



Instituto de
MATEMÁTICA
E ESTATÍSTICA

UFRGS



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DEPARTAMENTO DE ESTATÍSTICA

PREVISÃO DAS TAXAS DE OCUPAÇÃO DE HOTÉIS NO BRASIL

DANIELLE RODRIGUES GUIDINI

Porto Alegre
2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

PREVISÃO DAS TAXAS DE OCUPAÇÃO DE HOTÉIS NO BRASIL

Autor: Danielle Rodrigues Guidini

Trabalho de Conclusão de Curso
apresentado para a obtenção do grau de
Bacharel em Estatística

Banca Examinadora:
Professor Dr. Flávio Augusto Ziegelmann
Professor Dr. Eduardo de Oliveira Horta

Porto Alegre
2019

AGRADECIMENTOS

Agradeço a todos que, de alguma forma, me acompanharam e apoiaram durante a graduação.

Em especial, agradeço aos meus pais, Márcia e Elizeu, e ao meu avô, Alcendor, por entenderem minha mudança de cidade, minhas decisões e meus momentos de ausência e prover todos o suporte necessário para que eu alcançasse meus objetivos.

Aos meus amigos, obrigada pelo incentivo, pela paciência, por entenderem minhas frequentes ausências e pelo apoio nos momentos difíceis. Agradeço, especialmente, àquela que, além de amiga, dividiu comigo do início ao fim essa trajetória como colega. Camila, obrigada pelas incontáveis horas de estudo, palavras de incentivo e companheirismo ao longo desses cinco anos.

À Thais e Luna, por sempre estarem presentes e terem sido meu maior suporte no momento derradeiro da minha graduação. Não há palavras para expressar minha gratidão por todos os momentos de apoio, compreensão e carinho.

À Fernanda, por ter iniciado esse sonho da graduação junto comigo ainda no Ensino Médio. Muito me orgulha que tenhamos realizado, mesmo que de longe, tudo aquilo que planejamos.

Ao professor Flávio, por todo o conhecimento transmitido e por aceitar me orientar na construção desta monografia.

Nada disso seria possível sem cada um de vocês. Obrigada!

RESUMO

Este estudo tem por objetivo avaliar a capacidade preditiva de dois modelos consagrados, um da área de séries temporais, outro de machine learning. Utilizamos dados diários das regiões Centro-Oeste, Nordeste, Norte, Sudeste e Sul provenientes de uma empresa privada que usará os resultados do estudo como insumo. Propomos a utilização do modelo sazonal autorregressivo integrado de médias móveis (SARIMA) e o modelo de Redes Neurais para a previsão da taxa de ocupação hoteleira futura para dois grupos distintos de hotéis, em cada uma das regiões brasileiras. Para a comparação das previsões dos modelos candidatos foi utilizado o teste de Diebold-Mariano. Concluímos que, em geral, o modelo SARIMA produziu previsões superiores.

Palavras-chave: Previsão, Taxa de Ocupação, SARIMA, Redes Neurais

ABSTRACT

This study aims to evaluate the predictive capacity of two well-known models, one in the time series area, other in machine learning. We use daily data from a private company of the five Brazilian regions: Centro-Oeste, Nordeste, Norte, Sudeste and Sul. We propose the use of integrated autoregressive seasonal moving average model (SARIMA) and Neural Networks to predict the hotel occupancy rate for two distinct groups of hotels, in each of the Brazilian regions. To compare the predictions of the models we use Diebold-Mariano test. We conclude that, in general, the SARIMA model produced better predictions.

Keywords: Forecasting, Occupancy Rate, SARIMA, Neural Networks

LISTA DE ILUSTRAÇÕES

Figura 1 – Neurônio Biológico	22
Figura 2 – Modelo de um neurônio de McCulloch e Pitts	24
Figura 3 – Séries temporais da taxa de ocupação (%) da Região Nordeste (a)Tipo 1 (b) Tipo 2	29
Figura 4 – Previsão da Taxa de Ocupação (%) pelo modelo SARIMA, 60 pre- visões um passo à frente, para hotéis do Tipo 1. (a) Região Centro- Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.	32
Figura 5 – Previsão da Taxa de Ocupação (%) pelo modelo SARIMA, 60 pre- visões um passo à frente, para hotéis do Tipo 2. (a) Região Centro- Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.	33
Figura 6 – Previsão da Taxa de Ocupação (%) pelo modelo de Redes Neurais, 60 previsões um passo à frente, para hotéis do Tipo 1. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.	35
Figura 7 – Previsão da Taxa de Ocupação (%) pelo modelo de Redes Neurais, 60 previsões um passo à frente, para hotéis do Tipo 2. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.	36
Figura 8 – Séries temporais da taxa de ocupação (%) da Região Centro-Oeste (a)Tipo 1. (b) Tipo 2	46
Figura 9 – Séries temporais da taxa de ocupação (%) da Região Norte (a)Tipo 1. (b) Tipo 2	48
Figura 10 – Séries temporais da taxa de ocupação (%) da Região Sudeste (a)Tipo 1. (b) Tipo 2	49
Figura 11 – Séries temporais da taxa de ocupação (%) da Região Sul (a)Tipo 1. (b) Tipo 2	51
Figura 12 – Funções de autocorrelação e autocorrelação parcial da taxa de ocu- pação dos hotéis do tipo 1, para a Região Centro-Oeste.	53
Figura 13 – Funções de autocorrelação e autocorrelação parcial da taxa de ocu- pação dos hotéis do tipo 2, para a Região Centro-Oeste.	53
Figura 14 – Funções de autocorrelação e autocorrelação parcial da taxa de ocu- pação dos hotéis do tipo 1, para a Região Nordeste.	54
Figura 15 – Funções de autocorrelação e autocorrelação parcial da taxa de ocu- pação dos hotéis do tipo 2, para a Região Nordeste.	54
Figura 16 – Funções de autocorrelação e autocorrelação parcial da taxa de ocu- pação dos hotéis do tipo 1, para a Região Norte.	55

Figura 17 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Norte.	55
Figura 18 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Sudeste.	56
Figura 19 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Sudeste.	56
Figura 20 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Sul.	57
Figura 21 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Sul.	57
Figura 22 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Centro-Oeste	58
Figura 23 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Centro-Oeste	59
Figura 24 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Nordeste	60
Figura 25 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Nordeste	61
Figura 26 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Norte	62
Figura 27 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Norte	63
Figura 28 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Sudeste	64
Figura 29 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Sudeste	65
Figura 30 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Sul	66
Figura 31 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Sul	67

LISTA DE TABELAS

Tabela 1 – Taxa de Ocupação (%) média para 2017 e 2018, por região brasileira.	12
Tabela 2 – Taxa de ocupação (%) da região Centro-Oeste, por mês, em 2018 .	13
Tabela 3 – Taxa de ocupação (%) da região Nordeste, por mês, em 2018	14
Tabela 4 – Taxa de ocupação (%) da região Norte, por mês, em 2018	15
Tabela 5 – Taxa de ocupação (%) da região Sudeste, por mês, em 2018	16
Tabela 6 – Taxa de ocupação (%) da região Sul, por mês, em 2018	17
Tabela 7 – Tipos de funções de ativação	24
Tabela 8 – Parametrizações selecionadas	27
Tabela 9 – AIC dos modelos SARIMA sugeridos, Nordeste, hotéis tipo 1	30
Tabela 10 – AIC dos modelos SARIMA sugeridos, Nordeste, hotéis tipo 2	30
Tabela 11 – P-valor do teste de Diebold-Mariano para a Região Centro-Oeste .	37
Tabela 12 – Raiz do erro quadrado médio para a Região Centro-Oeste	37
Tabela 13 – P-valor do teste de Diebold-Mariano para a Região Nordeste	38
Tabela 14 – Raiz do erro quadrado médio para a Região Nordeste	38
Tabela 15 – P-valor do teste de Diebold-Mariano para a Região Norte	38
Tabela 16 – Raiz do erro quadrado médio para a Região Norte	39
Tabela 17 – P-valor do teste de Diebold-Mariano para a Região Sudeste	39
Tabela 18 – Raiz do erro quadrado médio para a Região Sudeste	39
Tabela 19 – P-valor do teste de Diebold-Mariano para a Região Sul	40
Tabela 20 – Raiz do erro quadrado médio para a Região Sul	40
Tabela 21 – AIC dos modelos SARIMA sugeridos, Centro-Oeste, hotéis tipo 1 .	47
Tabela 22 – AIC dos modelos SARIMA sugeridos, Centro-Oeste, hotéis tipo 2 .	47
Tabela 23 – AIC dos modelos SARIMA sugeridos, Norte, hotéis tipo 1	48
Tabela 24 – AIC dos modelos SARIMA sugeridos, Norte, hotéis tipo 2	49
Tabela 25 – AIC dos modelos SARIMA sugeridos, Sudeste, hotéis tipo 1	50
Tabela 26 – AIC dos modelos SARIMA sugeridos, Sudeste, hotéis tipo 2	50
Tabela 27 – AIC dos modelos SARIMA sugeridos, Sul, hotéis tipo 1	52
Tabela 28 – AIC dos modelos SARIMA sugeridos, Sul, hotéis tipo 2	52

SUMÁRIO

1	TAXA DE OCUPAÇÃO NO TURISMO	11
1.1	Introdução e Histórico Recente Brasileiro	11
1.2	Importância	12
1.3	Panorama por Região Brasileira	12
1.3.1	Região Centro-Oeste	12
1.3.2	Região Nordeste	13
1.3.3	Região Norte	14
1.3.4	Região Sudeste	15
1.3.5	Região Sul	16
2	MODELOS SARIMA	18
2.1	Modelos de Médias Móveis (MA)	18
2.2	Modelos autorregressivos (AR)	18
2.3	Modelos autorregressivos e de médias móveis (ARMA)	19
2.4	Modelos autorregressivos integrados de médias móveis (ARIMA)	19
2.5	Modelos sazonais autorregressivos integrados de médias móveis (SARIMA)	20
2.6	Estimação	20
2.7	Previsão	21
3	MODELOS DE REDES NEURAIS	22
3.1	Introdução	22
3.2	Neurônios Biológicos	22
3.3	Neurônios Artificiais	23
3.4	Estimação	24
3.5	Previsão	25
4	MODELAGEM	26
4.1	SARIMA	26
4.2	Redes Neurais	27
4.3	Comparação dos modelos	28
5	ANÁLISE EMPÍRICA	29
5.1	Resultados dos Modelos SARIMA	29
5.2	Resultados dos Modelos de Redes Neurais	33
5.3	Comparação das previsões dos modelos propostos	36
6	CONCLUSÃO	41

REFERÊNCIAS	42
APÊNDICE A	44
APÊNDICE B	46
APÊNDICE C	53
APÊNDICE D	58

1 TAXA DE OCUPAÇÃO NO TURISMO

1.1 Introdução e Histórico Recente Brasileiro

O setor de turismo ocupa papel relevante na economia mundial, sendo uma das atividades com maior representatividade econômica, ao lado da indústria do petróleo (GORINI; MENDES, 2005) . A razão pela qual isso ocorre, segundo Țițu, Răulea e Țițu (2016) é a receptividade do setor às mudanças do mundo contemporâneo, evoluindo sob seus impactos de maneira dinâmica.

