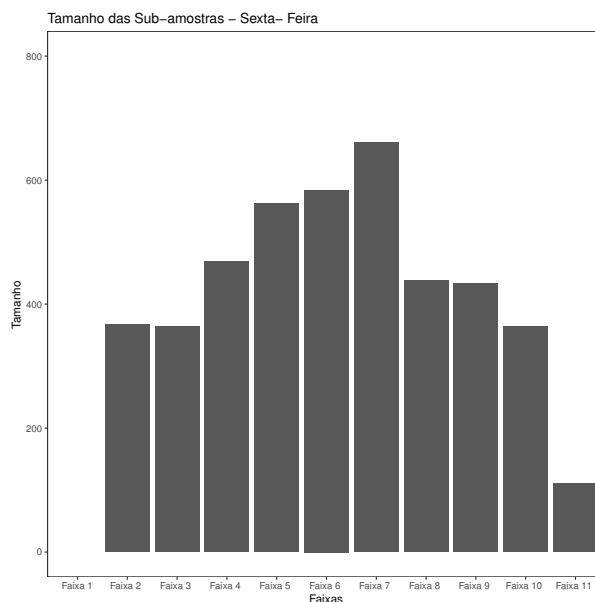


Figura 8: Tamanhos das sub-amostras - sexta-feira

Através do gráfico da figura 8, nota-se que, assim como os demais dias da semana, as faixas 4 a 7 possuem maiores frequências. A faixa 11 possui tamanho de sub-amostra igual a 110, é o maior valor em relação aos demais dias da semana para esse mesmo intervalo.

Nota-se que, em todas as tabelas, as faixas 1 e 11, que correspondem ao período que o restaurante abre e fecha, respectivamente, são pouco frequentes e/ou não possuem informações. Diante disso, elas serão retiradas dos gráficos *boxplots* e histogramas, representados nas próximas subseções.

3.1.2.6 Representações gráficas

São representados na figura 9, os *boxplots* dos tempos de serviço por dia da semana e faixa.

Observa-se no primeiro gráfico da figura 9, que a variabilidade dos tempos de serviço na segunda e quinta é semelhante. Na terça e sexta, percebe-se maiores variabilidades e tempos medianos superiores a 5 segundos.

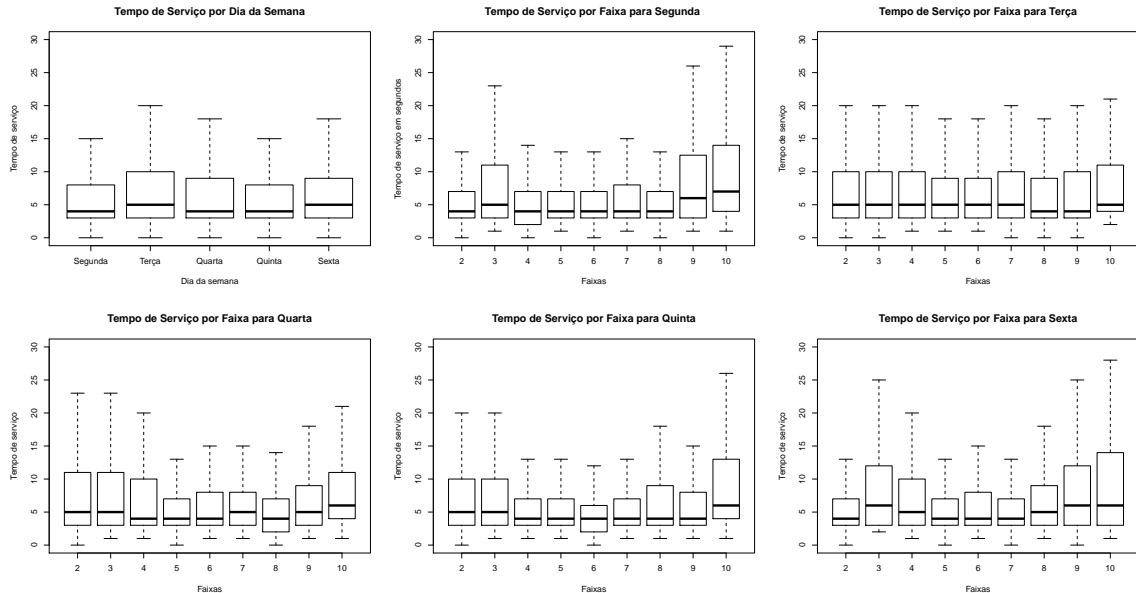


Figura 9: Boxplots dos tempos de serviço por dia e intervalo

Algumas observações relacionadas às faixas são feitas a partir dos gráficos exibidos na figura 9:

- faixa 2: apresenta grande variabilidade na terça, quarta e quinta. Em relação à segunda e sexta, essa faixa apresenta variação semelhante dos tempos de serviço, em que os tempos máximos são menores que 15 segundos;
- faixa 3: possui grandes variações em todos os dias da semana, em que as maiores referem-se à segunda, quarta e sexta. O tempo máximo para essa faixa é aproximadamente 25 segundos, que se refere à sexta;
- faixa 4: a menor variabilidade é observada na quinta. Todos os tempos medianos para essa faixa são próximos de 5 segundos. As variabilidades são semelhantes na terça e sexta. Os tempos máximos nesses dias são aproximadamente 20 segundos;

- faixa 5: possui menor variação do tempo de serviço em todos os dias da semana. Os tempos medianos são menores que 5 segundos, exceto na terça. Os tempos máximos para essa faixa são menores que 15 segundos, com exceção da terça, em que o tempo máximo é próximo de 20 segundos;
- faixa 6: percebe-se que a variabilidade é semelhante na segunda, quarta e sexta. Os tempos medianos são próximos de 5 segundos para todos os dias. Os tempos máximos são menores que 15 segundos, exceto na terça;
- faixa 7: apresenta variabilidade semelhante na quinta e sexta. Os tempos medianos são próximos de 5 segundos. Já os tempos máximos são menores que 15 segundos, exceto para terça, em que o tempo máximo é próximo de 20 segundos;
- faixa 8: possui menor variabilidade na segunda, em que o tempo de serviço máximo é próximo de 15 segundos. Nota-se que as variabilidades mais semelhantes referem-se à terça e quinta. Os tempos medianos para esses dias são próximos de 5 segundos e os tempos máximos são aproximadamente 20 segundos;
- faixa 9: observa-se maiores variabilidades nessa faixa, com exceção de quinta. Nota-se também que os tempos medianos referentes a essas faixas são próximos de 5 segundos para a maioria dos dias da semana. Os tempos máximos dessa faixa variam de aproximadamente 20 a 30 segundos;
- faixa 10: apresenta maiores variabilidades na maioria dos dias. Observa-se também que os tempos medianos referentes a essa faixa são maiores que 5 segundos para a maioria dos dias da semana. Os tempos máximos variam de 20 a 30 segundos.

Na maioria dos dias, notam-se pequenas variações das faixas 5 a 7, e as maiores são observadas nas faixas 2, 3, 4, 9 e 10. Na terça, diferente dos demais dias, para todas as faixas notam-se variabilidades semelhantes, assim como os tempos

medianos e máximos. Um importante passo no estudo das filas consiste em descobrir a distribuição dos tempos de serviço, para identificação do modelo de filas. Para isso, foram construídos os histogramas por dia e faixa.

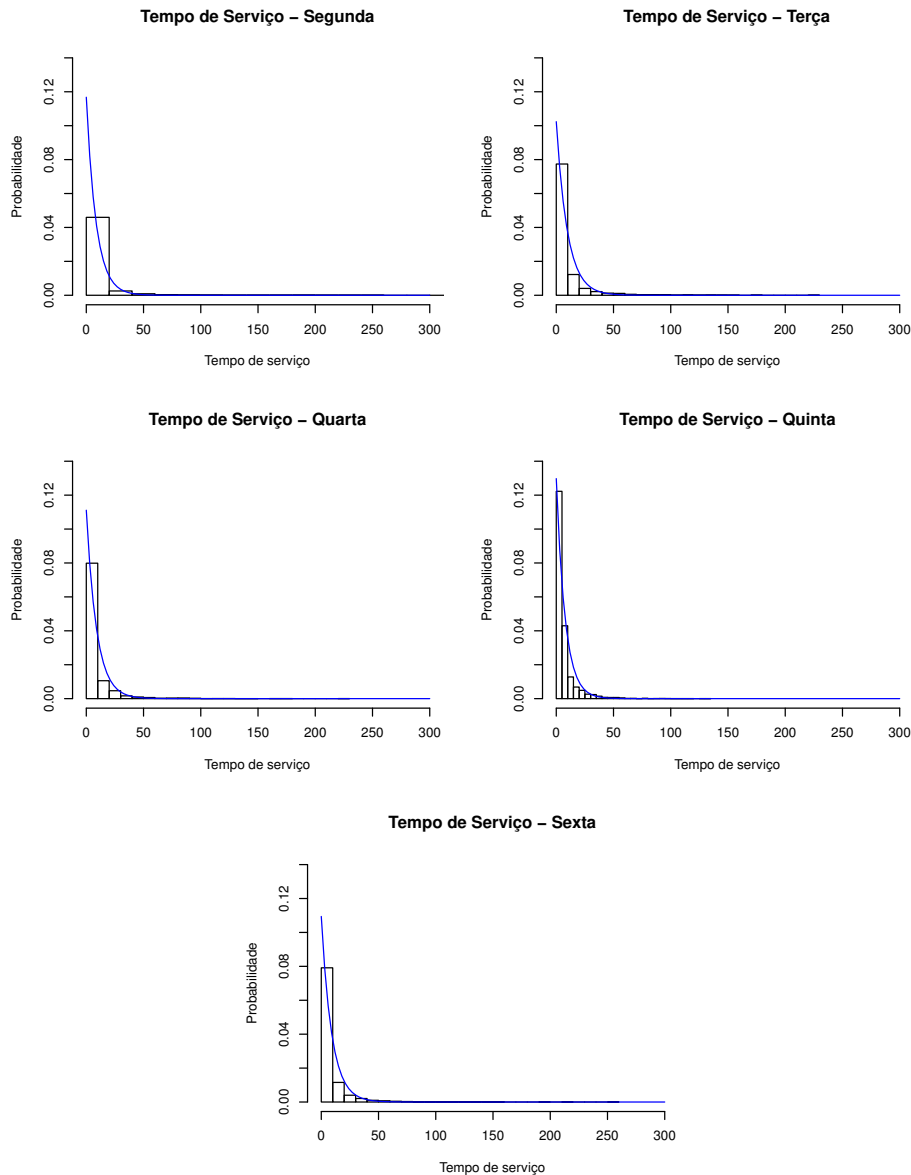


Figura 10: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas

Devido ao fato da distribuição Exponencial ser a mais comum e os gráficos sugerirem tal distribuição, as curvas exponenciais foram ajustadas com base em cada sub-amostra e representadas pela cor azul. São apresentados na figura 10, os histogramas dos tempos de serviço por dia da semana com as curvas exponenciais.

Através dos histogramas apresentados na figura 10, observa-se que os dados

têm comportamento semelhante à curva ajustada. Então, supõe-se que os tempos de serviço de todos os dias da semana seguem uma distribuição Exponencial, em que a taxa de atendimento é estimada com base na amostra de cada dia.

São apresentados a seguir, os histogramas por dia e faixa. Estão representados na figura 11, os histogramas por faixa referentes à segunda-feira.

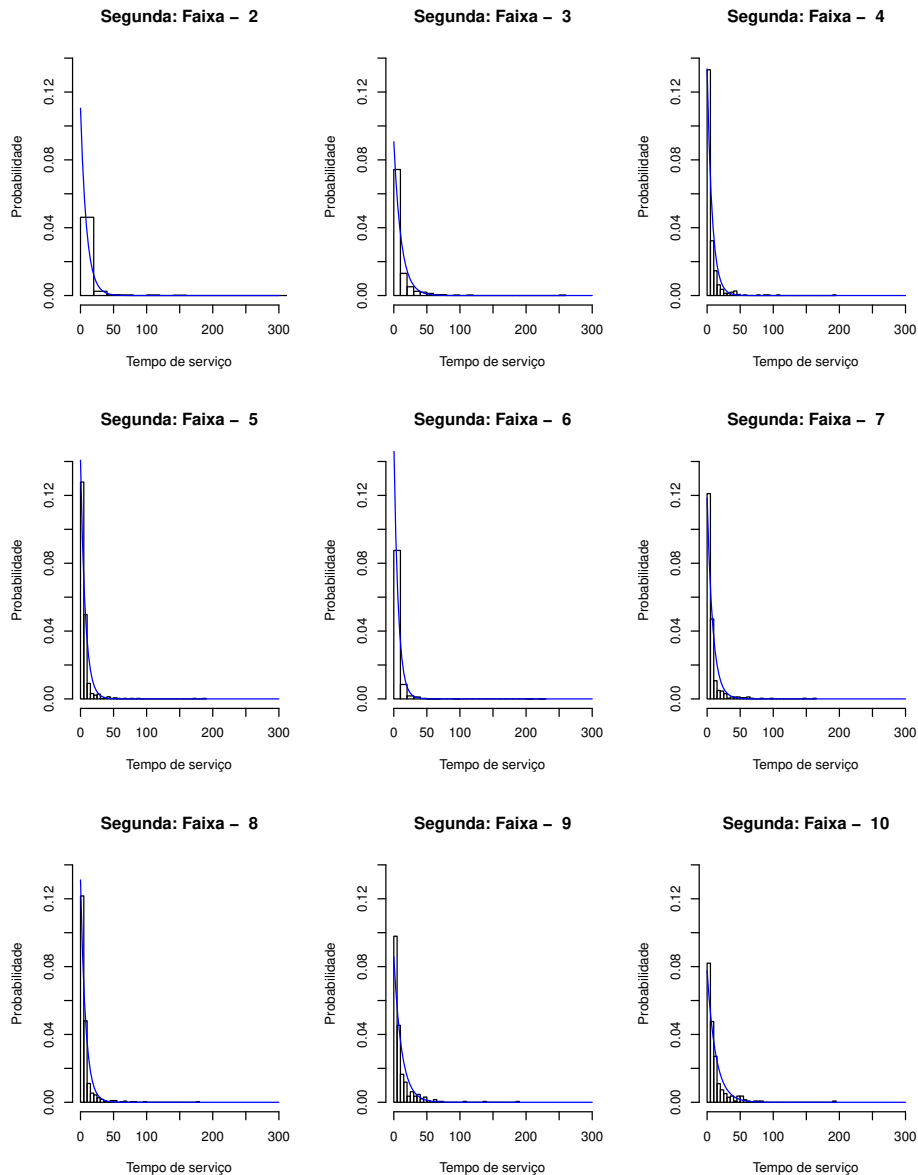


Figura 11: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Segunda-feira

Através dos histogramas apresentados na figura 11, é possível notar que, independentemente da faixa, os tempos de atendimento têm comportamento seme-

lhante à distribuição Exponencial. Portanto, supõe-se que sigam essa distribuição.

São mostrados na figura 12, os histogramas por faixa referentes à terça-feira.

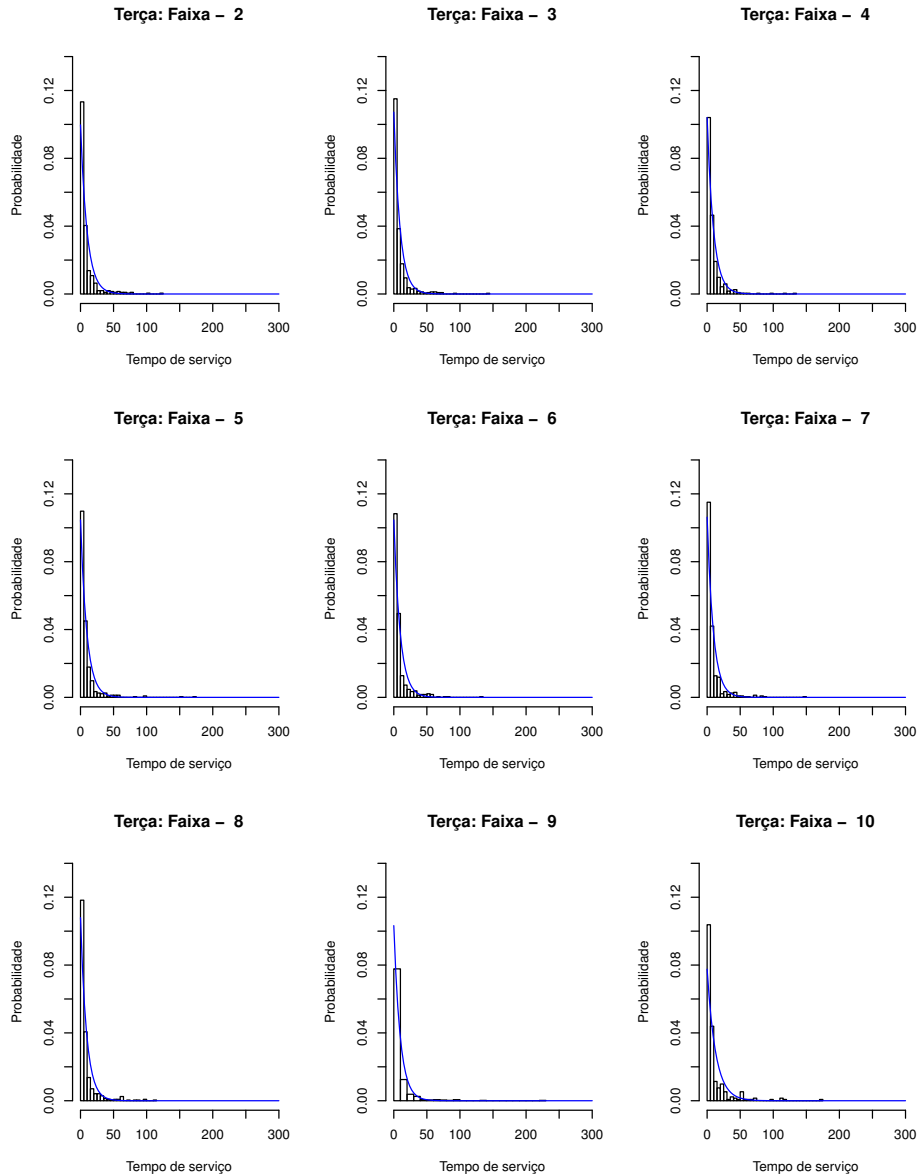


Figura 12: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Terça-feira

Através dos histogramas apresentados na figura 12, nota-se que, independentemente da faixa, os tempos de atendimento aparentemente seguem uma distribuição Exponencial, pois têm comportamento semelhante aos verificados através da representação das curvas ajustadas em azul.

Estão representados na figura 13, os histogramas por faixa referentes à quarta-feira.

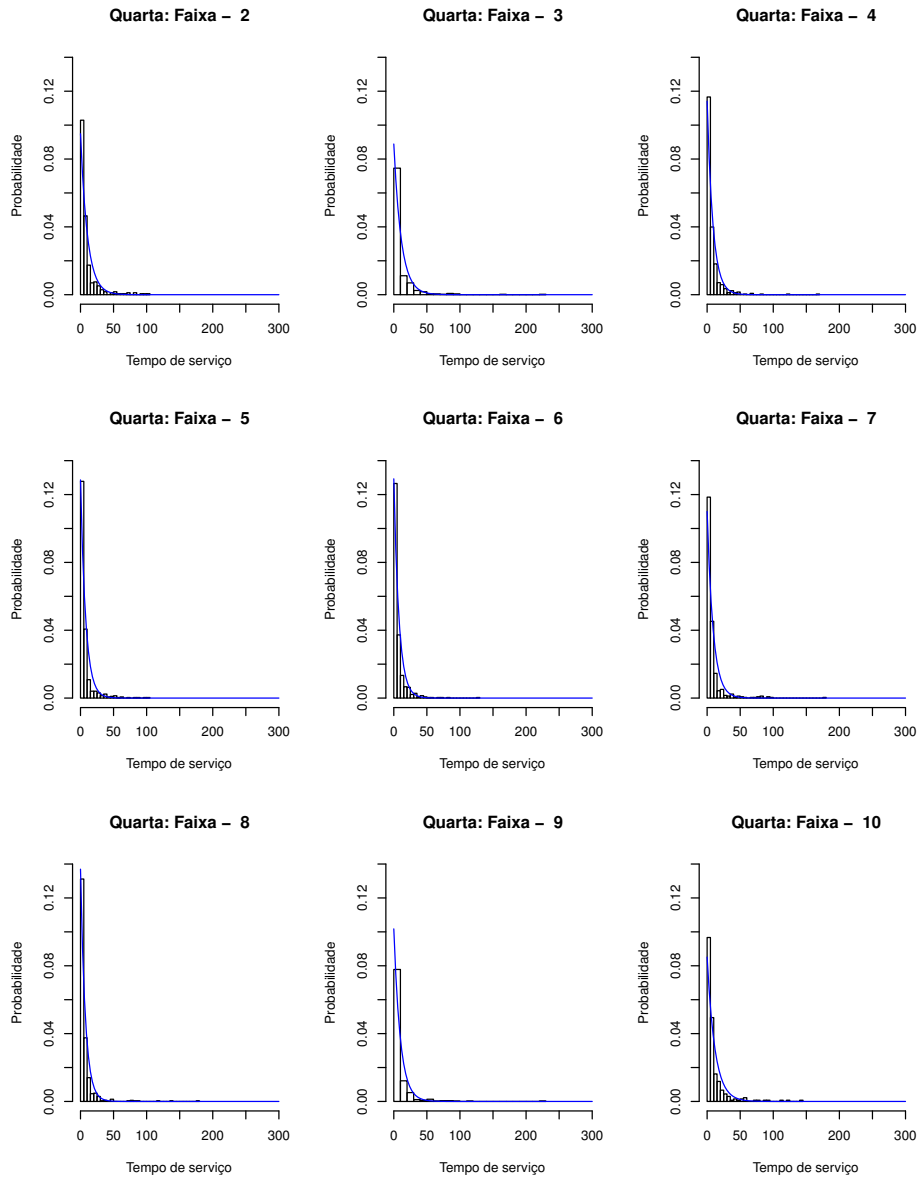


Figura 13: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Quarta-feira

Através dos histogramas apresentados na figura 13, nota-se que, independentemente da faixa, os tempos de atendimento têm comportamento semelhante à distribuição Exponencial. Portanto, supõe-se que sigam essa distribuição.

São mostrados na figura 14, os histogramas por faixa referentes à quinta-feira.

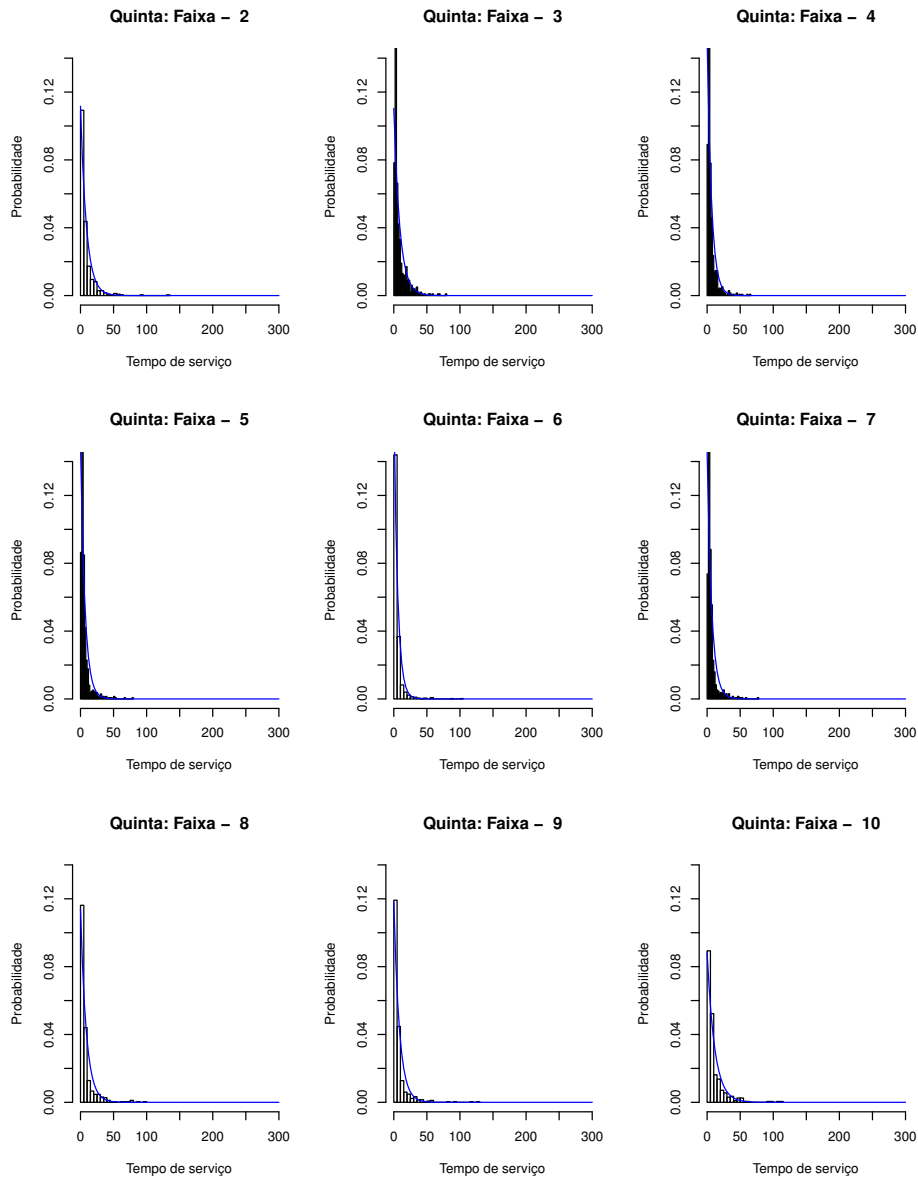


Figura 14: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Quinta-feira

Através dos histogramas apresentados na figura 14, nota-se que, independentemente da faixa, os tempos de atendimento têm comportamento semelhante à distribuição Exponencial. Portanto, supõe-se que sigam essa distribuição.

Estão representados na figura 15, os histogramas por faixa referentes à sexta-feira.