Diversas instituições nacionais e internacionais dedicam estudos à área turística. Um dos mais reconhecidos é o World Travel & Tourism Council (WTTC), órgão que representa o setor privado de viagens e turismo em todo o mundo. Nos últimos dez anos, segundo dados do WTTC (2018), a contribuição direta média do setor no país foi equivalente a aproximadamente 3% do PIB brasileiro. Além disso, o WTTC acrescenta que o segmento de viagens e turismo gerou diretamente 2 337 000 empregos em 2017, no Brasil, o que equivale a 2,6% do total de empregos gerados pela economia brasileira no ano em questão. Ainda segundo o WTTC, no entanto, apesar dos bons resultados, o país ainda ocupava apenas a 118ª posição da classificação mundial de importância relativa direta do turismo ao PIB, em 2017.

No Brasil, assim como em grande parte dos países emergentes, o turismo é um campo fundamental e em constante crescimento. Wada e Camargo (2006) ressaltam que, antigamente, no Brasil, viajar e frequentar hotéis era caro e privilégio de algumas classes sociais, apenas. No entanto, com o desenvolvimento da sociedade e o apelo turístico do país, viajar deixou de ser exclusivamente uma atividade de lazer e passou a integrar também o mundo do trabalho. Acrescentam ainda que a união dessas duas vertentes de interesse deu origem a uma segmentação do mercado hoteleiro, permitindo que todos os níveis de renda das mais diversas classes sociais fossem atendidos.

Essa transformação da hotelaria exigiu mudanças na forma de olhar para os serviços oferecidos e lidar com os novos desafios oriundos dessa diversidade. Neste contexto, um dos indicadores mais importantes para mensurar o crescimento dos meios de hospedagem é a taxa de ocupação. Obtida dividindo-se o número total de quartos ocupados pelo número de quartos disponíveis no período, ela serve como um termômetro do interesse dos turistas em visitar determinada localidade. Com isso, oferece subsídios para tomadas de decisão acerca das políticas de investimento da hotelaria.

Conforme dados do inFOBH (2018) na Tabela 1, o Brasil, no ano de 2018, apresentou aumento de 6,1% na taxa de ocupação em relação ao mesmo período de 2017. Quanto à análise por região, registrou-se acréscimo no indicador em todas as regiões brasileiras quando realizado comparativo com o ano de 2017. Destaca-se

a região Norte que computou aumento de 17,7% na taxa de ocupação. O estudo considerou 389 hotéis responsáveis por 60 709 unidades habitacionais (UHs).

Tabela 1 – Taxa de Ocupação (%) média para 2017 e 2018, por região brasileira.

Região	UHs	Hotéis	2017	2018	Variação
CO	5463	33	56,33	59,34	5,3%
NE	6371	44	60,74	63,27	4,2%
N	2458	16	47,71	56,16	17,7%
SE	34 973	211	54,21	57,80	6,6%
S	11 444	85	58,39	60,79	4,1%
Brasil	60 709	389	55,61	59,01	6,1%

Fonte: Adaptado da edição 137 do inFOHB, dezembro de 2018

1.2 Importância

O desenvolvimento da indústria do turismo não se sustenta a não ser que haja capacidade suficiente disponível. A previsão imprecisa geralmente leva a uma relação desequilibrada entre demanda e suprimento no setor, o que, com a crescente competição entre redes hoteleiras, significa grandes perdas para o hotel (CHOW; SHYU; WANG, 1998).

A previsão da taxa de ocupação é um passo muito importante no processo de tomada de decisão de hotéis. Algumas estratégias populares como o Gerenciamento de Receita, por exemplo, têm a previsão como uma atividade vital para preços dinâmicos. Sem uma previsão precisa, os erros de precificação podem afetar negativamente o desempenho financeiro do hotel (CAICEDO; PAYARES, 2016).

A partir de uma investigação histórica da taxa de ocupação, é possível prever o comportamento deste indicador e promover mudanças e investimentos que possibilitem o atendimento de futuras demandas. A sazonalidade presente no setor hoteleiro auxilia nesse processo, possibilitando a detecção de picos e quedas de ocupação em determinados períodos (MATTOS, 2004).

1.3 Panorama por Região Brasileira

1.3.1 Região Centro-Oeste

A região Centro-Oeste conta com importantes exemplos de patrimônios naturais da humanidade, como os Complexos de Áreas Protegidas do Pantanal e do Cerrado. Por

conta disso, vem recebendo a implantação de infra-estrutura urbana em locais isolados, buscando dar suporte ao setor turístico, com a oferta de produtos que respondam a padrões internacionais. (PAES; BRUHNS; SERRANO, 2007). Isso faz com que a região seja um ótimo destino para aqueles que desejam praticar turismo de aventura.

Conforme a Tabela 2, nota-se que a região Centro-Oeste manteve uma média de ocupação próxima a 60% durante o ano de 2018. Os meses de dezembro a março foram os de menor expressão do indicador. Além disso, com exceção de fevereiro e maio, todos os outros meses apresentaram crescimento na taxa de ocupação em comparação com o período homólogo.

Tabela 2 – Taxa de ocupação (%) da região Centro-Oeste, por mês, em 2018

Mês	UHs	Hotéis	2017	2018	Variação
Janeiro	5878	36	40,38	44,69	10,7%
Fevereiro	5814	35	51,98	51,29	-1,3%
Março	6214	37	53,09	58,92	11%
Abril	6333	37	53,18	60,42	13,6%
Mai	6963	37	58,35	54,94	-5,8%
Junho	6395	38	56,28	61,06	8,5%
Julho	6416	38	58,12	60,35	3,8%
Agosto	6384	38	60,12	65,56	9,0%
Setembro	6489	38	59,37	63,59	7,1%
Outubro	6078	36	61,28	61,45	0,3%
Novembro	6691	39	62,54	63,63	1,7%
Dezembro	6737	39	48,28	51,53	6,7%

Fonte: Adaptado das edições mensais do inFOHB do ano de 2018

1.3.2 Região Nordeste

O Nordeste é uma das regiões turísticas que mais cresce no país. No entanto, o turismo concentra-se nos litorais e áreas restritas, não atingindo todo o território nordestino. Há mais de quatro décadas, vêm se implantando, nessa região, experiências de desenvolvimento regional e local, com destaque ao setor turístico. Vê-se no turismo uma das saídas para a economia nordestina, desde que se invista no maior potencial

da região, a população, e se proteja, conserve ou recupere seu patrimônio natural e cultural. (CORIOLANO; ALMEIDA, 2007).

De acordo com a Tabela 3 a região Nordeste apresentou uma média de ocupação próxima a 63% durante o ano de 2018. Os meses de férias escolares no Brasil (julho, dezembro e janeiro) se destacaram em relação à ocupação tanto no ano de referência quanto no ano anterior. A taxa de ocupação apresentou variação anual positiva para a maioria dos meses, com destaque para abril que apresentou aumento de 18% em relação ao mesmo período de 2017.

Tabela 3 – Taxa de ocupação (%) da região Nordeste, por mês, em 2018

Mês	UHs	Hotéis	2017	2018	Variação
Janeiro	7091	51	68,99	76,04	10,2%
Fevereiro	6831	49	59,01	63,92	8,3%
Março	7459	53	60,12	64,18	6,8%
Abril	6615	47	52,70	62,18	18%
Mai	7195	47	53,62	53,94	0,6%
Junho	6657	47	52,20	56,20	7,7%
Julho	7283	51	64,31	66,18	2,9%
Agosto	7160	50	61,97	61,82	-0,2%
Setembro	7085	49	62,33	65,46	5%
Outubro	6540	46	65,75	64,75	-1,5%
Novembro	6588	46	65,38	68,64	5%
Dezembro	6886	47	60,13	62,38	3,7%

Fonte: Adaptado das edições mensais do inFOHB do ano de 2018

1.3.3 Região Norte

Um dos destinos mais famosos e com maior procura da região Norte é a Amazônia. Nesse contexto, os parques nacionais emergem como pontos focais para o desenvolvimento turístico na região e veículos potenciais de inclusão social. Como patrimônios naturais de impacto global, contribuem para o desenvolvimento regional, a economia local e a inclusão social (IRVING, 2006).

De acordo com a Tabela 4 a região Norte apresentou uma média de ocupação próxima a 56% durante o ano de 2018. Essa foi a menor média dentre as cinco regiões

brasileiras. Diferentemente de outras regiões, os meses de dezembro, janeiro e fevereiro são os que apresentam menor taxa de ocupação. Isso explica-se pelo fato de esse ser o início do período chuvoso na região. A taxa de ocupação apresentou variação anual positiva para a maioria dos meses, com destaque negativo para fevereiro que apresentou redução de 3,1% em relação ao mesmo período de 2017.

Tabela 4 – Taxa de ocupação (%) da região Norte, por mês, em 2018

Mês	UHs	Hotéis	2017	2018	Variação
Janeiro	2960	20	37,78	47,15	24,8%
Fevereiro	2694	17	47,92	46,46	-3,1%
Março	3298	21	46,95	54,79	16,7%
Abril	2521	17	41,00	54,22	32,2%
Maio	2778	17	42,73	56,64	32,5%
Junho	2646	17	46,66	58,01	24,3%
Julho	2633	17	49,08	58,97	20,1%
Agosto	2638	17	55,12	64,11	16,3%
Setembro	2653	17	50,90	62,31	22,4%
Outubro	2427	17	53,68	62,10	15,7%
Novembro	2637	17	61,56	63,55	3,2%
Dezembro	2643	17	45,22	47,48	5%

Fonte: Adaptado das edições mensais do inFOHB do ano de 2018

1.3.4 Região Sudeste

A região Sudeste é a principal receptora do turismo de negócios do Brasil, e conta com um turismo de lazer inter-regional vigoroso que continua a crescer, alimentado pela concentração de atividades econômicas (RESENDE et al., 2015). Os dois maiores exemplos dessas duas características distintas são as cidades do Rio de Janeiro e São Paulo. A capital fluminense é internacionalmente conhecida por suas belezas naturais e reconhecida como um grande polo de turismo cultural. Já a capital paulista é o principal centro financeiro e maior cidade do Brasil, contando com a maior rede hoteleira do país.