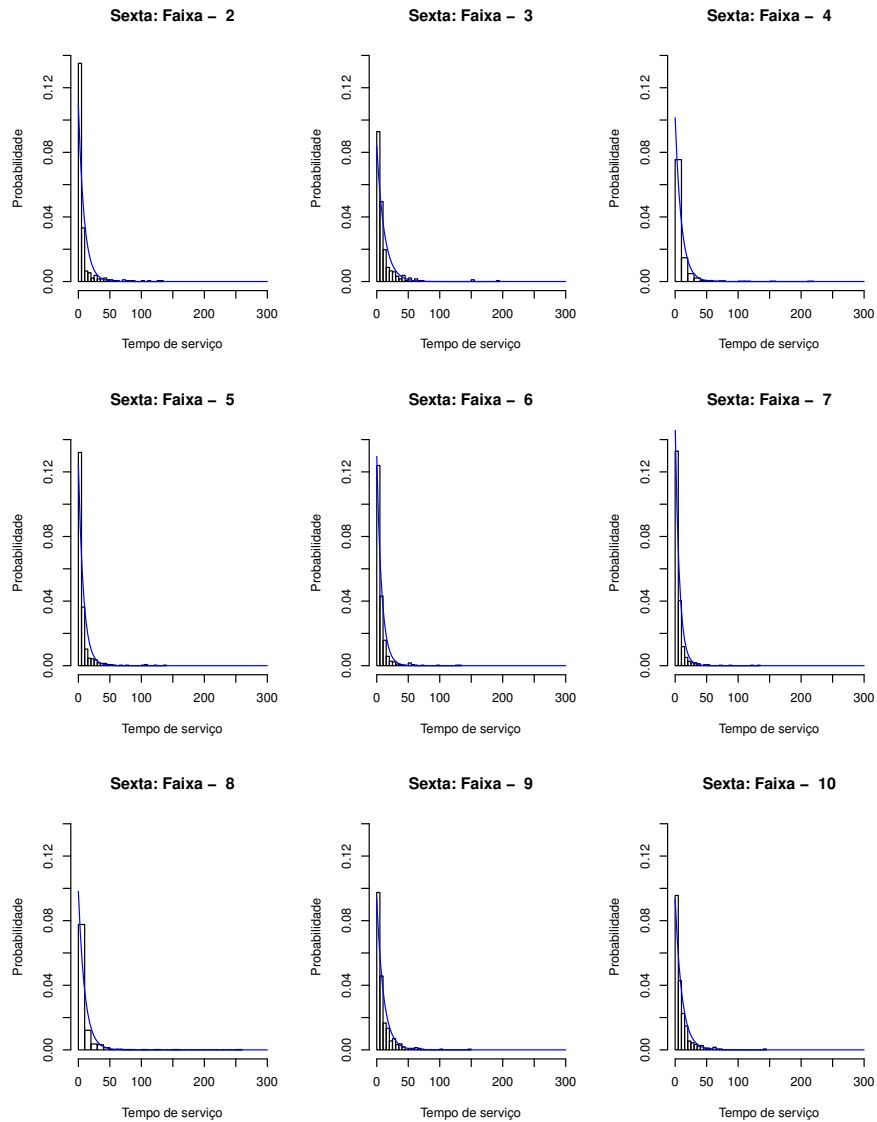


Figura 15: Histogramas dos tempos de serviço com as curvas exponenciais ajustadas - Sexta-feira

Através dos histogramas apresentados na figura 15, nota-se que, independentemente da faixa, os tempos de atendimento têm comportamento semelhante à distribuição Exponencial. Portanto, supõe-se que sigam essa distribuição.

Diante de todas essas análises, uma constatação vem a tona, o procedimento de chegadas para atendimento no Restaurante Universitário da UFOP se comporta como um Processo de Poisson, entretanto as taxas mudam a cada dia e a cada faixa

de horário. Dessa forma, uma verificação bastante relevante, seria tentar descrever as taxas como funções temporais, ou seja, verificar a estrutura de um processo de Poisson não homogêneo. Verificar essa natureza de função temporal, seria na prática, tentar descrever os intervalos entre chegadas de clientes ao Restaurante Universitário da UFOP.

A seção seguinte toma por base todo conjunto de dados, dividido por dias de pesquisa e busca ajustar para cada dia sob investigação um modelo de série temporal para descrever o comprimento do intervalo de tempo entre chegadas como uma função dependente do tempo. A verificação de um grande volume de ajustes bem sucedidos será tratada com indício suficiente para corroborar a tese de que o procedimento de chegadas no Restaurante Universitário da UFOP se comporta como um processo de Poisson não homogêneo.

As tentativas de ajuste de modelos de séries temporais partira de modelos AR, MA e ARMA, mas em muitos casos se concentrará nos modelos ARIMA, isso porque, a considerar as informações teóricas relatadas na seção 2.2 é a escolha mais segura para uma investigação robusta para tais ajustes. Vale ressaltar, que não existia previamente nenhuma informação que garantisse ajustes plenamente bem sucedidos, apenas é possível visualizar que sucesso nesses ajustes são indicativos para corroborar a tese de uma investigação acerca de processos de Poisson não homogêneos.

3.2 AJUSTE DO MODELO DE SÉRIES TEMPORAIS

Serão apresentadas a seguir as tentativas de ajuste de modelos ARIMA. Ressalta-se que há 5 amostras por dia da semana no banco de dados, e, para cada uma delas, foi ajustado um modelo de série temporal para estimar a taxa do tempo de atendimento. Além dos modelos, serão mostrados os gráficos dos resíduos e medidas de acurácia para verificação da qualidade dos ajustes.

As medidas de acurácia serão apresentadas por siglas, que têm os seguintes significados: ME: erro médio (*Mean Error*), RMSE: raiz do erro quadrático médio (*Root*

Mean Squared Error), MAE: erro médio absoluto (*Mean Absolute Error*), MPE: erro médio percentual (*Mean Percentage Error*), MAPE: erro percentual médio absoluto (*Mean Absolute Percentage Error*), MASE: erro escalonado médio absoluto (*Mean Absolute Scaled Error*) e ACF1: autocorrelação de erros no *lag 1* (*Autocorrelation of errors at lag 1*). Essas medidas avaliam o erro. Portanto, menores valores de acurácia indicam bom ajuste.

A investigação será iniciada com dados referentes às segundas-feiras, cada data no conjunto de dados será discutida individualmente, isso visto que o interesse é garantir que qualquer dia sob análise se comporta como um processo de Poisson Não homogêneo e que na prática as variações ao longo dos dias e dentro dos dias se devem a padrão aleatório em função do tempo. Para a série relacionada a esse dia 22/01/2018, uma segunda-feira, estão representados na figura 16 os gráficos da FAC (função de autocorrelação) e FACP (função de autocorrelação parcial).

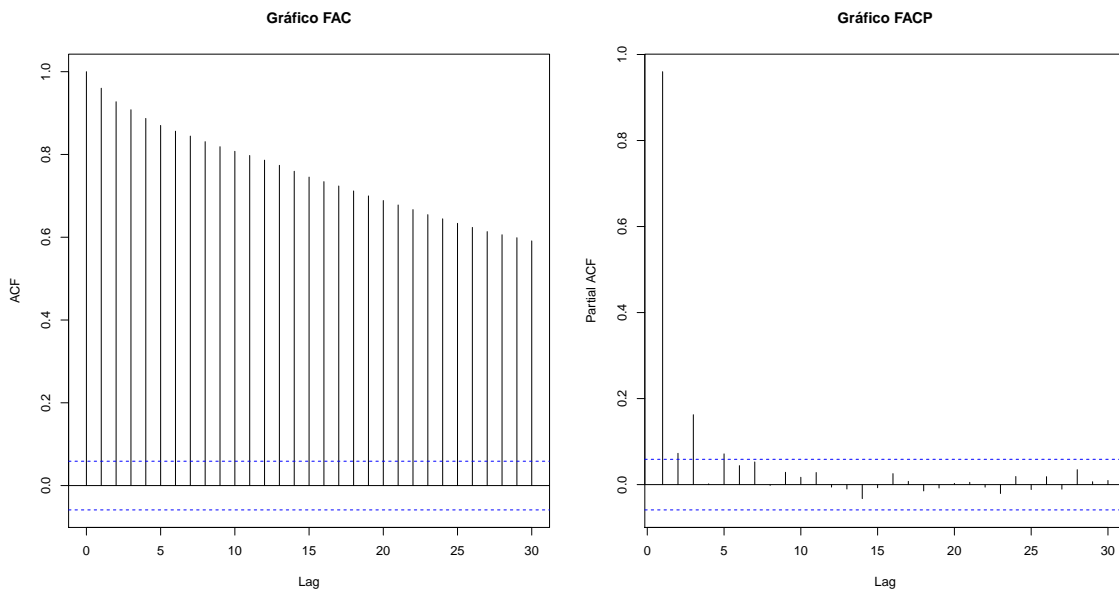


Figura 16: FAC e FACP da série (22/01/2018)

Através da figura 16, nota-se um decaimento exponencial no gráfico da FAC e um pico no *lag 1* no gráfico da FACP. Isso sugere que o modelo deve ser ajustado com termos autorregressivos.

Inicialmente tentou-se o ajuste de um modelo AR(1) que apresentou intercepto não significativo e posteriormente um modelo AR(2) também não adequado. Após algumas tentativas, foi escolhido o modelo ARIMA(1,1,0), em que os critérios adotados foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. O coeficiente do modelo é mostrado na tabela 10.

Tabela 10: Coeficiente e medida AIC (22/01/2018) - modelo ARIMA(1,1,0)

	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,4805	0,0303	0,0000	-9074,0000

Nota-se na tabela 10 que o coeficiente é significativo, com valor-p menor que um nível de significância de 0,05. Além disso, observa-se que o valor de AIC é muito baixo, portanto, segundo este critério, o modelo está bem ajustado. Posteriormente, verificou-se a função de autocorrelação e função de autocorrelação parcial para tentar detectar possíveis relações entre os resíduos. Os resultados são exibidos nos gráficos da figura 17.

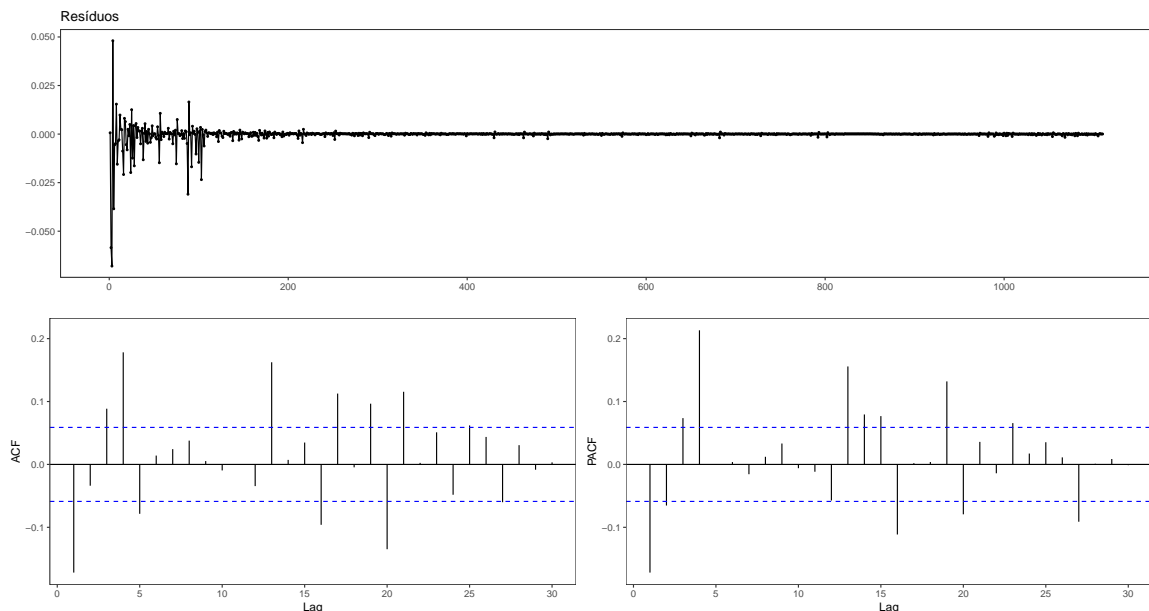


Figura 17: FAC e FACP dos resíduos (22/01/2018)

Nota-se na figura 17 que a maioria dos picos não são significativos na FAC e FACP. Portanto, considera-se que os resíduos são ruído branco. As exigências para o uso do modelo são a independência e normalidade dos resíduos. Para ratificar

a suposição de independência, foi feito o gráfico dos resíduos padronizados e está representado na figura 18.

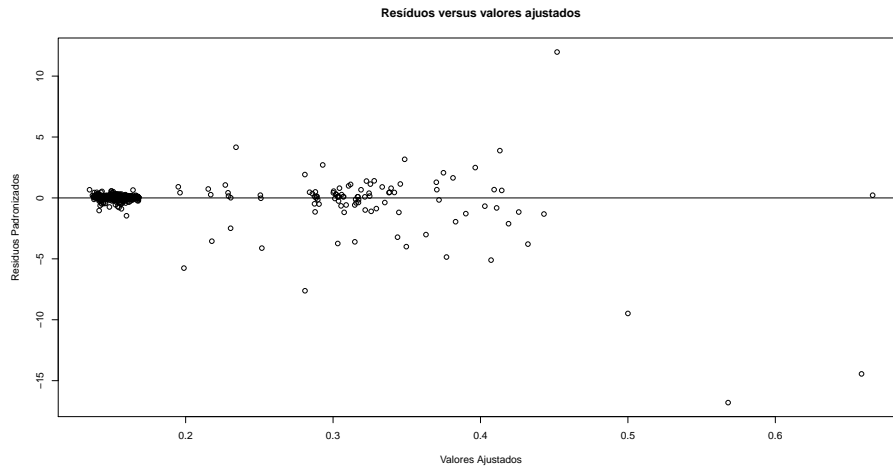


Figura 18: Resíduos Padronizados (22/01/2018)

Através da figura 18, nota-se que os resíduos estão dispostos aleatoriamente em torno do zero. Logo, a suposição de independência parece ser atendida. Para verificação da normalidade dos resíduos, foram feitos os gráficos histograma e QQplot. Eles são exibidos na figura 19.