Conforme a Tabela 5, a região Sudeste registrou aumento na taxa de ocupação para todos os meses de 2018 em relação ao ano anterior. Os meses de janeiro,

novembro e dezembro despontaram com variações anuais acima de 11%. A média de ocupação da região para o ano de referência foi de cerca de 58%.

Tabela 5 – Taxa de ocupação (%) da região Sudeste, por mês, em 2018

Mês	UHs	Hotéis	2017	2018	Variação
Janeiro	38 050	234	46,89	52,29	11,5%
Fevereiro	37 417	232	49,77	52,88	6,2%
Março	42 246	251	54,43	58,46	7,4%
Abril	41 284	250	53,20	58,01	9,0%
Maiο	43 570	249	53,83	54,16	0,6%
Junho	41 222	246	53,45	55,10	3,1%
Julho	42 849	258	55,12	58,19	5,6%
Agosto	43 418	260	56,63	61,96	9,4%
Setembro	42 098	261	55,52	61,20	10,2%
Outubro	40 568	254	56,24	59,12	5,1%
Novembro	41 411	258	57,03	63,43	11,2%
Dezembro	41 617	261	48,51	54,85	13,1%

Fonte: Adaptado das edições mensais do inFOHB do ano de 2018

1.3.5 Região Sul

A região sul tem a forte presença do patrimônio do imigrante no Brasil dentro do contexto de diversidade étnica e pluralidade cultural (PAIVA, 2015).

De acordo com a Tabela 6, a região Sul registrou aumento na taxa de ocupação para todos os meses de 2018 em relação ao ano anterior. A média do indicador para o ano de referência foi de aproximadamente 61%, o segundo maior entre as regiões brasileiras. Os meses de inverno, em geral, apresentaram taxas de ocupação acima de 60%, o que pode ser explicado pelos atrativos oriundos do frio rigoroso nesta época do ano.

Tabela 6 – Taxa de ocupação (%) da região Sul, por mês, em 2018

Mês	UHs	Hotéis	2017	2018	Variação
Janeiro	12 562	96	56,35	61,56	9,2%
Fevereiro	12 520	96	53,03	55,35	4,4%
Março	12 880	99	57,40	60,21	4,9%
Abril	13 084	99	56,19	61,10	8,7%
Maiο	13 515	98	52,07	52,22	0,3%
Junho	13 237	100	54,86	56,62	3,2%
Julho	13 318	101	61,58	64,19	4,2%
Agosto	13 447	102	58,39	62,68	7,4%
Setembro	13 455	102	58,53	62,23	6,3%
Outubro	13 254	100	62,17	63,64	2,4%
Novembro	12 544	95	63,82	66,87	4,8%
Dezembro	12 604	95	55,95	59,04	5,5%

Fonte: Adaptado das edições mensais do inFOHB do ano de 2018

2 MODELOS SARIMA

Uma série temporal é um conjunto de pontos coletados em sequência ao longo do tempo. Por conta disso, é esperado que a série tenha correlação seriada no tempo. Com isso, o propósito dos modelos ARIMA (Autorregressivos Integrados de Médias Móveis) é justamente identificar o comportamento da autocorrelação entre os valores da série temporal (MANTOVANI, 2004).

Esses modelos são compostos por uma combinação de três componentes: autorregressivo (AR), médias móveis (MA) e filtro de integração (I).

2.1 Modelos de Médias Móveis (MA)

Supondo que a_t seja um ruído branco (ver Apêndice A). Então a série Y_t é um processo de médias móveis de ordem q se:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

onde θ_i é uma constante e $\mu = E(Y_t)$.

Neste trabalho consideramos $\mu = 0$. A variância é dada por $Var(Y_t) = \sigma_a^2 \sum_{i=0}^q \theta_i^2$, se os a_t 's forem independentes.

O operador defasagem, representado por B, e definido por $B^j Y_t = Y_{t-j}$, para todo j , quando aplicado a uma variável, fornece o valor anterior na série temporal.

Box e Jenkins (1970) propõem restrições a θ_i através da chamada condição de invertibilidade. Podemos expressar essa condição através do operador defasagem como

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t = Y_t,$$

onde $\theta(B)$ é o polinômio de ordem q em B. Para que a condição de invertibilidade seja atendida, as raízes da equação

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0$$

devem ser maiores que um.

2.2 Modelos autorregressivos (AR)

Os modelos autorregressivos são modelos de regressão múltipla, nos quais Y_t é regredido de seus valores passados. Daí o termo autorregressivo (MANTOVANI, 2004).

Suponha que a_t seja um ruído branco. Então Y_t é um processo autorregressivo de ordem p se

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t,$$

onde ϕ_i é uma constante para $i = 1, 2, \dots, p$, e c é um valor constante.

Podemos reescrever a equação anterior utilizando o operador defasagem B , descrito em 2.1 como

$$(1 - \phi_1 B - \phi_2 B - K - \phi_p B^p)Y_t = \phi(B)Y_t = a_t.$$

2.3 Modelos autorregressivos e de médias móveis (ARMA)

Em alguns casos pode ser necessário utilizar um alto número de parâmetros em modelos unicamente autorregressivos ou unicamente de médias móveis. Nessas situações, é mais adequado que utilizemos uma mistura de componentes AR e MA, gerando um modelo ARMA (MANTOVANI, 2004).

Seja Y_t uma série temporal descrita por

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + K + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - K - \theta_q a_{t-q} + c,$$

então dizemos que Y_t é um ARMA(p, q).

A função de autocorrelação (ver Apêndice A) dos modelos ARMA(p, q) possui características semelhantes às da função MA(q) para as defasagens $k < q$, dado que a “memória” do componente de médias móveis dura exatos q períodos. Para defasagens maiores que $k + 1$, as características se assemelham a um AR(p) (MANTOVANI, 2004).

2.4 Modelos autorregressivos integrados de médias móveis (ARIMA)

Caso a série temporal não tenha a condição de estacionariedade atendida (ver Apêndice A), é necessário transformá-la. Box e Jenkins (1970) sugerem que tomemos sucessivas diferenças da série original até que se torne estacionária. Normalmente, basta tomar uma ou duas diferenças para que a série se torne estacionária (MANTOVANI, 2004).

A primeira diferença é definida por

$$\nabla Y_t = Y_t - Y_{t-1},$$

enquanto a segunda diferença é dada por

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} - Y_{t-2}.$$

A inclusão do termo de diferenças dá origem ao modelo ARIMA(p, d, q), definido como

$$W_t = \phi_1 W_{t-1} + K + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - K - \theta_q a_{t-q}$$

em que $W_t = \nabla^d Y_t$, onde $\nabla^d Y_t$ é o operador de diferenças.

Podemos reescrever o modelo acima utilizando o operador de defasagem B como

$$(1 - \phi_1 B - K - \phi_p B^p)W_t = (1 - \theta_1 B - K - \theta_q B^q)a_t.$$

2.5 Modelos sazonais autorregressivos integrados de médias móveis (SARIMA)

Frequentemente há componentes em séries temporais que se repetem com um período de s unidades de tempo. Box e Jenkins (1970) generalizaram os modelos ARIMA para lidar com a sazonalidade. Esses modelos são conhecidos como SARIMA (Seasonal Autoregressive Integrated Moving Average) (MANTOVANI, 2004).

O modelo mais geral é dado por:

$$(1 - \phi_1 B - K - \phi_p B^p)(1 - \Phi_1 B^s - K - \Phi_P B^{P_s})(1 - B)^d(1 - B^s)^D Y_t = (1 - \theta_1 B - K - \theta_q B^q)(1 - \Theta_1 B^s - K - \Theta_Q B^{Q_s}) a_t,$$

em que:

$(1 - \phi_1 B - K - \phi_p B^p)$ é o polinômio auto-regressivo não-sazonal de ordem p ;

$(1 - \Phi_1 B^s - K - \Phi_P B^{P_s})$ é o polinômio auto-regressivo sazonal de ordem P e período sazonal s ;

$(1 - B)^d$ é o polinômio de integração não-sazonal de ordem d ;

$(1 - B^s)^D$ é o polinômio de integração sazonal de ordem D e período sazonal s ;

$(1 - \theta_1 B - K - \theta_q B^q)$ é o polinômio não-sazonal de médias móveis de ordem q ;

$(1 - \Theta_1 B^s - K - \Theta_Q B^{Q_s})$ é o polinômio sazonal de médias móveis de ordem Q e período sazonal s ;

No Apêndice A encontramos alguns conceitos básicos de séries temporais.

2.6 Estimação

Considere um modelo ARIMA(p, d, q) e coloque seus $p+q+1$ parâmetros no vetor $\underline{\xi} = (\underline{\phi}, \underline{\theta}, \sigma_a^2)$, onde $\underline{\phi} = (\phi_1, K, \phi_p)$ e $\underline{\theta} = (\theta_1, K, \theta_q)$. Quando $d > 0$, supomos $\mu = 0$. Caso contrário, μ é incluído como mais um parâmetro a ser estimado.

Para estimar $\underline{\xi}$, utiliza-se o método de máxima verossimilhança. Dadas as observações X_1, \dots, X_n , considere a função de verossimilhança $L(\underline{\xi} | X_1, K, X_n)$, como função de $\underline{\xi}$. Os estimadores de máxima verossimilhança (EMV) de $\underline{\xi}$ serão os valores de maximizam L ou $l = \log L$.

Tomando-se d diferenças para alcançar estacionariedade, fica-se com $n = N - d$ observações W_t, \dots, W_n , onde $W_t = \Delta^d Y_t$. Como o modelo ARMA(p, q) resultante é estacionário e invertível, tem-se a equação

$$a_t = \tilde{W}_t - \phi_1 \tilde{W}_{t-1} - K - \phi_p \tilde{W}_{t-p} + \theta_1 a_{t-1} + K + \theta_q a_{t-q},$$

onde $\tilde{W}_t = W_t - \mu_W$.

Para calcular os a_t 's, é necessário obter valores iniciais para os \tilde{W} 's e para os a 's. Podemos obter de forma condicional, substituindo por valores que supomos razoáveis; ou incondicionalmente, estimando valores iniciais de uma amostra de dados (ver Mantovani, 2004).

2.7 Previsão

Suponha que temos as observações $(\dots, X_{t-2}, X_{t-1}, X_t)$, até o instante t , $t = 1, 2, \dots, T$. O interesse é prever um valor para X_{t+h} , $h \geq 1$. A previsão de origem t e horizonte h é denotada por $\hat{x}_t(h)$ (ver Mantovani, 2004).