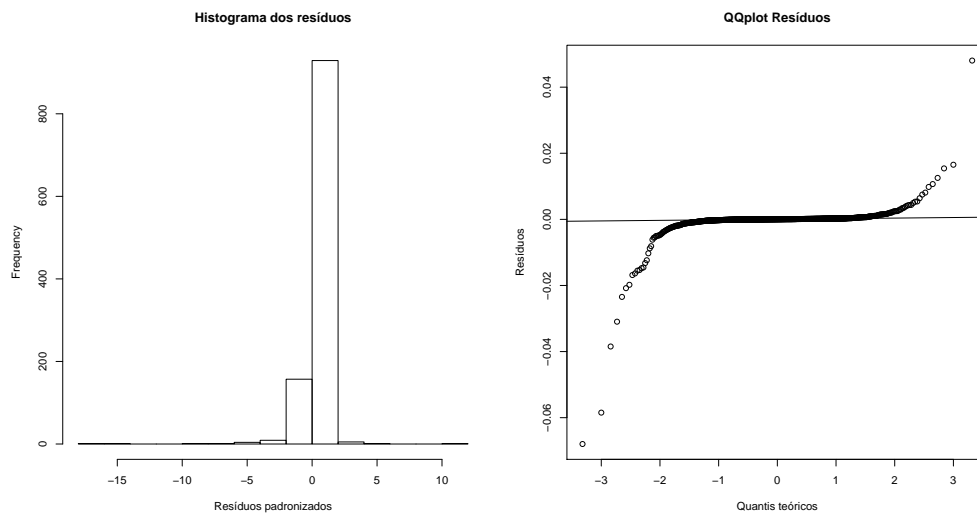


Figura 19: Histograma e QQplot dos resíduos (22/01/2018)

Observa-se no histograma da figura 19 que o comportamento dos resíduos é semelhante a uma distribuição Normal. No gráfico QQplot, nota-se que a maioria dos

pontos estão próximos ou sobre a reta. Logo, a suposição de normalidade parece ser satisfeita.

Outra medida importante para testar a qualidade do ajuste consiste em obter as medidas de acurácia. Elas são apresentadas na tabela 11.

Tabela 11: Medidas de acurácia (22/01/2018) - modelo ARIMA(1,1,0)

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-0,0002	0,0040	0,0009	-0,0708	0,3447	1,0520	-0,1720

Observa-se na tabela 11 que a maioria das medidas de acurácia são baixas, ou seja, próximas de zero. Logo, por esse e os demais itens verificados, o modelo ajustado é adequado.

Os mesmos passos feitos anteriormente serão adotados para os demais dias. Diante disso, para a série relacionada ao dia 29/01/2018, uma segunda-feira, estão representados na figura 20 os gráficos da FAC e FACP.

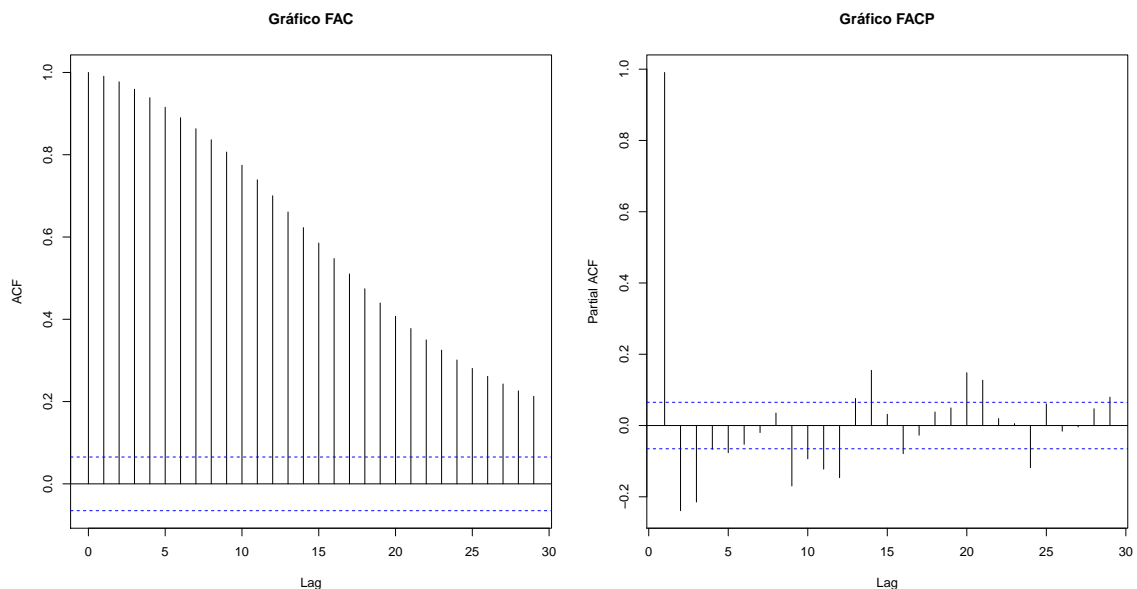


Figura 20: FAC e FACP da série (29/01/2018)

Através da figura 20, nota-se um decaimento exponencial no gráfico da FAC e um pico no *lag* 1 no gráfico da FACP. Isso sugere que o modelo deve ser ajustado com termos autorregressivos, uma análise bastante similar aos dados anteriormente

discutidos.

Após algumas tentativas, foi escolhido o modelo AR(2), em que os critérios adotados novamente foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. Os coeficientes desse modelo são mostrados na tabela 12.

Tabela 12: Coeficientes e medida AIC (29/01/2018) - modelo AR(2)

	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	1,3238	0,0314	0,0000	-8017,8400
AR(2)	-0,3326	0,0315	0,0000	
Intercepto	0,1400	0,0101	0,0000	

Nota-se na tabela 12 que todos os coeficientes são significativos, pois os valores-p são menores que 0,05. Além disso, observa-se que o valor de AIC é muito baixo, portanto, segundo este critério, o modelo está bem ajustado.

Posteriormente, verificou-se a função de autocorrelação e função de autocorrelação parcial para tentar detectar possíveis relações entre os resíduos. Os resultados são exibidos nos gráficos da figura 21.

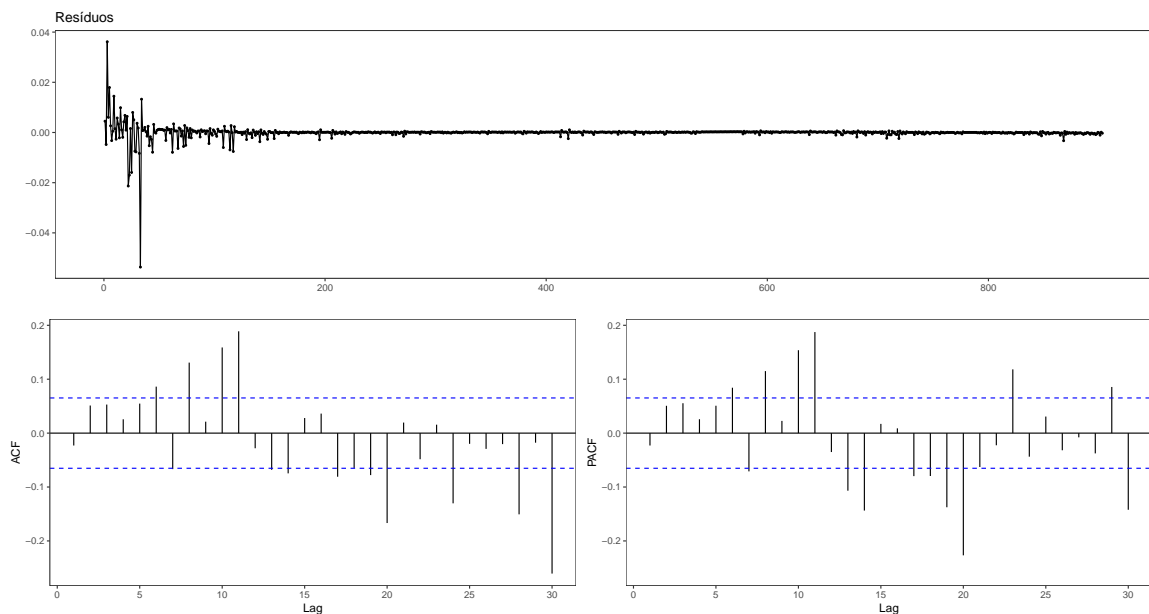


Figura 21: FAC e FACP dos resíduos (29/01/2018)

Nota-se na figura 21 que a maioria dos picos não são significativos na FAC e

FACP. Portanto, considera-se que os resíduos são ruído branco.

As exigências para o uso do modelo são a independência e normalidade dos resíduos. A confirmação da suposição de independência, foi feita através do gráfico dos resíduos padronizados representado na figura 22.

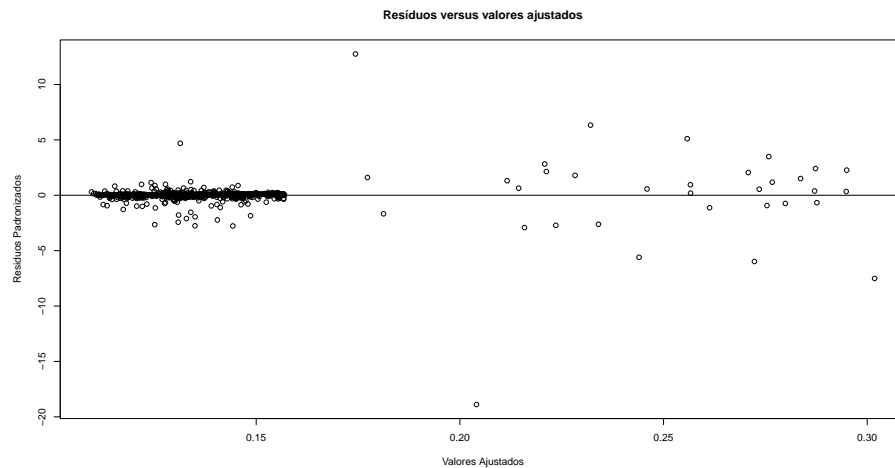


Figura 22: Resíduos Padronizados (29/01/2018)

A figura 22, apresenta resíduos dispostos aleatoriamente em torno do zero. Confirmando a suposição de independência. Já a verificação da normalidade dos resíduos, foi feita através de gráficos histograma e QQplot que podem ser visualizados na figura 23.

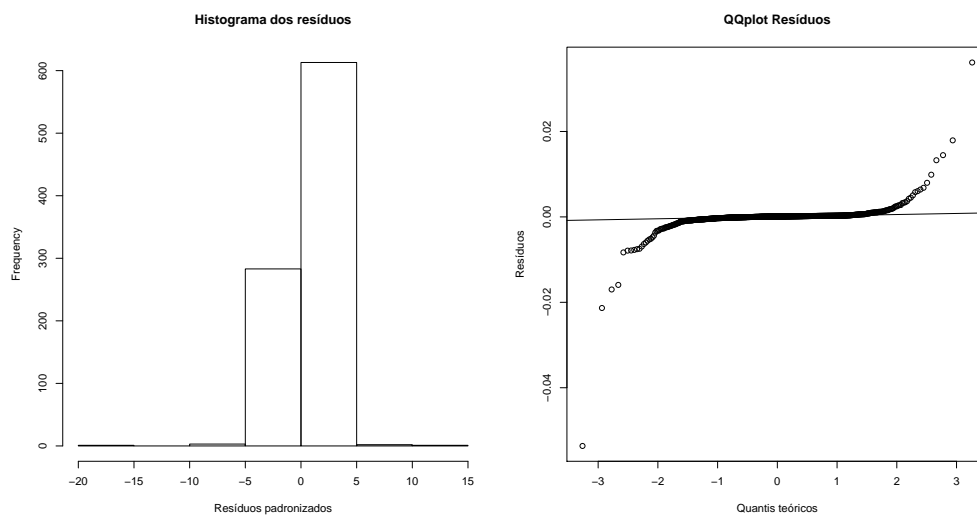


Figura 23: Histograma e QQplot dos resíduos (29/01/2018)

O histograma da figura 23 mostra que o comportamento dos resíduos é semelhante a uma distribuição Normal. No gráfico QQplot, verifica-se que boa parte dos pontos estão próximos ou sobre a reta atestando a suposição de normalidade. Outro passo importante para testar a qualidade do ajuste consiste em obter as medidas de acurácia como pode ser visualizado na tabela 13.

Tabela 13: Medidas de acurácia (29/01/2018) - modelo AR(2)

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
0,0000	0,0028	0,0007	-0,0489	0,4536	0,9960	-0,0230

A tabela 13 apresenta a maioria das medidas de acurácia com valores baixos, ou seja, próximas de zero. Esse conjunto de itens verificados atesta que o modelo ajustado é adequado.

Para a série relacionada ao dia 05/03/2018, uma segunda-feira, estão representados na figura 24 os gráficos da FAC e FACP. Através da figura 24, novamente verifica-se um decaimento exponencial no gráfico da FAC e um pico significativo no gráfico da FACP.