Em um modelo ARIMA, é razoável supor que $\hat{x}_t(h)$ é uma função das observações até o instante t , conseqüentemente, uma função de a_t, a_{t-1}, \dots . Sendo a melhor previsão indicada por

$$\hat{x}_t(h) = \Psi_h^* a_t + \Psi_{h+1}^* a_{t-1} + \Psi_{h+2}^* a_{t-2} + K$$

os pesos Ψ_j^* devem ser determinados a fim de minimizar o erro quadrático médio (EQM), que é dado por

$$E(X_{t+h} - \hat{x}_t(h))^2 = E\left(\sum_{j=0}^{\infty} \Psi_j a_{t+h+j} - \sum_{j=0}^{\infty} \Psi_{h+j}^* a_{t-j}\right)^2$$

O erro de previsão é dado por

$$e_t(h) = X_{t+h} - \hat{x}_t(h)$$

$$= \Psi_0 a_{t+h} + \Psi_1 a_{t+h-1} + K + \Psi_{h-1} a_{t+1} - \sum_{j=0}^{\infty} (\Psi_{h+j} - \Psi_{h+j}^*) a_{t-j},$$

O EQM pode ser escrito como

$$E(e_t(h))^2 = (1 + \Psi_1^2 + K + \Psi_{h-1}^2) \sigma_a^2 + \sum_{j=0}^{\infty} (\Psi_{h+j} - \Psi_{h+j}^*)^2 \sigma_a^2,$$

uma vez que os a_t 's não são autocorrelacionados e têm média zero.

Assim, a previsão de EQM mínimo é dada por

$$\hat{x}_t(h) = \Psi_h a_t + \Psi_{h+1} a_{t-1} + \Psi_{h+2} a_{t-2} + K.$$

3 MODELOS DE REDES NEURAIS

3.1 Introdução

Redes neurais artificiais têm o propósito de fornecer um modelo que realize processamento de informações como o cérebro humano. O cérebro recebe informações e as converte em respostas, por meio de um sistema massivo de computação paralela (altamente conectado) baseado em neurônios, sua célula básica (OLIVEIRA, 2010).

Segundo Fausett (1994), uma rede neural artificial é um sistema de processamento de informações que possui certas características de desempenho em comum com redes neurais biológicas, com base nas seguintes suposições:

- 1) O processamento de informação ocorre em muitos elementos simples chamados neurônios.
- 2) Os sinais são transmitidos entre os neurônios pelas conexões.
- 3) Cada conexão tem um peso associado, o qual, em uma rede neural típica, multiplica o sinal transmitido.
- 4) Cada neurônio aplica uma função de ativação (geralmente não linear) à sua entrada para determinar seu sinal de saída.

3.2 Neurônios Biológicos

Para Fausett (1994), um neurônio biológico possui três tipos de componentes particularmente interessantes para o entendimento de um neurônio artificial: dendritos, soma e axônio. Conforme a Figura 1, os dendritos são responsáveis por receber os sinais de outros neurônios. A soma é o corpo do neurônio e realiza a função de somar os sinais que chegam. Quando são recebidas entradas suficientes, as células são acionadas e transmitem um sinal sobre seu axônio para outras células.

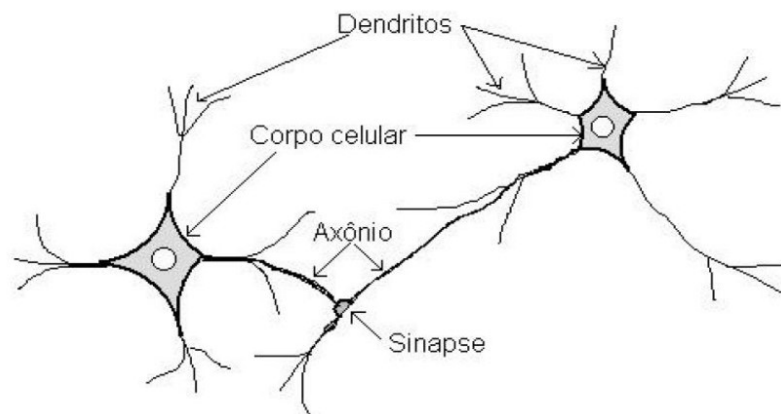


Figura 1 – Neurônio Biológico

Fonte: BONINI NETO; BONINI (2010)

Uma característica em comum entre os neurônios artificiais e biológicos é a tolerância dos sistemas ao erro. Os humanos nascem com cerca de 100 bilhões de neurônios. Muitos desses se localizam no cérebro e não são substituídos quando morrem. No entanto, apesar da perda contínua de neurônios, nós continuamos aprendendo. Isso ocorre porque outros neurônios podem ser treinados para assumir as funções das células danificadas. De maneira similar, redes neurais artificiais podem ser desenhadas para serem insensíveis a pequenos danos à rede, podendo ela ser retreinada em casos de dano significativo, como perda de dados ou de conexões, por exemplo (FAUSETT, 1994).

3.3 Neurônios Artificiais

Segundo Ferneda (2006), uma rede neural é caracterizada por três elementos:

- 1) Um conjunto de D conexões de entrada (x_1, x_2, \dots, x_D) , caracterizadas por pesos (w_1, w_2, \dots, w_D) .
- 2) Um somador (Σ) para acumular os sinais de entrada.
- 3) Função de ativação que limita o intervalo permissível de amplitude do sinal de saída (y) a um valor fixo.

A Figura 2 mostra o modelo de um neurônio artificial de McCulloch e Pitts. Conforme Rauber (2005), a informação fornecida por outros neurônios entra em D entradas x_j (sinapses) no neurônio processador. Esse processamento é uma combinação linear das entradas. A cada uma dessas entradas x_j associa-se um valor w_j que reflete a sua importância. O resultado dessa combinação linear é o valor *net*:

$$net = w_1x_1 + w_2x_2 + \dots + w_Dx_D = \sum_{j=1}^D w_jx_j$$

O valor *net* resultante é enviado para a função de ativação, que define a saída y do neurônio.

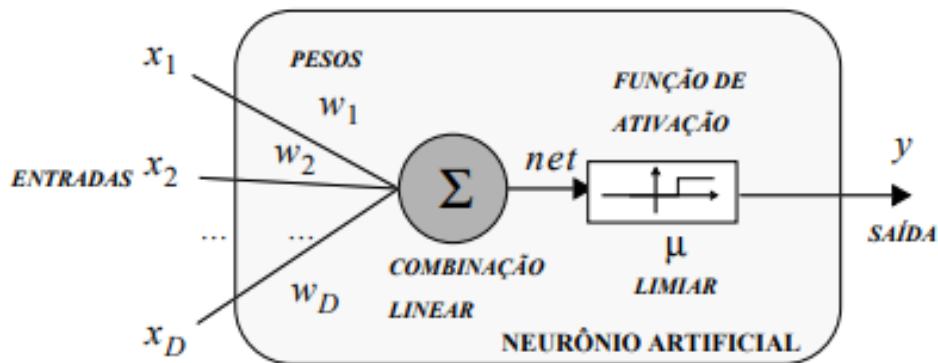


Figura 2 – Modelo de um neurônio de McCulloch e Pitts
Fonte: RAUBER (2005)

Existem diferentes tipos de funções de ativação, como mostrado na Tabela 7:

Tabela 7 – Tipos de funções de ativação

Tipo	Função de Ativação	Derivada
Linear	$f(x) = px$	$f'(x) = p$
Logística	$f(x) = \frac{e^{px}}{1+e^{px}} = \frac{1}{1+e^{-px}}$	$f'(x) = pf(x)(1 - f(x))$
Tangente Hiperbólica	$f(x) = \frac{e^{px} - e^{-px}}{e^{px} + e^{-px}} = tg(px)$	$f'(x) = p(1 - f(x)^2)$
Arco-Tangente	$f(x) = arctg(px)$	$f'(x) = p \frac{1}{1+x^2}$

A função logística é especialmente destacável por possuir um conjunto de propriedades úteis nos cálculos relacionados à aprendizagem dos pesos e ao mapeamento realizado pela rede (RAUBER, 2005).

Quanto à classificação, redes neurais podem ser recorrentes ou não-recorrentes. Neste estudo, trabalharemos com redes recorrentes, aquelas que contêm laços de realimentação, ou seja, contêm conexões das saídas de uma determinada camada para a entrada da mesma ou de camadas anteriores (BONINI NETO; BONINI, 2010).

3.4 Estimação

Os parâmetros no modelo RNA(k,q)

$$y_t = x_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j) + a_t$$

podem ser estimados pela minimização da função soma dos quadrados dos resíduos

$$Q_n(\theta) = \sum_{t=1}^n [y_t - F(x_t; \theta)]^2,$$

onde

$$F(x_t; \theta) = x_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j),$$

com θ sendo o vetor que consiste dos $k+1+q(k+2)$ parâmetros em $\phi, \beta_1, K, B_q, \gamma_1, K, \gamma_q$.

Através da estimação recursiva, dados valores iniciais $\hat{\theta}_0$, o valor do gradiente para a primeira observação, $\nabla q_1(\hat{\theta}_0)$, é calculado e usado para obter um novo conjunto de estimativas $\hat{\theta}_1$, que são usadas para se obter o gradiente para a segunda observação, $\nabla q_2(\hat{\theta}_1)$, e assim por diante. Esse processo se repete até que os parâmetros estimados tenham convergido. Como $\nabla q_t(\theta) = -2\nabla F(x_t; \theta)[y_t - F(x_t; \theta)]$, a fórmula de atualização das estimativas tem a forma

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \lambda \nabla F(x_{t+1}; \hat{\theta}_t)[y_{t+1} - F(x_{t+1}; \hat{\theta}_t)],$$

onde $\hat{\theta}_t$ representa as estimativas dos parâmetros baseadas nas t primeiras observações (ver Mantovani, 2004).

3.5 Previsão

Uma previsão y_{n+1} um passo a frente pode ser calculada diretamente de um modelo RNA(p, q) como

$$\hat{y}_{t+1|t} = x_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j),$$

onde $x_t = (1, y_t, K, y_{t-p+1})'$. Sob a suposição inicial adicional de que os erros a_t são normalmente distribuídos, o erro de previsão um passo a frente $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$ é também normalmente distribuído e os intervalos de predição podem ser construídos de maneira usual.

Para previsões diversos passos à frente, não existem expressões prontas para $\hat{y}_{t+h|t}$, onde $h > 1$, e deve-se utilizar simulações para a obtenção das previsões e intervalos de predição (MANTOVANI, 2004).

4 MODELAGEM

Neste estudo utilizamos dados diários da taxa de ocupação hoteleira das cinco regiões brasileiras: Centro-Oeste, Nordeste, Norte, Sudeste e Sul. As informações foram analisadas com o consentimento de uma empresa privada que utilizará os resultados do estudo como insumo.

Em concordância com o objetivo da empresa provedora dos dados citados acima, optou-se por modelar as observações separadamente conforme o tipo de hotel. Os tipos de hotéis não são definidos explicitamente por conta da sensibilidade dos dados disponibilizados pela empresa. Sendo assim, optamos por representá-los pelas indicadoras 1 e 2. Para cada região brasileira, modelamos duas séries temporais, a primeira para hotéis do tipo 1 e a segunda para hotéis do tipo 2. Cada conjunto de dados possui 699 observações, com exceção dos hotéis do tipo 1 da região Norte, para os quais obtivemos apenas 363 dias sequenciais. Não houve dados faltantes.