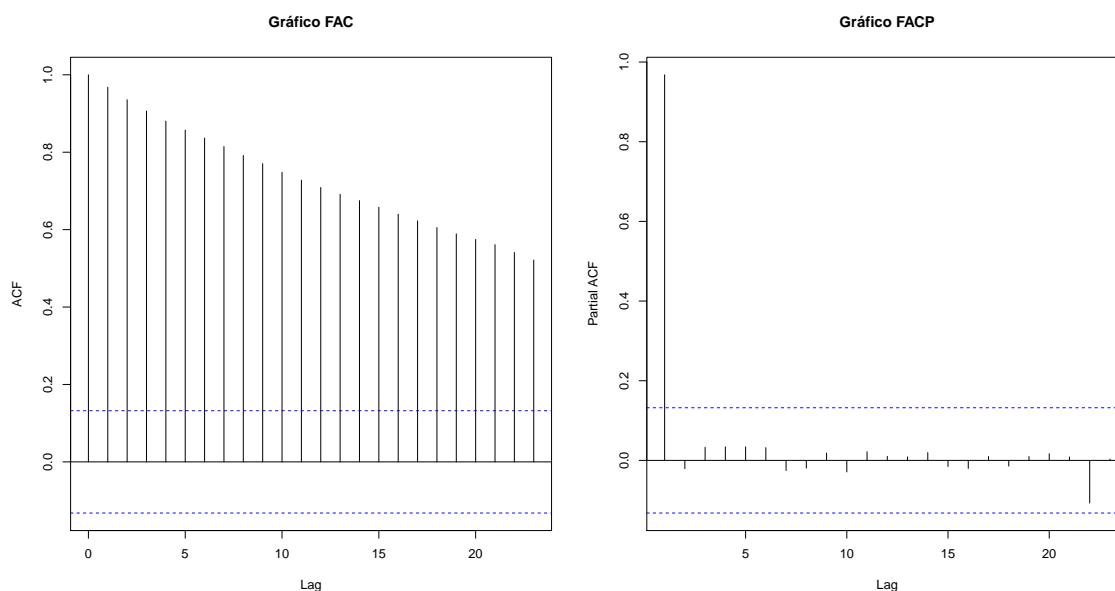


Figura 24: FAC e FACP da série (05/03/2018)

Estas características indicam que a série deve ser modelada por um modelo AR(1). Na proposta de modelo AR(1), o intercepto não é significativo. Na tentativa

do modelo AR(2) ocorrem problemas de convergência, e o modelo AR(3) apresentou somente o 1 coeficiente significativo.

Após algumas tentativas, foi escolhido o modelo ARIMA(1,1,0), considerando os mesmos critérios de parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. O coeficiente do modelo pode ser visualizado na tabela 14.

Tabela 14: Coeficiente e medida AIC (05/03/2018) - modelo ARIMA(1,1,0)

	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,2843	0,0650	0,0000	-2743,1300

Nota-se na tabela 14 que o coeficiente é significativo segundo avaliação através do valor-p. O valor de AIC é muito baixo, portanto, segundo este critério, o modelo está bem ajustado.

Verificou-se ainda a função de autocorrelação e função de autocorrelação parcial buscando tentar detectar possíveis relações entre os resíduos. Os resultados são exibidos nos gráficos da figura 25.

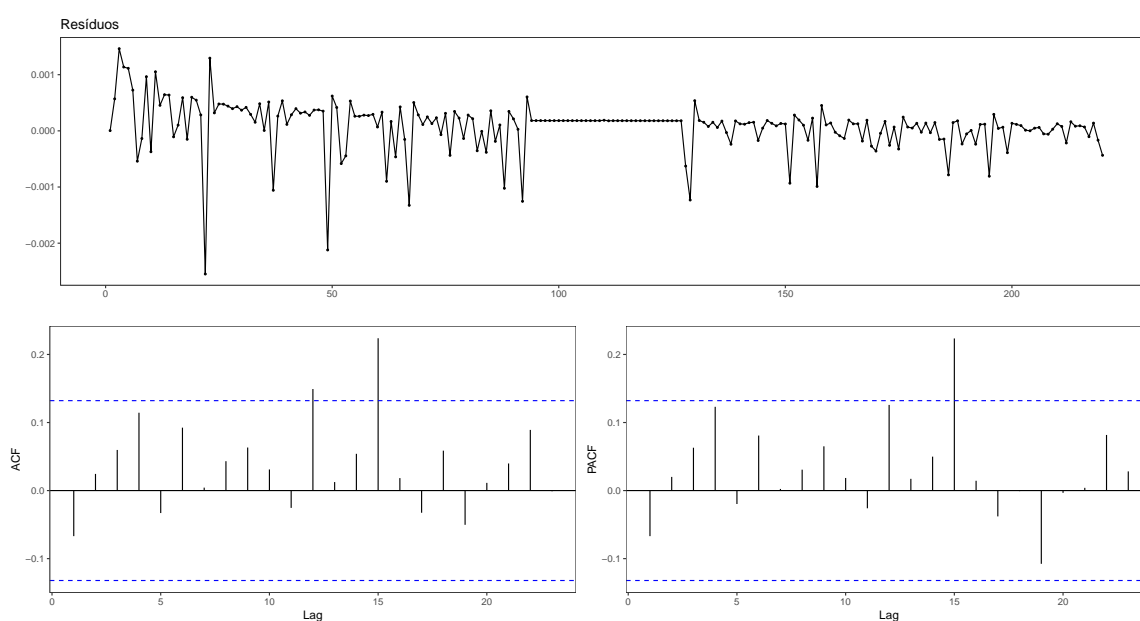


Figura 25: FAC e FACP dos resíduos (05/03/2018)

Nota-se na figura 25 que a maioria dos picos não são significativos na FAC e

FACP. Portanto, considera-se que os resíduos são ruído branco.

Para ratificar a suposição de independência, foi feito o gráfico dos resíduos padronizados que pode ser visto na figura 26.

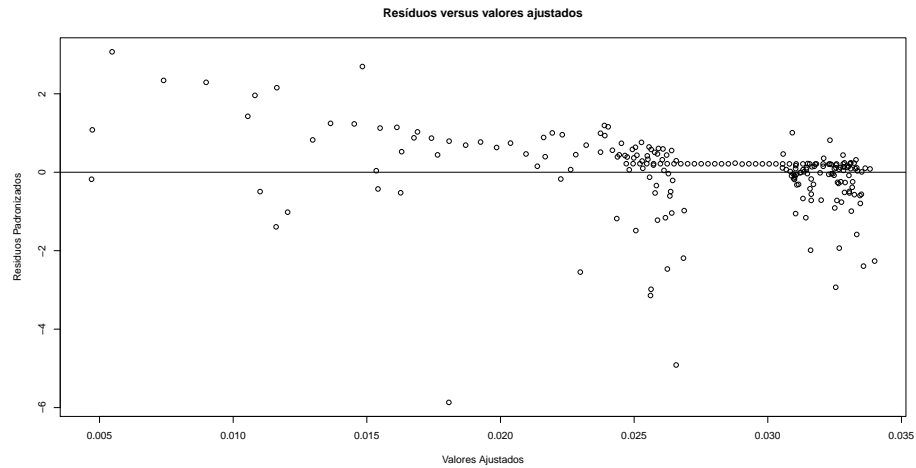


Figura 26: Resíduos Padronizados (05/03/2018)

Segundo análise gráfica realizada através da figura 26, a suposição de independência parece ser atendida. Gráficos histograma e QQplot são exibidos na figura 27. Observa-se no histograma da figura 27 que o comportamento dos resíduos revela alguma tendência de normalidade.

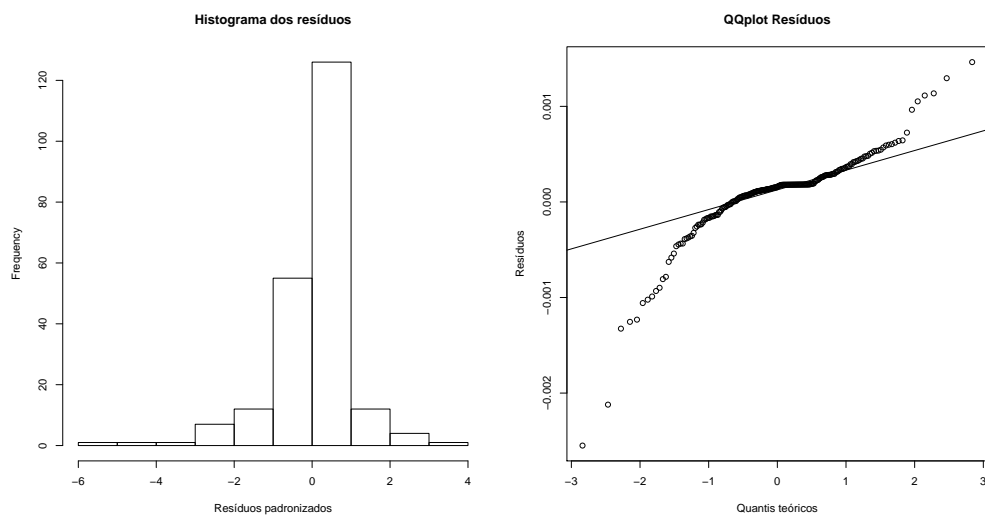


Figura 27: Histograma e QQplot dos resíduos (05/03/2018)

O gráfico QQplot apresenta alguns pontos escapando da reta de quantis teó-

ricos, principalmente na cauda inicial, algum efeito de assimetria, mas nada suficiente para rechaçar completamente a suposição de normalidade.

A tabela 15 apresenta dados com as medidas de acurácia desse caso. Observa-se na tabela 15 que a maioria das medidas de acurácia são suficientemente baixas.

Tabela 15: Medidas de acurácia (05/03/2018) - modelo ARIMA(1,1,0)

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
0,0001	0,0005	0,0003	0,5845	1,5190	0,9396	-0,0669

A análise da tabela 15 e os demais itens verificados confirmam que o modelo ajustado pode ser considerado adequado.

Ainda considerando dados referentes à coletas em segundas-feiras, foi analisada a série relacionada ao dia 09/04/2018. A figura 28 apresenta os gráficos da FAC e FACP.

Novamente verifica-se através da figura 28 um decaimento quase exponencial no gráfico da FAC e um pico significativo no gráfico da FACP. Essa constatação sugere que o modelo deve ser ajustado com termos autorregressivos.

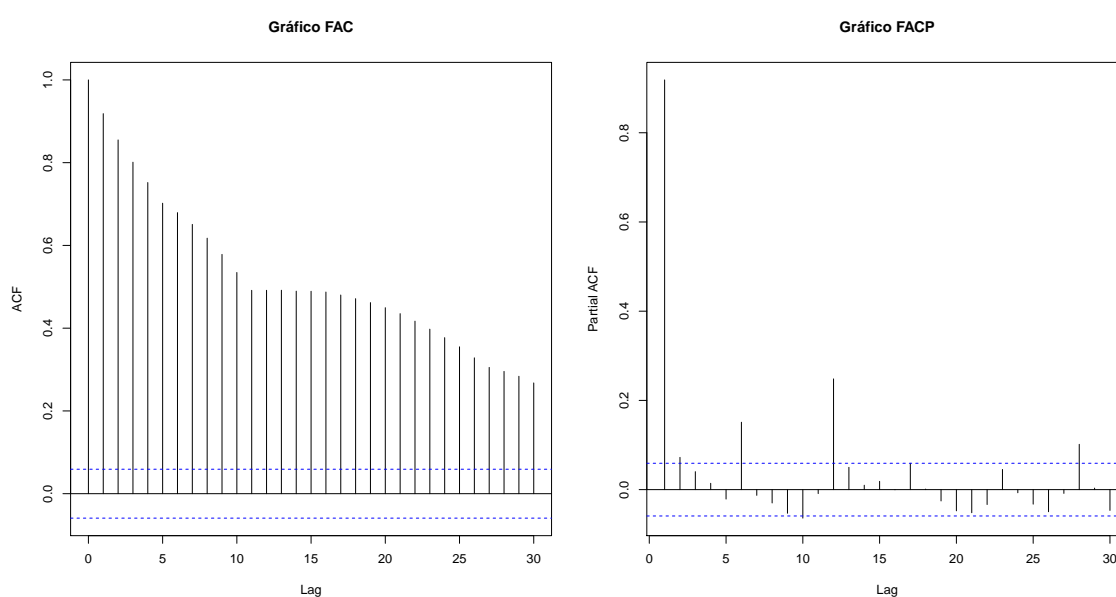


Figura 28: FAC e FACP da série (09/04/2018)

A tentativa inicial foi de buscar o ajuste de um modelo AR(1), posteriormente AR(2) e AR(3). Em todos os casos o ajuste não foi plenamente adequado, por fim o melhor ajuste obtido foi através de um modelo ARIMA(1,1,0).

Nesse caso, foi escolhido o modelo ARIMA(1,1,0), novamente com os mesmos critérios adotados: parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia.

O coeficiente desse modelo é mostrado na tabela 16. Novamente é apresentado coeficiente significativo a um nível de 0,05. O valor de AIC é muito baixo, ou seja, um critério que confirma qualidade na tentativa de ajuste de modelo proposta.

Tabela 16: Coeficiente e medida AIC (09/04/2018) - modelo ARIMA(1,1,0)

	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,2889	0,0317	0,0000	-8436,6100

Posteriormente, verificou-se a função de autocorrelação e função de autocorrelação parcial para tentar detectar possíveis relações entre os resíduos. Os resultados são exibidos nos gráficos da figura 29.

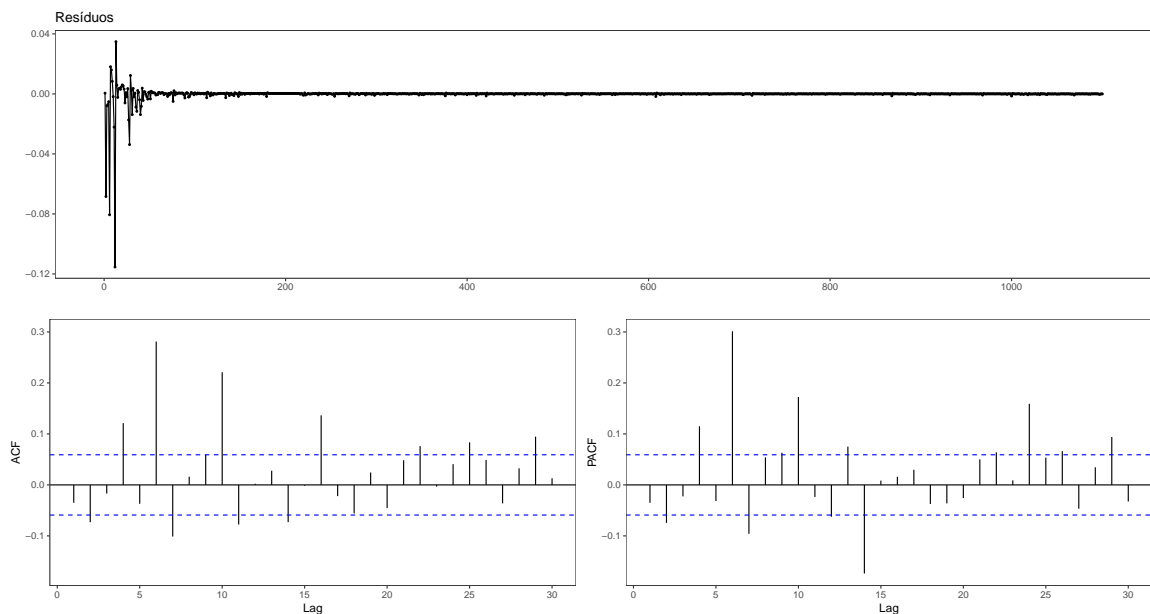


Figura 29: FAC e FACP dos resíduos (09/04/2018)

Observa-se através da figura 29 que não se apresentam picos significativos

na análise da FAC e FACP. Isso é um indício para a consideração de que os resíduos se comportam como ruído branco.

Como já mencionado, são exigências para o uso do modelo a suposição de independência e de normalidade dos resíduos. Novamente a validação de tais pressupostos precisa ser verificada. A suposição de independência é avaliada através de um gráfico dos resíduos padronizados que é representado na figura 30.

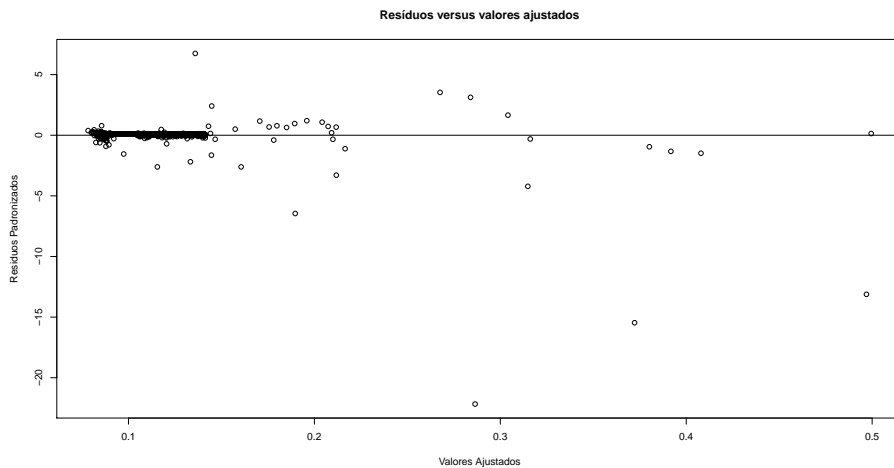


Figura 30: Resíduos Padronizados (09/04/2018)

Através da figura 30, nota-se que os resíduos estão dispostos aleatoriamente em torno do zero confirmando que a suposição de independência parece ser atendida.

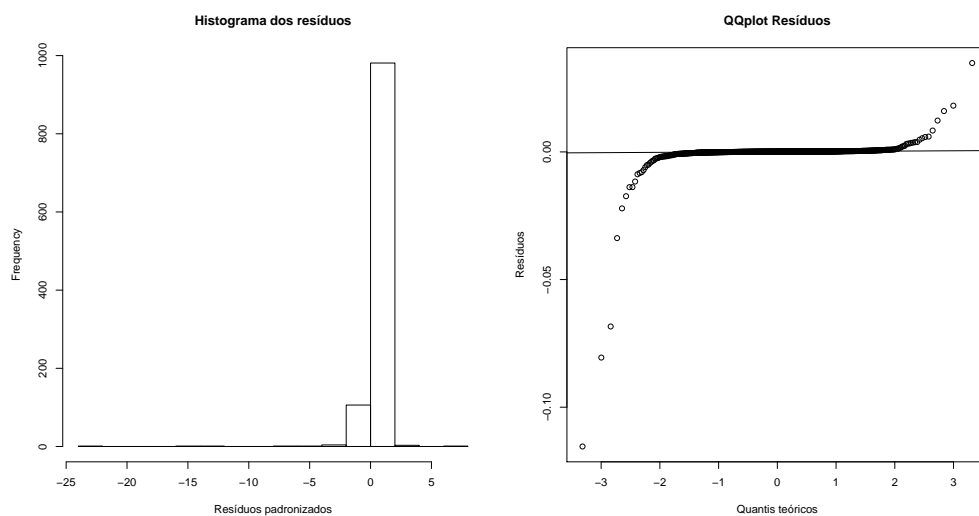


Figura 31: Histograma e QQplot dos resíduos (09/04/2018)

Além disso, para a verificação da suposição de normalidade dos resíduos,

foram feitos os gráficos histograma e QQplot. Esses gráficos para análise de normalidade se encontram exibidos na figura 31.

A figura 31 revela que o comportamento dos resíduos parece ser aproximadamente Normal, principalmente quando considerado o gráfico QQplot com a maioria dos pontos suficientemente próximos da reta. Portanto, parece razoável admitir que a suposição de normalidade esteja satisfeita.

Por fim, a tabela 17 apresenta as medidas de acurácia com valores sempre baixos, próximos de zero. Esse conjunto de análises corrobora a qualidade do ajuste proposto nesse caso.

Tabela 17: Medidas de acurácia (09/04/2018) - modelo ARIMA(1,1,0)

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-0,0002	0,0052	0,0007	-0,0994	0,4628	0,9930	-0,0351

O último conjunto de dados analisados referentes a uma segunda-feira, dia 16/04/2018, podem ser visualizados na figura 32 os gráficos da FAC e FACP.

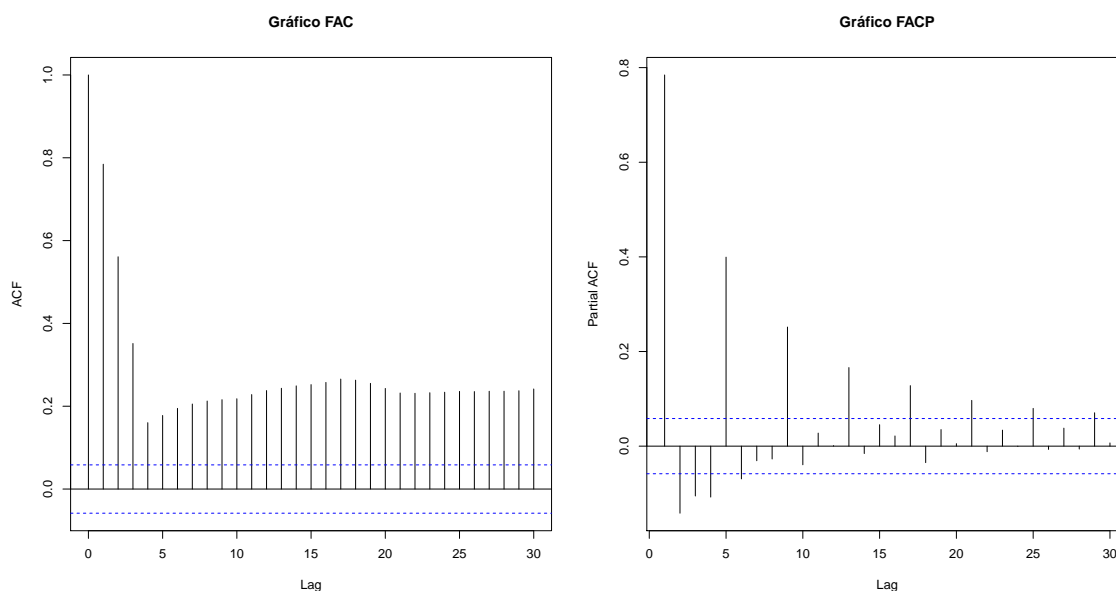


Figura 32: FAC e FACP da série (16/04/2018)

Através da figura 32, nota-se picos nos 2 primeiros *lags* no gráfico da FAC e decrescimento alternado no gráfico da FACP. Isso sugere que o modelo deve ser

ajustado com termos autorregressivos e médias móveis.

Após algumas tentativas, foi escolhido o modelo $ARIMA(1,0,0)(0,0,1)_5$. Essa estratégia de modelo se aplica à maioria das séries sazonais reais, ou seja, realizações de processos que apresentam correlação serial “dentro” e “entre” períodos sazonais. Novamente seguindo critérios de parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. Os coeficientes desse modelo são mostrados na tabela 18, verifica-se que os coeficientes do modelo são significativos a um nível de significância de 0,05. O critério AIC também confirma a tese anterior pois apresenta um valor muito baixo. Configurando indícios favoráveis a qualidade do ajuste proposto para esse conjunto de dados.

Tabela 18: Coeficientes e medida AIC (16/04/2018) - modelo $ARIMA(1,0,0)(0,0,1)_5$

	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,9983	0,0035	0,0000	-7433,3400
SMA(1)	-0,8216	0,0231	0,0000	
Intercepto	0,1604	0,0318	0,0000	

Verificou-se ainda a função de autocorrelação e função de autocorrelação parcial para tentar detectar possíveis relações entre os resíduos como visto na figura 33.

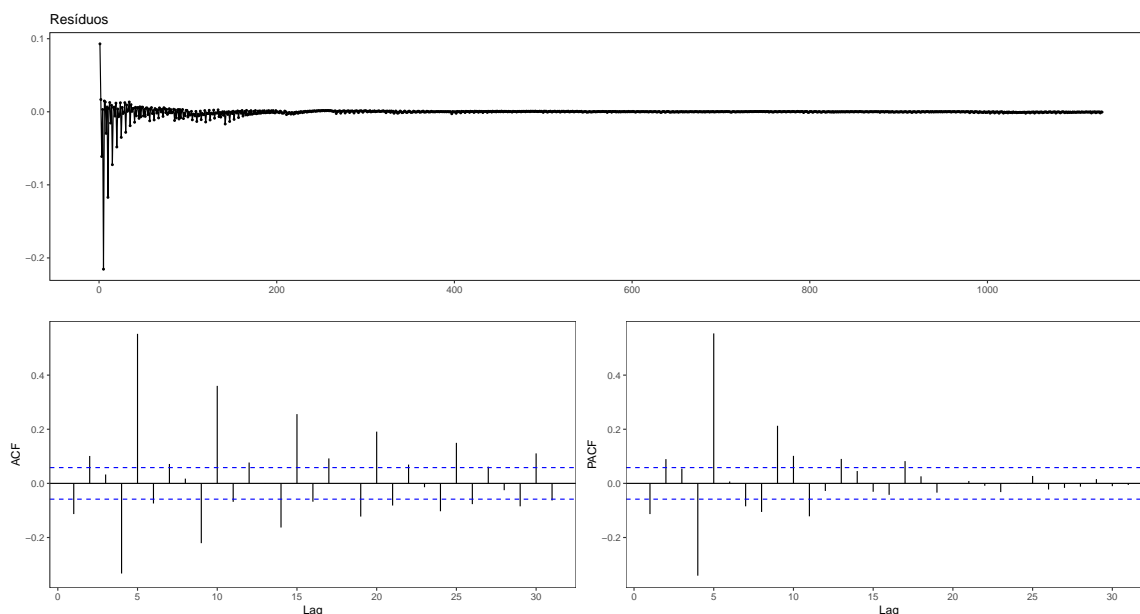


Figura 33: FAC e FACP dos resíduos (16/04/2018)

Através dos gráficos apresentados na figura 33 verifica-se que a maioria dos picos não são significativos na FAC e FACP.

As exigências de independência dos resíduos foram também avaliadas através de gráfico de resíduos padronizados e está representado na figura 34, nota-se que os resíduos estão dispostos aleatoriamente em torno do zero. Logo, a suposição de independência parece ser atendida.