Utilizamos dois modelos para o ajuste e previsão das séries temporais em questão, SARIMA e Redes Neurais Artificiais (RNAs), para os hotéis dos grupos 1 e 2 de cada uma das cinco regiões brasileiras. Separamos *datasets* para comparação posterior e realizamos 60 previsões um passo à frente.

As análises estatísticas foram realizadas através do *software* estatístico R. Para o ajuste dos modelos SARIMA e Redes Neurais foram utilizadas as funções *Arima* e *nnetar*, respectivamente, ambas do pacote *forecast*. A função *nnetar* ajusta redes neurais feed-forward com uma única camada oculta e entradas defasadas para prever séries temporais univariadas.

É importante ressaltar que as estimativas não levam em consideração possíveis mudanças estruturais nas regiões em questão como, por exemplo, abertura de novos hotéis ou maior disponibilidade de voos.

4.1 SARIMA

Várias parametrizações do modelo foram testados para cada um dos grupos estudados. Para realizarmos as previsões, selecionamos as parametrizações com menor valor no Critério de Informação de Akaike (AIC). A Tabela 8 apresenta as parametrizações selecionados para cada tipo de hotel das cinco regiões brasileiras.

Tabela 8 – Parametrizações selecionadas

Região	Tipo de Hotel	Modelo Selecionado
Centro-Oeste	1	SARIMA(3,1,3)x(1,1,1) ₇
	2	SARIMA(5,1,1)x(1,1,1) ₇
Nordeste	1	SARIMA(5,0,1)x(3,1,1) ₇
	2	SARIMA(5,1,1)x(1,1,1) ₇
Norte	1	SARIMA(1,1,1)x(0,1,1) ₇
	2	SARIMA(2,1,1)x(1,1,1) ₇
Sudeste	1	SARIMA(2,1,3)x(0,1,2) ₇
	2	SARIMA(2,1,2)x(1,1,1) ₇
Sul	1	SARIMA(5,0,1)x(1,1,1) ₇
	2	SARIMA(3,1,3)x(1,1,1) ₇

Após a seleção pelo critério de AIC, checamos as suposições para os resíduos do modelo selecionado. Para a estacionariedade, realizamos o teste da raiz unitária de Dickey-Fuller. Para a autocorrelação, o teste de Ljung-Box. A normalidade checamos graficamente através do gráfico de QQ-Plot dos resíduos (ver Apêndice A).

4.2 Redes Neurais

Podemos utilizar os valores defasados de uma série temporal como entradas para uma rede neural. Chamamos esse modelo de NNAR. Um NNAR(k, q) indica que há k entradas e q nós na camada oculta. Portanto, um NNAR(k, q) é um modelo que utiliza as k últimas observações como entradas para prever a saída e possui q neurônios na camada oculta.

A função *nnetar* ajusta um modelo NNAR(k, q). Se não especificados, os valores de k e q são selecionados automaticamente. Para a realização de previsões, a rede é aplicada iterativamente. Os intervalos de predição são computados simulando amostras do modelo selecionado, gerando aleatoriamente valores de ε_t de uma distribuição normal. Se gerarmos repetidas simulações, podemos adquirir conhecimento acerca da distribuição dos valores futuros com base na rede neural ajustada. A função *nnetar* gera, por padrão, 1000 simulações (HYNDMAN, 2016).

4.3 Comparação dos modelos

Para a comparação dos modelos propostos, utilizamos o teste de Diebold-Mariano, através da função *dm.test* do *software* R. A função implementa o teste modificado proposto por Harvey, Leybourne e Newbold em 1997 (HYNDMAN, 2016).

A hipótese nula do teste é que os dois métodos possuem a mesma acurácia de previsão, portanto, rejeitar a hipótese nula significa que um dos métodos tem maior precisão que o outro. No caso de rejeição da hipótese nula, para selecionar o melhor modelo, verificamos a raiz do erro quadrático médio (RMSE) de previsão dos modelos.

5 ANÁLISE EMPÍRICA

5.1 Resultados dos Modelos SARIMA

Como método de avaliação inicial, analisamos as séries temporais diárias das taxas de ocupação dos dois grupos de hotéis estudados. A Figura 3 mostra a série da Região Nordeste.

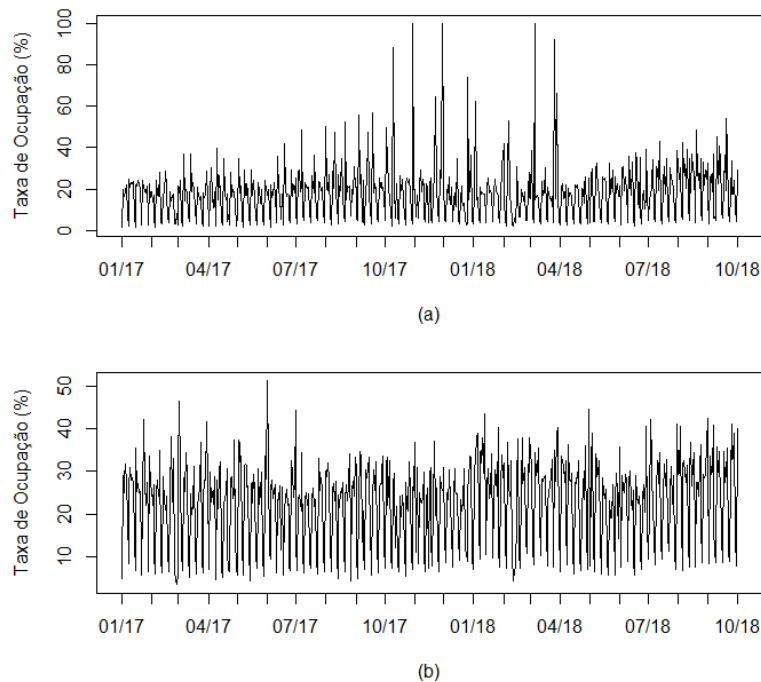


Figura 3 – Séries temporais da taxa de ocupação (%) da Região Nordeste
(a) Tipo 1 (b) Tipo 2

Realizamos o teste da raiz unitária de Dickey-Fuller. Na Região Nordeste, para as séries de ambos os tipos de hotéis, obtivemos um p-valor < 0.01 e rejeitamos a hipótese de presença de raiz unitária, assumindo as séries das taxas de ocupação dos dois grupos como estacionárias.

No Apêndice C, a Figura 14 mostra as funções de autocorrelação e autocorrelação parcial da série (a), enquanto a Figura 15 mostra as funções de autocorrelação e autocorrelação parcial da série (b). Atendidas as suposições de estacionariedade, diversos modelos ARIMA foram testados para ambas as séries.

Para os hotéis do tipo 1, a Tabela 9 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo $SARIMA(5,0,1) \times (3,1,1)_7$ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 9 – AIC dos modelos SARIMA sugeridos, Nordeste, hotéis tipo 1

Modelo Sugerido	AIC
SARIMA(5,0,1)x(3,1,1)₇	4749.11
SARIMA(5,1,1)x(1,1,2) ₇	4750.14
SARIMA(5,1,1)x(2,1,2) ₇	4750.75

Para os hotéis do tipo 2, a Tabela 10 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(5,1,1)x(1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Sazonalidade semanal nos dados também foi detectada. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 10 – AIC dos modelos SARIMA sugeridos, Nordeste, hotéis tipo 2

Modelo Sugerido	AIC
SARIMA(5,1,1)x(1,1,1)₇	3934.11
SARIMA(5,1,1)x(2,1,2) ₇	3936
SARIMA(5,1,1)x(3,1,1) ₇	3935.83

As análises de seleção de modelos para as demais regiões encontram-se no Apêndice B. Foram realizadas 60 previsões um passo à frente para as taxas de ocupação (%) dos dois grupos de hotéis, das cinco regiões brasileiras, seguindo os modelos selecionados.

As previsões para os hotéis do Tipo 1, para as cinco regiões brasileiras, se mostraram satisfatórias. A Figura 5a mostra as previsões para a Região Centro-Oeste. O intervalo de predição comportou as observações reais na maioria dos casos e o modelo foi capaz de captar os períodos de picos na série. A Figura 5b apresenta as previsões para a Região Nordeste. As observações permaneceram dentro do intervalo de predição e o modelo detectou os períodos de variação da série. Na Figura 5c, as observações da Região Norte foram contempladas pelo intervalo de predição, no entanto, o modelo não detectou totalmente as variações dentro dos períodos de pico. A Figura 5d mostra que o intervalo de predição para a Região Sudeste comportou a maioria das observações e o modelo foi capaz de identificar os períodos de pico da série. A Figura 5e apresenta as previsões para a Região Sul. As observações ficaram, em sua maioria, dentro do intervalo de predição e as estimativas se aproximaram dos dados reais em muitos casos.

Para os hotéis do Tipo 2, as previsões, para as cinco regiões brasileiras, foram satisfatórias. A Figura 6a mostra as previsões para a Região Centro-Oeste. O intervalo de predição comportou as observações reais na maioria dos casos, com exceção de três picos destoantes do resto do período, e as previsões se mostraram próximas às observações. A Figura 6b apresenta as previsões para a Região Nordeste. O modelo captou a forma da série. Na Figura 6c, as observações da Região Norte foram contempladas pelo intervalo de predição, e o modelo gerou estimativas próximas aos valores observados em grande parte dos casos. A Figura 6d mostra que o intervalo de predição para a Região Sudeste comportou a maioria das observações e o modelo gerou estimativas próximas ao observado. A Figura 6e apresenta as previsões para a Região Sul. O intervalo de predição contemplou boa parte das observações e as estimativas se aproximaram dos dados reais em muitos casos.

A Figura 5a mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo SARIMA para a Região Centro-Oeste. A Figura 5b mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo SARIMA para a Região Nordeste. A Figura 5c mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo SARIMA para a Região Norte. A Figura 5d mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo SARIMA para a Região Sudeste. A Figura 5e mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo SARIMA para a Região Sul.