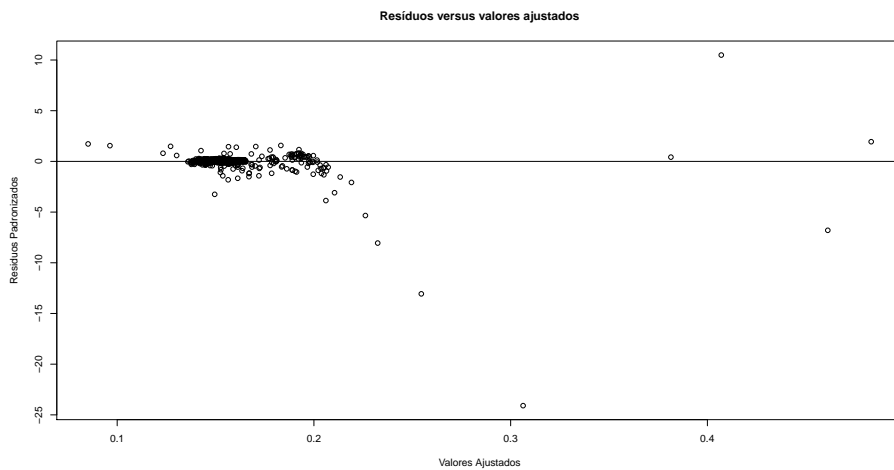


Figura 34: Resíduos Padronizados (16/04/2018)

Além disso, uma análise sobre o pressuposto de normalidade foi realizada através de avaliação gráfica de histograma e também gráfico QQplot.

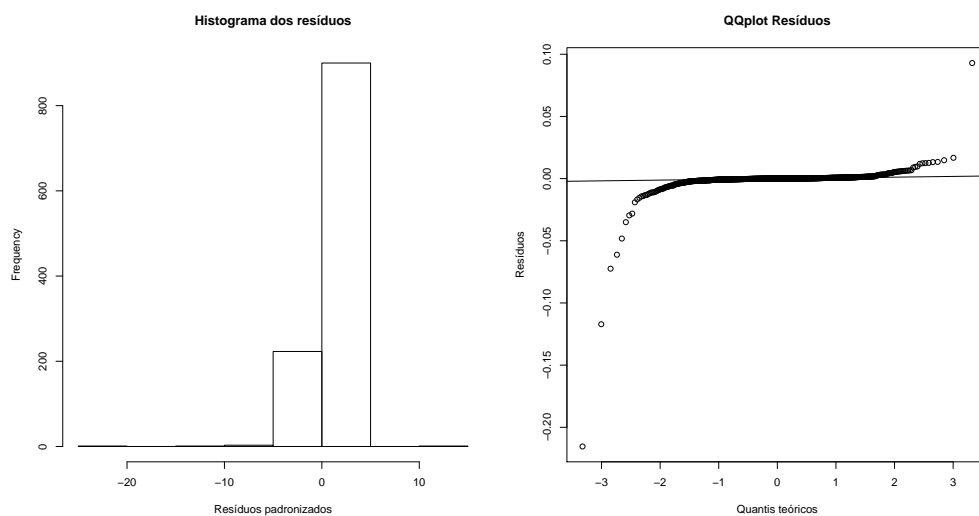


Figura 35: Histograma e QQplot dos resíduos (16/04/2018)

Observa-se no histograma da figura 35 que o comportamento dos resíduos

é aparentemente semelhante a uma distribuição Normal. No gráfico QQplot, nota-se que a maioria dos pontos na vizinhança da reta de quantis teóricos.

Não menos relevante para aferir a qualidade do ajuste é considerar a avaliação em termos de medidas de acurácia. Essas informações são apresentadas na tabela 19. Os valores verificados são suficientemente baixos a ponto de ser razoável admitir que o ajuste seja satisfatório.

Tabela 19: Medidas de acurácia (16/04/2018) - modelo ARIMA(1,0,0)(0,0,1)₅

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-0,0006	0,0089	0,0017	-0,4963	1,1216	2,0999	-0,1134

A análise dos ajustes para dados referentes às segundas-feiras foi feita de forma bastante detalhada e individualizada, com o intuito de ser além de informativo, suficientemente didático nesse texto. Para os demais dias a análise será apresentada de forma mais condensada.

Serão apresentados a seguir as tentativas de ajuste dos modelos referentes às terças-feiras 06/02/2018, 27/02/2018, 06/03/2018, 13/03/2018 e 03/04/2018.

Considerando o cenário de tentativas de ajuste, para o dia 06/02/2018 foi escolhido o modelo ARIMA(0,1,2), para o dia 27/02/2018 foi escolhido o modelo ARIMA(1,1,0), para o dia 06/03/2018 foi escolhido o modelo ARIMA(2,1,0), para o dia 13/03/2018 foi escolhido o modelo ARIMA(0,1,2) e para o dia 03/04/2018 foi escolhido o modelo ARIMA(0,1,2). Os critérios adotados para avaliação da qualidade do ajuste foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. Os coeficientes desses modelos são apresentados na tabela 20.

Nota-se na tabela 20 que todos os coeficientes são significativos quando considerado um nível de significância de 0,05. Além disso, observa-se para todos os ajustes propostos que o valor de AIC é muito baixo, portanto, segundo este critério, os modelos podem ser avaliados como propostas de ajuste satisfatórias.

Tabela 20: Coeficientes e medidas AIC (terça-feira)

modelo ARIMA(0,1,2) - 06/02/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	0,1324	0,0350	0,0002	-8453,4200
MA(2)	0,0870	0,0385	0,0239	
modelo ARIMA(1,1,0) - 27/02/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,1918	0,0927	0,0386	-2327,1800
modelo ARIMA(2,1,0) - 06/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	1,9848	0,0038	0,0000	-874,0500
AR(2)	-1,0000	0,0000	0,0000	
modelo ARIMA(0,1,2) - 13/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	0,2003	0,0283	0,0000	-8429,6900
MA(2)	0,2423	0,0295	0,0000	
modelo ARIMA(0,1,2) - 03/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	0,4430	0,0328	0,0000	-9095,6600
MA(2)	0,8030	0,0211	0,0000	

Através de análises gráficas, verificou-se que os resíduos estão dispostos de forma aleatória em torno de zero. Logo, a suposição de independência parece estar satisfeita em todos os cenários. Além disso, não se observou constatações suficientemente fortes para atestar a violação da suposição de normalidade dos resíduos.

Tabela 21: Medidas de acurácia (terça-feira)

Data	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
06/02/2018	0,0000	0,0023	0,0007	0,0166	0,4359	0,9931	-0,0036
27/02/2018	-0,0009	0,0133	0,0016	-0,6095	1,5913	1,0377	-0,1820
06/03/2018	-0,0037	0,0663	0,0073	-1,0938	4,5824	1,1979	-0,0580
13/03/2018	-0,0002	0,0070	0,0007	-0,0981	0,4650	1,0032	-0,0051
03/04/2018	-0,0002	0,0052	0,0011	-0,0547	0,5284	1,3148	-0,1601

Ainda para testar a qualidade do ajuste as medidas de acurácia estão apresentadas na tabela 21. Observa-se que a maioria das medidas de acurácia são baixas, ou seja, suficientemente próximas de zero. Logo, por esse e pelos demais itens verificados conclui-se que é bastante razoável admitir que todos os cinco modelos ajustados sejam adequados.

Serão apresentados a seguir as tentativas de ajuste dos modelos referentes

às quartas-feiras 24/01/2018, 28/02/2018, 07/03/2018, 14/03/2018 e 11/04/2018. Os coeficientes dos modelos escolhidos são apresentados na tabela 22.

Tabela 22: Coeficientes e medidas AIC (quarta-feira)

modelo ARIMA(1,1,2) - 24/01/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,7358	0,0605	0,0000	-8472,9500
MA(1)	-0,8748	0,0533	0,0000	
MA(2)	0,5406	0,0303	0,0000	
modelo ARIMA(2,1,0) - 28/02/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,3485	0,0404	0,0000	-2351,0400
AR(2)	0,6444	0,0409	0,0000	
modelo ARIMA(1,1,2) - 07/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	-0,5415	0,1953	0,0056	-2952,5400
MA(1)	0,6376	0,1842	0,0005	
MA(2)	0,2327	0,0607	0,0001	
modelo ARIMA(2,1,1) - 14/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,2412	0,0290	0,0000	-10196,7400
AR(2)	0,7521	0,0288	0,0000	
MA(1)	0,1231	0,0382	0,0013	
modelo ARIMA(2,1,0) - 11/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	-0,1530	0,0281	0,0000	-11266,3000
AR(2)	-0,1436	0,0371	0,0001	

Tomando em conta as possibilidades de tentativas de ajuste, para o dia 24/01/2018 foi escolhido o modelo ARIMA(1,1,2), para o dia 28/02/2018 foi escolhido o modelo ARIMA(2,1,0), para o dia 07/03/2018 foi escolhido o modelo ARIMA(1,1,2), para o dia 14/03/2018 foi escolhido o modelo ARIMA(2,1,1) e para o dia 11/04/2018 foi escolhido o modelo ARIMA(2,1,0). Os critérios adotados para avaliação da qualidade do ajuste foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia.

A tabela 22 mostra em todos os casos coeficientes significativos em relação a um nível de significância de 0,05. Verifica-se ainda para todos os ajustes propostos que o valor de AIC é muito baixo, portanto, segundo este critério, os modelos podem ser avaliados como propostas de ajuste satisfatórias.

A utilização desses modelos é condicionada a requisitos de independência e normalidade dos resíduos. Com intenção de ratificar a suposição de independência foram avaliados gráficos de resíduos padronizados. Para verificação da normalidade dos resíduos, foram analisados os gráficos histograma e QQplot nos cinco ajustes. A análise do dia 07/03/2018, 14/03/2018 e do dia 11/04/2018 apresentou caudas pesadas no gráfico QQplot, esse efeito parece mais amenizado nos dias 24/01/2018 e 28/02/2018. Entretanto não são constatações suficientemente fortes para atestar a violação da suposição de normalidade dos resíduos.

Ainda para testar a qualidade do ajuste as medidas de acurácia estão apresentadas na tabela 23. Verifica-se que a maioria das medidas de acurácia são baixas, ou seja, suficientemente próximas de zero. Logo, por esse, e pelos demais itens verificados, conclui-se que é bastante razoável admitir que todos os cinco modelos ajustados sejam adequados.

Tabela 23: Medidas de acurácia (quarta-feira)

Data	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
24/01/2018	-0,0001	0,0067	0,0008	-0,0364	0,5136	1,1140	0,0333
28/02/2018	0,0005	0,0085	0,0014	0,4169	1,8963	0,7740	0,0018
07/03/2018	0,0000	0,0017	0,0006	0,0501	1,4804	0,9889	-0,0153
14/03/2018	0,0001	0,0034	0,0007	0,0379	0,3934	1,0275	0,0025
11/04/2018	-0,0001	0,0028	0,0006	-0,0714	0,3328	1,0480	-0,0009

Serão apresentados a seguir as tentativas de ajuste dos modelos referentes às quintas-feiras 18/01/2018, 22/02/2018, 15/03/2018, 05/04/2018 e 26/04/2018.

Observando as tentativas de ajuste verificou-se: para o dia 18/01/2018 foi escolhido o modelo ARIMA(1,1,2), para o dia 22/02/2018 foi escolhido o modelo ARIMA(0,1,2), para o dia 15/03/2018 foi escolhido o modelo ARIMA(0,1,1), para o dia 05/04/2018 foi escolhido o modelo ARIMA(1,1,2) e para o dia 26/04/2018 foi escolhido o modelo ARIMA(2,1,0). Os critérios adotados para avaliação da qualidade do ajuste foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. Os coeficientes desses modelos são apresentados na tabela 24.

Tabela 24: Coeficientes e medidas AIC (quinta-feira)

modelo ARIMA(1,1,2) - 18/01/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,7109	0,0386	0,0000	-7230,1800
MA(1)	0,1824	0,0316	0,0000	
MA(2)	0,6299	0,0286	0,0000	
modelo ARIMA(0,1,2) - 22/02/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	1,0030	0,0294	0,0000	-5790,2700
MA(2)	0,7003	0,0268	0,0000	
modelo ARIMA(0,1,1) - 15/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	-0,1956	0,0288	0,0000	-7770,8500
modelo ARIMA(1,1,2) - 05/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,9854	0,0168	0,0000	-8900,5800
MA(1)	-0,6921	0,0350	0,0000	
MA(2)	0,2030	0,0413	0,0000	
modelo ARIMA(2,1,0) - 26/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,2477	0,0314	0,0000	-8852,1900
AR(2)	-0,4475	0,0339	0,0000	

A tabela 24 apresenta sempre coeficientes significativos para a um nível de significância de 0,05. Observa-se também que para todos os ajustes propostos um valor de AIC muito baixo, dessa forma, de acordo com esse critério, os modelos podem ser avaliados como ajustes adequados.

Os requisitos de independência e normalidade dos resíduos são suposições para a utilização desses modelos. A suposição de independência foi verificada através da análise de gráficos de resíduos padronizados. Para verificação da normalidade dos resíduos, foram verificados os gráficos histograma e QQplot nos cinco ajustes analisados.

A análise do dia 18/01/2018, 22/02/2018 e do dia 26/04/2018 apresenta caudas pesadas no gráfico QQplot, esse efeito parece mais amenizado nos dias 15/03/2018 e 05/04/2018. Mas não se tratam de constatações suficientemente robustas para aferir uma violação quanto a suposição de normalidade dos resíduos.

Por fim, ainda com interesse em garantir a qualidade do ajuste, as medidas de acurácia estão apresentadas na tabela 25.