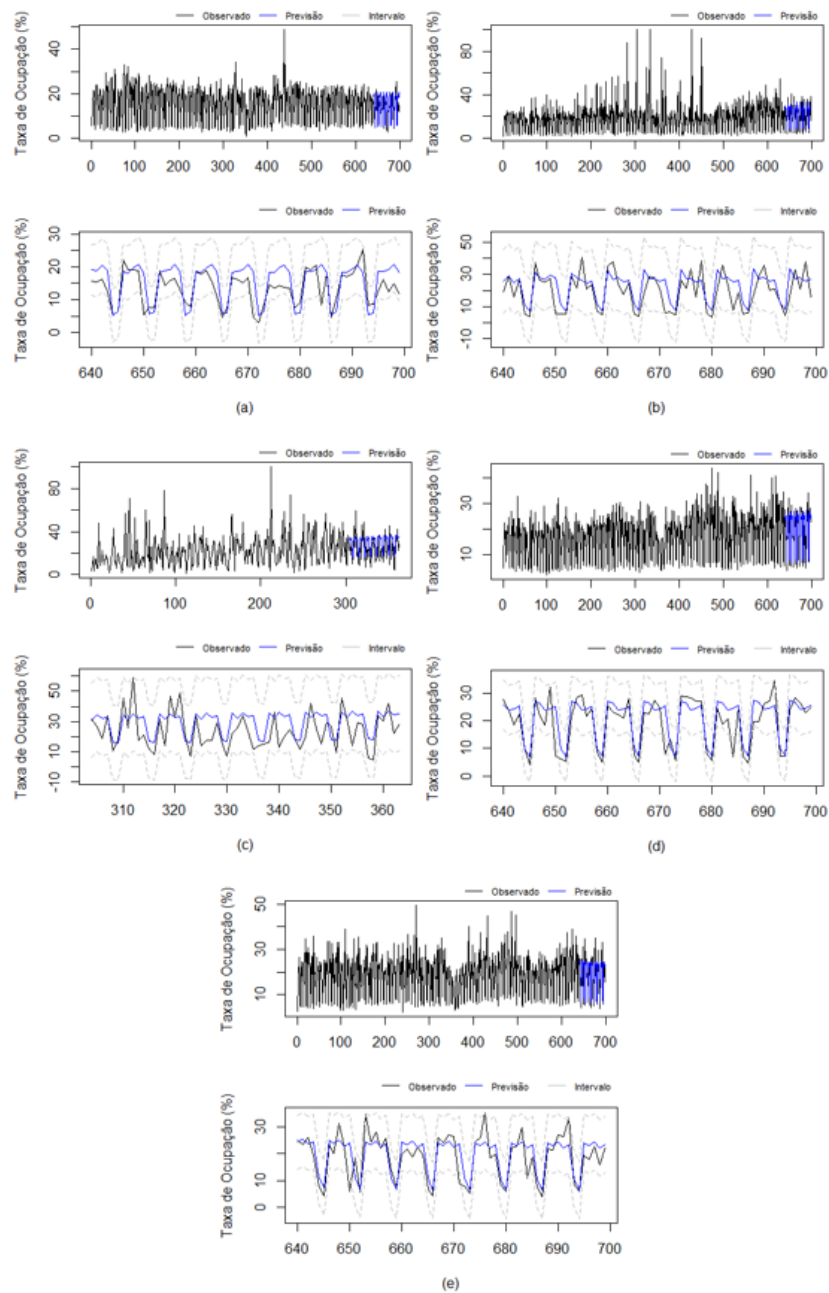


Figura 4 – Previsão da Taxa de Ocupação (%) pelo modelo SARIMA, 60 previsões um passo à frente, para hotéis do Tipo 1. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.

A Figura 6a mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo SARIMA para a Região Centro-Oeste. A Figura 6b mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo SARIMA para a Região Nordeste. A Figura 6c mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo SARIMA para a Região Norte. A Figura 6d mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo SARIMA para a Região Sudeste. A Figura 6e mostra a previsão da taxa de ocupação dos hotéis

do tipo 2, 60 previsões um passo à frente, com o modelo SARIMA para a Região Sul.

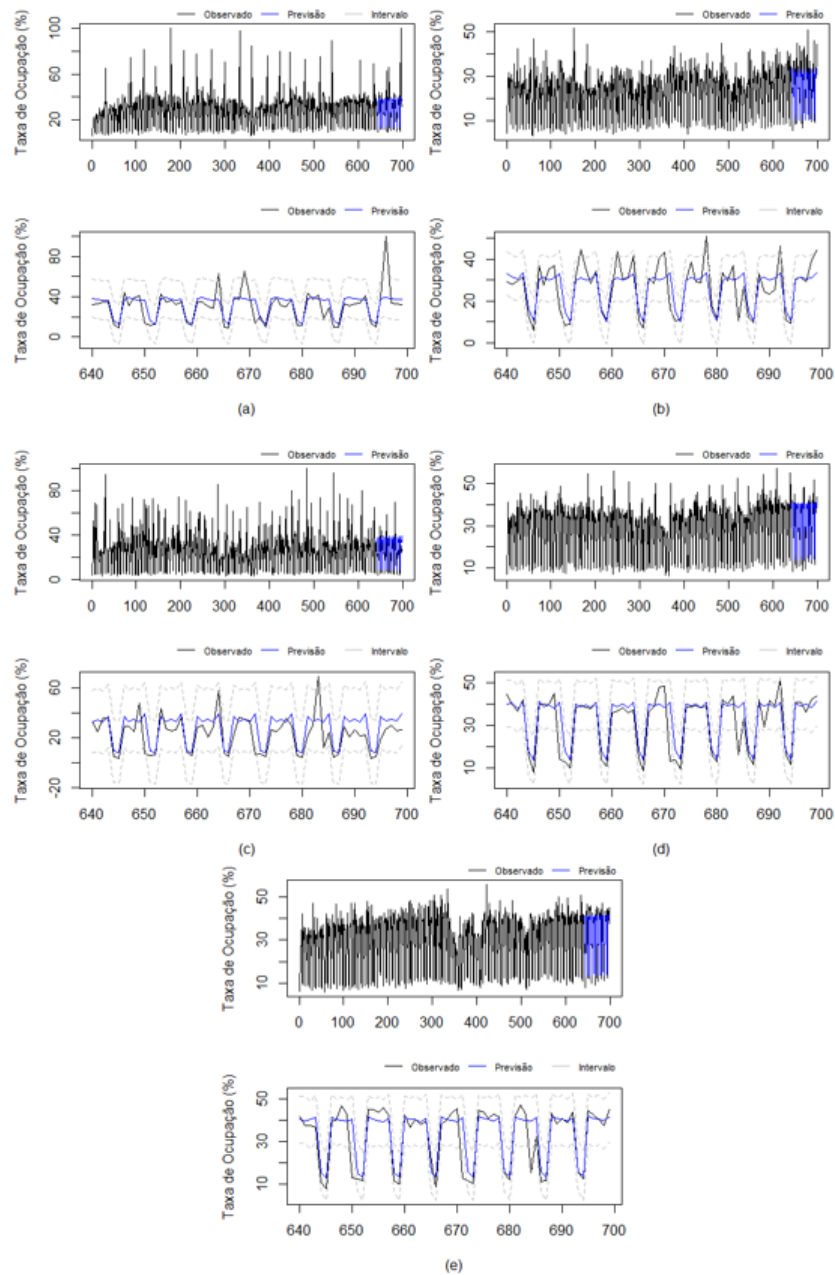


Figura 5 – Previsão da Taxa de Ocupação (%) pelo modelo SARIMA, 60 previsões um passo à frente, para hotéis do Tipo 2. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.

5.2 Resultados dos Modelos de Redes Neurais

Para os hotéis do tipo 1 da Região Norte, modelamos um NNAR (21,11), ou seja, uma rede neural com as últimas 21 observações usadas como insumo para prever a saída y e com 11 neurônios na camada oculta. Para o restante dos grupos das demais regiões (Centro-Oeste, Nordeste, Sudeste e Sul) e os hotéis do tipo 2 da Região Norte, modelamos um NNAR (28,14). As taxas de ocupação (%) foram

realizadas 60 previsões um passo à frente para os hotéis dos dois grupos, das cinco regiões brasileiras, seguindo os modelos selecionados.

As previsões para os hotéis do Tipo 1, para as cinco regiões brasileiras, se mostraram razoáveis, em geral. A Figura 7a mostra as previsões para a Região Centro-Oeste. A maior parte das observações ficaram no limite do intervalo de predição e o modelo foi capaz de captar os períodos variação da série. A Figura 7b apresenta as previsões para a Região Nordeste. As observações permaneceram dentro do intervalo de predição, em sua maioria e as estimativas, em geral, se aproximaram do observado. Na Figura 7c, as observações da Região Norte não foram contempladas pelo intervalo de predição em alguns casos, o que é um comportamento esperado. A Figura 7d mostra que o intervalo de predição para a Região Sudeste comportou a maioria das observações e o modelo gerou estimativas razoáveis. A Figura 7e apresenta as previsões para a Região Sul. As estimativas, em geral, ficaram bastante próximas do observado.

Para os hotéis do Tipo 2, as previsões, para as cinco regiões brasileiras, foram, em geral, satisfatórias. A Figura 8a mostra as previsões para a Região Centro-Oeste. As predições se mostraram próximas às observações em muitos casos. A Figura 8b apresenta as previsões para a Região Nordeste. O modelo captou a forma da série, porém apresentou estimativas, em geral, mais baixas que os valores observados. Na Figura 8c, as observações da Região Norte foram contempladas pelo intervalo de predição na maior parte dos casos, e o modelo gerou estimativas próximas aos valores observados. A Figura 8d mostra que o intervalo de predição para a Região Sudeste comportou a maioria das observações e o modelo gerou estimativas próximas ao observado. A Figura 8e apresenta as previsões para a Região Sul. O intervalo de predição contemplou boa parte das observações e as estimativas se aproximaram dos dados reais em muitos casos.

A Figura 7a mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Centro-Oeste. A Figura 7b mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Nordeste. A Figura 7c mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Norte. A Figura 7d mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Sudeste. A Figura 7e mostra a previsão da taxa de ocupação dos hotéis do tipo 1, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Sul.

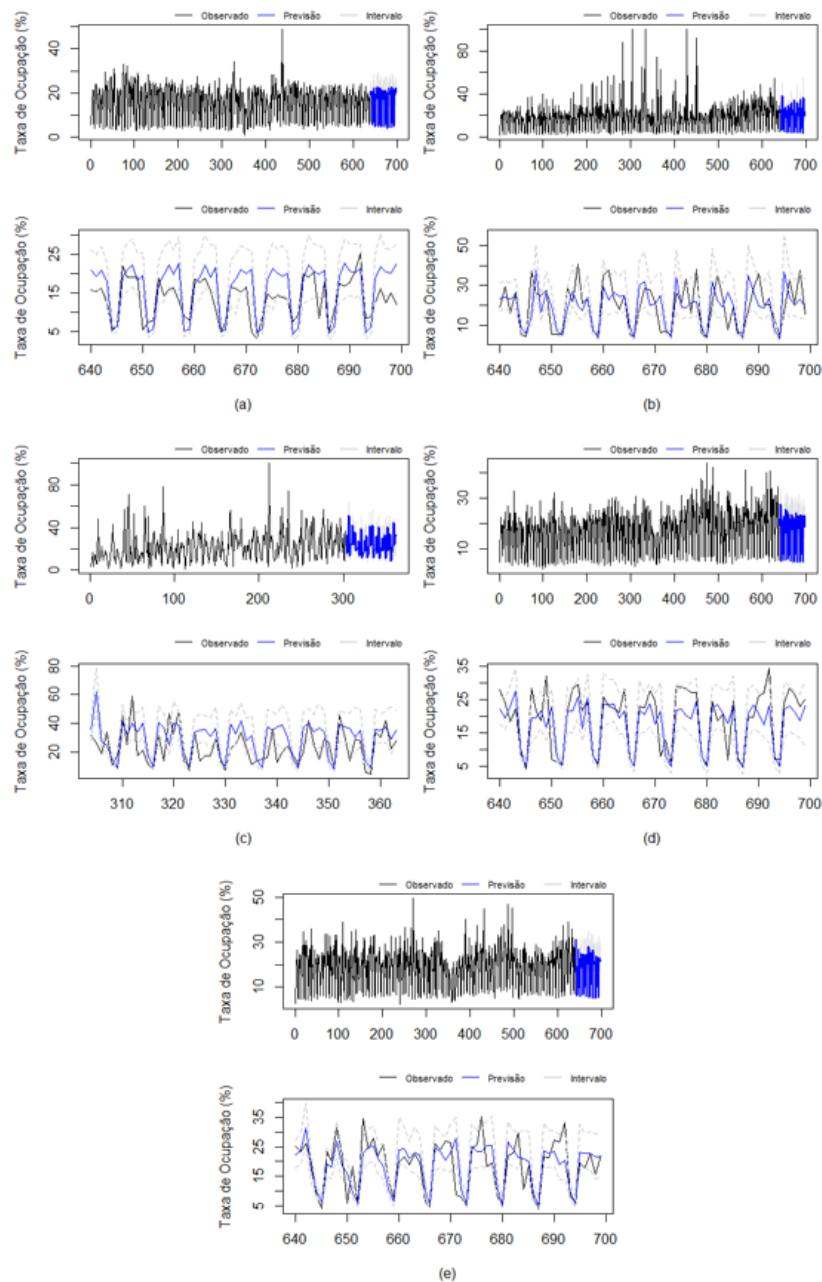


Figura 6 – Previsão da Taxa de Ocupação (%) pelo modelo de Redes Neurais, 60 previsões um passo à frente, para hotéis do Tipo 1. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.

A Figura 8a mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Centro-Oeste. A Figura 8b mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Nordeste. A Figura 8c mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Norte. A Figura 8d mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente, com o modelo de Redes Neurais para a Região Sudeste. A Figura 8e mostra a previsão da taxa de ocupação dos hotéis do tipo 2, 60 previsões um passo à frente,

com o modelo de Redes Neurais para a Região Sul.

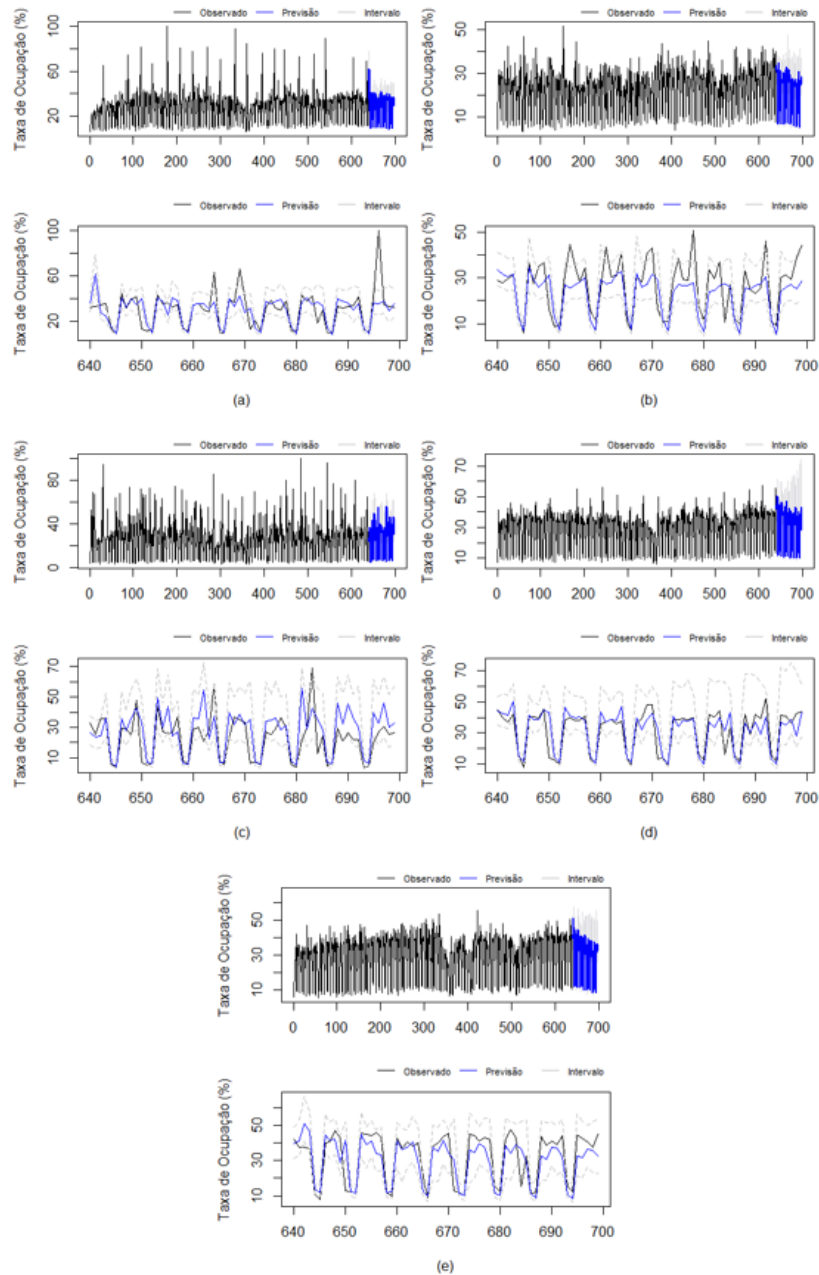


Figura 7 – Previsão da Taxa de Ocupação (%) pelo modelo de Redes Neurais, 60 previsões um passo à frente, para hotéis do Tipo 2. (a) Região Centro-Oeste. (b) Região Nordeste. (c) Região Norte. (d) Região Sudeste. (e) Região Sul.

5.3 Comparação das previsões dos modelos propostos

Realizamos o teste de Diebold-Mariano para a comparação das previsões dos modelos. Como mencionado no tópico 4.3, a hipótese nula do teste é de que os dois modelos possuem a mesma capacidade de previsão. Quando a hipótese nula é rejeitada, um dos modelos deve ter melhor capacidade de previsão que o outro. Foram feitas 60 previsões 1 passo à frente, 54 previsões 7 passos à frente e 30 previsões 30

passos à frente para aplicação do teste.

Para a região Centro-Oeste, para os hotéis do tipo 1, nas comparações 1 e 7 passos à frente, rejeitamos a hipótese nula de que os dois modelos têm a mesma capacidade de previsão. Olhando a raiz do erro quadrado médio, observamos que o modelo mais preciso nos dois casos é o SARIMA, pois tem os menores valores de RMSE, 4.576 e 4.374, respectivamente. Para os hotéis do tipo 2, não rejeitamos a hipótese nula para nenhuma das comparações.

A Tabela 11 apresenta os p-valores obtidos pelo teste de Diebold-Mariano para a comparação entre os modelos SARIMA e Redes Neurais.

Tabela 11 – P-valor do teste de Diebold-Mariano para a Região Centro-Oeste

	Tipo 1	Tipo 2
1 passo à frente	0.0474	0.3244
7 passos à frente	4.889e-07	0.9476
30 passos à frente	0.1699	0.06055

Na Tabela 12 são apresentadas as raizes dos erros quadrados médios dos modelos SARIMA e Redes Neurais para os dois grupos de hotéis.

Tabela 12 – Raiz do erro quadrado médio para a Região Centro-Oeste

	Tipo 1		Tipo 2	
	SARIMA	RNs	SARIMA	RNs
1 passo à frente	4.576	5.04	11.345	12.013
7 passos à frente	4.374	5.066	11.843	13.305
30 passos à frente	7.573	8.668	18.029	25.62

Para a região Nordeste, para os hotéis do tipo 1, para as comparações 7 e 30 passos à frente, rejeitamos a hipótese nula de que os dois modelos têm a mesma capacidade de previsão. Olhando a raiz do erro quadrado médio, observamos que o modelo mais preciso nestes casos é o SARIMA, pois tem os menores valores de RMSE, 7.881 e 11.135. Para os hotéis do tipo 2, rejeitamos a hipótese nula para as comparações 1 e 7 passos à frente. Novamente, observando os valores de RMSE, 7.021 e 7.366, respectivamente, concluímos que o modelo SARIMA tem melhor capacidade de previsão nos dois casos. Para a comparação 30 passos à frente não rejeitamos a hipótese nula.

A Tabela 13 apresenta os p-valores obtidos pelo teste de Diebold-Mariano para a comparação entre os modelos SARIMA e Redes Neurais.

Tabela 13 – P-valor do teste de Diebold-Mariano para a Região Nordeste

	Tipo 1	Tipo 2
1 passo à frente	0.1834	0.005646
7 passos à frente	0.0007398	0.006797
30 passos à frente	0.01025	0.1912

Na Tabela 14 são apresentadas as raízes dos erros quadrados médios dos modelos SARIMA e Redes Neurais para os dois grupos de hotéis.

Tabela 14 – Raiz do erro quadrado médio para a Região Nordeste

	Tipo 1		Tipo 2	
	SARIMA	RNs	SARIMA	RNs
1 passo à frente	7.562	8.433	7.021	8.461
7 passos à frente	7.881	9.139	7.366	8.738
30 passos à frente	11.135	17.164	13.503	16.169

Para a região Norte, para os hotéis do tipo 1, rejeitamos a hipótese nula somente para a comparação 1 passo à frente. Olhando a raiz do erro quadrado médio, observamos que o modelo mais preciso é o SARIMA, pois tem o menor valor de RMSE, 11.238. Para os hotéis do tipo 2, rejeitamos a hipótese nula somente para as comparações 1 e 7 passos à frente. Novamente, observando os valores de RMSE, 11.228 e 11.595, concluímos que o modelo SARIMA tem melhor capacidade de previsão nos dois casos.

A Tabela 15 apresenta os p-valores obtidos pelo teste de Diebold-Mariano para a comparação entre os modelos SARIMA e Redes Neurais.