Tabela 25: Medidas de acurácia (quinta-feira)

Data	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
18/01/2018	0,0000	0,0055	0,0011	-0,0013	0,8229	1,0904	-0,2165
22/02/2018	-0,0001	0,0087	0,0016	-0,0717	1,2535	1,6057	-0,5655
15/03/2018	-0,0004	0,0082	0,0009	-0,1691	0,4366	1,0800	0,0005
05/04/2018	0,0002	0,0063	0,0009	0,0693	0,5952	1,1522	0,0319
26/04/2018	-0,0001	0,0040	0,0009	-0,0917	0,4937	1,1200	-0,0991

A tabela 25 mostra para a maioria das medidas de acurácia valores baixos, ou seja, suficientemente próximos de zero. Isto posto, por esse, e pelas demais averiguações, conclui-se que é bastante razoável admitir que todos os cinco modelos ajustados sejam adequados.

Serão apresentados a seguir as tentativas de ajuste dos modelos referentes às sextas-feiras 16/02/2018, 09/03/2018, 06/04/2018, 13/04/2018 e 27/04/2018.

Tabela 26: Coeficientes e medidas AIC (sexta-feira)

modelo ARIMA(0,1,2) - 16/02/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	1,6333	0,0252	0,0000	-5684,3500
MA(2)	0,7352	0,0237	0,0000	
modelo ARIMA(1,1,1) - 09/03/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,8298	0,0769	0,0000	-1025,07
MA(1)	-0,9497	0,0295	0,0000	
modelo ARIMA(0,1,2) - 06/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	1,1578	0,0363	0,0000	-6680,3100
MA(2)	0,6489	0,0377	0,0000	
modelo ARIMA(1,1,1) - 13/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
AR(1)	0,9951	0,0064	0,0000	-7219,0300
MA(1)	-0,1813	0,0518	0,0005	
modelo ARIMA(0,1,2) - 27/04/2018				
	Coeficiente	Erro Padrão	Valor - p	AIC (modelo)
MA(1)	1,3803	0,0453	0,0000	-5394,8200
MA(2)	0,4368	0,0447	0,0000	

Observando as tentativas de ajuste verificou-se: para o dia 16/02/2018 foi

escolhido o modelo ARIMA(0,1,2), para o dia 09/03/2018 foi escolhido o modelo ARIMA(1,1,1), para o dia 06/04/2018 foi escolhido o modelo ARIMA(0,1,2), para o dia 13/04/2018 foi escolhido o modelo ARIMA(1,1,1) e para o dia 27/04/2018 foi escolhido o modelo ARIMA(0,1,2). Os critérios adotados para avaliação da qualidade do ajuste foram a parcimônia, significância dos coeficientes, valor de AIC, gráficos de resíduos e acurácia. Os coeficientes desses modelos são apresentados na tabela 26.

A análise da tabela 26 mostra coeficientes significativos considerando um nível de significância de 0,05. Verifica-se ainda que dentre todos os ajustes propostos valores de AIC muito baixo, dessa forma, partindo desse critério, os modelos podem ser interpretados como ajustes adequados.

Quanto a independência e normalidade dos resíduos, essas são suposições relevantes para a utilização desses modelos. A suposição de independência foi verificada através da análise de gráficos de resíduos padronizados.

Além disso, para avaliação da suposição de normalidade para os resíduos, foram observados os gráficos histograma e também QQplot para as cinco propostas de ajustes analisadas. A análise de todos os cinco casos apresenta caudas com algum peso no gráfico QQplot, entretanto não parecem constatações suficientemente robustas para admitir uma violação quanto a suposição de normalidade dos resíduos.

Novamente as medidas de acurácia são analisadas para garantir a qualidade do ajuste e estão apresentadas na tabela 27.

Tabela 27: Medidas de acurácia (sexta-feira)

Data	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
16/02/2018	-0,0003	0,0116	0,0029	-0,0574	1,6468	2,1205	-0,6612
09/03/2018	-0,0043	0,0307	0,0056	-11,9114	12,5715	1,4333	0,0736
06/04/2018	-0,0003	0,0117	0,0017	-0,0623	0,8120	1,4139	-0,5971
13/04/2018	0,0003	0,0093	0,0010	0,1230	0,6150	1,1408	0,0089
27/04/2018	-0,0003	0,0137	0,0032	-0,0788	1,5515	2,2736	-0,6964

Através da tabela 27 verifica-se que para a maioria das medidas de acurácia os valores são baixos, ou seja, suficientemente próximos de zero. Com isso, por esse,

e pelas demais constatações, é bastante viável admitir que todos os cinco modelos ajustados sejam adequados.

Em todos os cenários investigados, sempre foi possível obter um ajuste de modelo que minimamente pode ser considerado satisfatório. Esse indício confirma a tese inicialmente aventada de que o procedimento de chegadas no Restaurante Universitário da UFOP se comporta como um processo de Poisson não homogêneo.

4 CONSIDERAÇÕES FINAIS

A execução desse estudo auxiliou sobremaneira no senso de apurar conhecimentos teóricos aliados ao contexto prático da tentativa de modelagem de um problema real, a saber, a discussão sobre a modelagem para o sistema de atendimento do Restaurante Universitário da Universidade Federal de Ouro Preto (UFOP) por meio de Teoria das Filas, uma relevante área de investigação em Estatística e Pesquisa Operacional.

Foi executado um estudo de caso condicionado à disponibilidade de dados de utilização do Restaurante Universitário da Universidade Federal de Ouro Preto (UFOP). Os dados de passagem de usuários pela catraca com horário exato estavam disponíveis. Infelizmente não existe um controle de horário de saída de indivíduos do restaurante, o que dificulta a produção de resultados mais robustos e abrangentes, mas a investigação realizada já é capaz de contribuir para o planejamento e melhor funcionamento do local.

Avaliando os objetivos descritos anteriormente para esse estudo, foi executada a construção de um pequeno referencial teórico acerca desse assunto de pesquisa, esse referencial poderá ser ampliado em estudos complementares e de continuidade dessa investigação. Enquanto auxílio na formação técnica, a utilização da linguagem $\text{T}_\text{E}_\text{X}$, que é padrão na confecção de textos estatísticos em vários níveis de pesquisa foi introduzida e alcançou níveis de fluência para sua aplicação futura em outros textos científicos. Discussões teóricas sobre Processos de contagem, Processos de Poisson homogêneos e não homogêneos foram abordadas. Bem como estudos para modelos de séries temporais dos tipos AR, MA, ARMA, ARIMA além de estudos sobre Teoria de Filas.

Foi realizada uma rica análise descritiva dos dados disponíveis, esse estudo se encontra descrito e detalhado no capítulo 3. É importante salientar que em avaliações estatísticas, muitas vezes a análise descritiva é subjulgada pelos pesquisadores,

mas esse trabalho reafirma o poder desse tipo de análise. O presente material descritivo direciona de forma contundente para as constatações mais relevantes desse estudo.

Através da análise descritiva, verificou-se que o padrão de entradas de usuários no Restaurante Universitário da UFOP se ajusta bem ao modelo teórico de distribuição exponencial de probabilidades, entretanto a taxa média de ocorrências por intervalo de tempo, parâmetro fundamental na distribuição exponencial de probabilidades apresenta diferenças entre dias de estudo e faixas de horário no estudo.

A mera constatação do ajuste exponencial, já é por si só, bastante relevante, pois caracteriza que o padrão de entrada de usuários se comporta como um Processo de Poisson. Por outro lado, a variação da taxa ao longo do tempo mostra que esse comportamento markoviano não é de um processo de Poisson homogêneo.

Essa análise induz a possibilidade de se aventar que trata-se de um processo de Poisson não homogêneo. Uma estratégia para confirmar tal fato seria tentar descrever a taxa λ das entradas de usuários no Restaurante Universitário da Universidade Federal De Ouro Preto (UFOP) como uma função $\lambda(t)$ dependente do tempo. Isso poderia ser visualizado através da tentativa de modelar a média do processo por meio de um modelo de séries temporais.

A seção 3.2 apresenta uma vasta investigação na tentativa de propor ajustes de modelos de séries temporais para os diferentes dias sob investigação, não era um propósito desse estudo a obtenção de um modelo único, tampouco isso teria algum senso lógico. O objetivo central seria garantir que sempre é possível encontrar algum modelo de série temporal com ajuste satisfatório. Isso seria resultado importante para corroborar que a taxa tende a ser descrita como uma variação dependente do tempo. Isso caracteriza o comportamento de um processo de Poisson não homogêneo.

O fato de em todos os dias investigados existir um modelo apresentando um ajuste adequado é notoriamente o resultado mais importante dessa investigação con-

firmando a tese de que a entrada de indivíduos no Restaurante Universitário da Universidade Federal De Ouro Preto (UFOP) se comporta como um processo de Poisson não homogêneo.

O grande trabalho no ajuste de diversos modelos de série temporal também revela um importante papel de desenvolvimento teórico no processo inicial de investigação científica. A habilidade em ajustar esses modelos sob investigação é com certeza uma grande e importante valência para um graduado em Estatística.

Esse relato ainda deixa lacunas abertas para investigações futuras e novos objetivos. Uma compilação desse estudo seria um trabalho com potencial claro para submissão em eventos científicos da área estatística. A busca por tabular dados com o instante temporal de saída dos usuários do Restaurante Universitário da Universidade Federal De Ouro Preto (UFOP) será de extrema valia para uma investigação mais ampla sobre o atendimento prestado e para o fornecimento de estratégias de melhoria.

Por fim, todos os envolvidos nesse estudo, agradecem enormemente o suporte fornecido pela Universidade Federal de Ouro Preto através da Pró-reitoria de Pesquisa e Pós-graduação que incentivou o início desse estudo através de um projeto de iniciação científica. Além disso, um agradecimento aos profissionais do restaurante universitário da UFOP que desde o princípio se mostraram disponíveis em auxiliar a execução do estudo. Todos os envolvidos nesse estudo esperam que futuramente incentivos dessa natureza possam ser mantidos e ampliados.

REFERÊNCIAS

- ABENSUR, E. O. Banking operations using queuing theory and genetic algorithms. *Produto & Produção*, v. 12 (2), p. 69–86, 2011.
- ARENALES, M. N.; ARMENTANO, V. A.; MORABITO, R. *Pesquisa operacional*. [S.l.]: Elsevier, 2007. (Coleção Campus-ABEPRO Engenharia de Produção). ISBN 9788535214543.
- BASS, R. F. *Stochastic Processes*. 1st. ed. New York, NY, USA: Cambridge University Press, 2011. ISBN 110700800X, 9781107008007.
- BHAT, U. N. *Elements of Applied Stochastic Processes*. Wiley, 1984. (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series). ISBN 9780471878261. Disponível em: <<https://books.google.com.br/books?id=LQOqHrsRwagC>>.
- CRUZ, F. R. B.; DUARTE, A. R.; SOUZA, G. L. Multi-objective performance improvements of general finite single-server queueing networks. *Journal of Heuristics*, v. 24, n. 5, p. 757–781, Oct 2018.
- CRUZ, F. R. B.; DUARTE, A. R.; WOENSEL, T. V. Buffer allocation in general single-server queueing networks. *Computers & Operations Research*, Elsevier, v. 35, n. 11, p. 3581–3598, 2008.
- CRUZ, F. R. B. et al. Throughput maximization of queueing networks with simultaneous minimization of service rates and buffers. *Mathematical Problems in Engineering*, Hindawi, Article ID 692593, 19 pages, 2012.
- GROSS, D.; HARRIS, C. M. *Fundamentals of queueing theory*. Wiley, 1985. (Wiley series in probability and mathematical statistics). ISBN 9780471890676. Disponível em: <<https://books.google.com.br/books?id=9mRRAAAAMAAJ>>.
- HILLIER, F. S.; LIEBERMAN, G. J. *Introdução à Pesquisa Operacional*. AMGH, 2013. ISBN 9788580551198. Disponível em: <<https://books.google.com.br/books?id=-A88a0-KxQ0C>>.
- KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded markov chains. *Annals Mathematical Statistics*, v. 24, p. 338–354, 1953.
- MARTINS, H. S. R. et al. Modeling and optimization of buffers and servers in finite queueing networks. *OPSEARCH*, 2019.
- MORETTIN, P. A. *Econometria financeira: um curso em séries temporais financeiras*. Edgard Blücher, 2008. ISBN 9788521204589. Disponível em: <<https://books.google.com.br/books?id=IWgmOwAACAAJ>>.
- MORETTIN, P. A.; TOLOI, C. M. de C. *Análise de séries temporais*. Edgard Blucher, 2006. (ABE - Projeto Fisher). ISBN 9788521203896. Disponível em: <<https://books.google.com.br/books?id=Q7bJAAAACAAJ>>.

- ROSS, S. M. *Applied Probability Models with Optimization Applications*. Dover Publications, 1992. (Dover Books on Advanced Mathematics). ISBN 9780486673141. Disponível em: <<https://books.google.com.br/books?id=Hftunnl-1zwC>>.
- SOARES, M. d. S. Aplicação de um modelos de teoria de filas em um restaurante universitário: estudo do tempo de atendimento. *Trabalho de Conclusão de Curso em Engenharia de Produção (Universidade Federal da Grande Dourados)*, 2016.
- YANG, X.-S. *Engineering Optimization: An Introduction with Metaheuristic Applications*. 1st. ed. [S.l.]: Wiley Publishing, 2010. ISBN 0470582464, 9780470582466.