Tabela 15 – P-valor do teste de Diebold-Mariano para a Região Norte

	Tipo 1	Tipo 2
1 passo à frente	0.01274	0.0009141
7 passos à frente	0.6072	0.003129
30 passos à frente	1	1

Na Tabela 16 são apresentadas as raízes dos erros quadrados médios dos modelos SARIMA e Redes Neurais para os dois grupos de hotéis.

Tabela 16 – Raiz do erro quadrado médio para a Região Norte

	Tipo 1		Tipo 2	
	SARIMA	RNs	SARIMA	RNs
1 passo à frente	11.238	14.871	11.228	11.425
7 passos à frente	11.486	13.683	11.595	12.857
30 passos à frente	13.637	18.427	17.295	23.981

Para a região Sudeste, para ambos os tipos de hotéis, não rejeitamos a hipótese nula de que os dois modelos têm a mesma capacidade de previsão em nenhum dos casos.

A Tabela 17 apresenta os p-valores obtidos pelo teste de Diebold-Mariano para a comparação entre os modelos SARIMA e Redes Neurais.

Tabela 17 – P-valor do teste de Diebold-Mariano para a Região Sudeste

	Tipo 1	Tipo 2
1 passo à frente	0.1808	0.05565
7 passos à frente	0.6168	0.8
30 passos à frente	1	1

Na Tabela 18 são apresentadas as raízes dos erros quadrados médios dos modelos SARIMA e Redes Neurais para os dois grupos de hotéis.

Tabela 18 – Raiz do erro quadrado médio para a Região Sudeste

	Tipo 1		Tipo 2	
	SARIMA	RNs	SARIMA	RNs
1 passo à frente	5.256	6.09	6.912	7.97
7 passos à frente	5.398	5.776	7.193	7.359
30 passos à frente	10.237	13.653	15.743	20.936

Para a região Sul, para os hotéis do tipo 1, não rejeitamos a hipótese nula de que os dois modelos têm a mesma capacidade de previsão em nenhum dos casos. Para os hotéis do tipo 2, rejeitamos a hipótese nula para comparação 1 passo à frente. Observando os valores de RMSE, concluímos que o modelo SARIMA tem melhor capacidade de previsão.

A Tabela 19 apresenta os p-valores obtidos pelo teste de Diebold-Mariano para a comparação entre os modelos SARIMA e Redes Neurais.

Tabela 19 – P-valor do teste de Diebold-Mariano para a Região Sul

	Tipo 1	Tipo 2
1 passo à frente	0.4778	0.01549
7 passos à frente	0.7126	0.07383
30 passos à frente	1	0.08324

Na Tabela 20 são apresentadas as raízes dos erros quadrados médios dos modelos SARIMA e Redes Neurais para os dois grupos de hotéis.

Tabela 20 – Raiz do erro quadrado médio para a Região Sul

	Tipo 1		Tipo 2	
	SARIMA	RNs	SARIMA	RNs
1 passo à frente	5.08	5.336	6.892	9.673
7 passos à frente	5.256	5.3	7.037	10.754
30 passos à frente	9.918	14.845	16.744	22.035

6 CONCLUSÃO

Esse trabalho teve como objetivo avaliar a capacidade preditiva de dois modelos consagrados, um da área de séries temporais, outro de machine learning. Os modelos escolhidos foram SARIMA e Redes Neurais. As séries de cada uma das cinco regiões: Centro-Oeste, Nordeste, Norte, Sudeste e Sul foram divididas em dois grupos: 1 e 2 por solicitação da empresa provedora dos dados analisados.

Utilizando-se o teste de comparação de previsão Diebold-Mariano, as conclusões não foram homogêneas para todos os tipos de hotéis das referidas regiões. Em alguns casos não foi possível afirmar a existência de diferença na capacidade de previsão entre os dois modelos testados. No entanto, em todos os casos em que esta diferença foi detectada, o modelo SARIMA apresentou menores valores de RMSE e foi, portanto, indicado como melhor alternativa.

Para trabalhos futuros, uma sugestão seria incorporar covariáveis nos modelos testados. Vale ressaltar que, no caso da hotelaria, além de variáveis macroeconômicas relacionadas à região em questão, informações com relação à origem dos hóspedes podem oferecer bons ganhos. Outra sugestão é testar outros modelos como, por exemplo, o modelo beta autorregressivo de médias móveis (β ARMA). O modelo em questão assume que a variável possui distribuição beta, que tem suporte restrito ao intervalo (0;1), logo, é adequado para prever taxas.

REFERÊNCIAS

- BONINI NETO, A.; BONINI, C. dos S. B. Redes Neurais Artificiais: Apresentação e utilização do algoritmo Perceptron em Biossistemas. **Revista Brasileira de Engenharia de Biossistemas**, v. 4, n. 2, p. 87 – 95, 12 2010.
- CAICEDO, W.; PAYARES, F. Machine Learning Model for Occupancy Rates and Demand Forecasting in the Hospitality Industry. In: PROCEEDINGS, 2016, San José. **IBERAMIA 2016**: 15th ibero-american conference on ai. San José, 2016. p. 201 – 211.
- CHOW, W. S.; SHYU, J.; WANG, K. Developing a Forecast System for Hotel Occupancy Rate Using Integrated ARIMA Models. **Journal of International Hospitality, Leisure & Tourism Management**, v. 1, n. 3, p. 55 – 80, 26 ago. 1998.
- CORIOLOANO, L. N. M. T.; ALMEIDA, H. M. de. O turismo no nordeste brasileiro: dos resorts aos núcleos de economia solidária. **Revista Electrónica de Geografía e y Ciencias Sociales**, v. 11, n. 245, Ago 2007. ISSN 1138-9788. Disponível em: <http://www.ub.edu/geocrit/sn/sn-24557.htm>. Acesso em: 22 jun. 2019.
- FAUSETT, L. V. **Fundamentals of Neural Networks**: Architectures, Algorithms And Applications. Londres: Pearson, 1994.
- FERNEDA, E. **Redes neurais e sua aplicação em sistemas de recuperação de informação**. 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000100003&lng=en&nrm=iso>. Acesso em: 28/10/2019.
- GORINI, A. P. F.; MENDES, E. da F. Setor de turismo no Brasil: segmento de hotelaria. **BNDES Setorial**, Rio de Janeiro, [S.l.], n. 22, p. 111 – 150, 09 2005. Disponível em: https://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/bnset/set2205.pdf. Acesso em: 06 Jun. 2019.
- HYNDMAN, R. **Forecasting Functions for Time Series and Linear Models**. Package forecast R, 2016.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting**: Principles and Practice. 2. ed. .: OTexts, 2018.
- INFOBH. Fórum de Operadores Hoteleiros do Brasil, São Paulo, n. 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 140, 2018.
- IRVING, M. de A. Áreas protegidas de fronteira e turismo sustentável na Amazônia: entre o surrealismo e a invenção. **Revista de Desenvolvimento Econômico**, v. 8, n. 13, p. 35 – 49, Jan 2006.
- MANTOVANI, G. F. **Previsão de séries temporais : redes neurais artificiais vs. modelos ARIMA**. 2004. 62 p. Monografia (Estatística) — UFRGS.
- MATTOS, A. C. de. **Diretrizes para o dimensionamento do número de unidades habitacionais de hotéis resort**. 2004. 108 p. Dissertação (Engenharia) — USP.
- OLIVEIRA, A. B. Um estudo comparativo de Redes Neurais e Modelos GARCH para Previsão da Volatilidade de Séries Temporais Financeiras. 2010.

PAES, M. T. D.; BRUHNS, H. T.; SERRANO, C. (org.). **Patrimônio, natureza e cultura**. [S.l.]: Papirus Editora, 2007.

PAIVA, O. da C. Imigração, patrimônio cultural e turismo no Brasil. **Anais do Museu Paulista**, São Paulo, v. 23, n. 2, p. 211 – 237, Dez. 2015.

RAUBER, T. W. **Redes neurais artificiais**. Vitória: Universidade Federal do Espírito Santo, 2005.

RESENDE, G. M. et al. Fatos Recentes do Desenvolvimento Regional no Brasil. **Texto para Discussão**, IPEA, Brasília, n. 2054, Mar 2015. ISSN 1415-4765.

SAS INSTITUTE INC. **Residual Normal QQ Plot**. Cary, NC, 2004.

SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications: With R Examples**. 3. ed. [S.l.]: Springer, 2011. ISBN 978-1441978646.

WADA, E. K.; CAMARGO, L. O. de L. Os desafios da hotelaria. **GV Executivo**, v. 5, n. 1, p. 53 – 57, Fev./Abr. 2006.

WEI, W. W. **Time Series Analysis: Univariate and Multivariate Methods**. 2. ed. [S.l.]: Pearson Education, 2006.

WTTC. **Travel & Tourism Economic Impact 2018 Brazil**. [S.l.], 2018.

ȚÎȚU, M. A.; RĂULEA, A. S.; ȚÎȚU Ștefan. Measuring Service Quality in Tourism Industry. **Procedia - Social And Behavioral Sciences**, v. 221, n. [s.l.], p. 294 – 301, 6 2016. Disponível em: <http://dx.doi.org/10.1016/j.sbspro.2016.05.118>. Acesso em: 06 Jun. 2019.

APÊNDICE A

Conceitos Básicos de Séries Temporais

Shumway e Stoffer (2011) definem uma série temporal como uma coleção de variáveis aleatórias indexadas de acordo com a ordem em que são obtidas no tempo.

Para atingirmos um entendimento apropriado de séries temporais e estimação, é importante definirmos alguns conceitos básicos.

Processos Estocásticos

Para Wei (2006), um processo estocástico é uma família de variáveis $Z(\omega, t)$ indexadas no tempo, onde ω pertence a um espaço de amostra e t pertence a um conjunto de índices. Para um t fixo, $Z(\omega, t)$ é uma variável aleatória. Para um dado ω , $Z(\omega, t)$, como uma função de t , é chamada de função de amostra ou realização. Assim, uma série temporal é uma função de realização ou amostra de um determinado processo estocástico.

Ainda segundo Wei (2006), considere um conjunto finito de variáveis aleatórias $\{Z_{t_1}, Z_{t_2}, \dots, Z_{t_n} \leq x_n\}$ de um processo estocástico $\{Z(\omega, t) : t = 0, \pm 1, \pm 2, \dots\}$. A função de distribuição é definida por:

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = P\{\omega : Z_{t_1} \leq x_1, \dots, Z_{t_n} \leq x_n\}$$

onde $x_i, i = 1, \dots, n$ são quaisquer números reais.

Um processo é dito ser estacionário de primeira ordem se sua função de distribuição unidimensional for invariante no tempo, ou seja, se $F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1)$ para qualquer inteiro t_1, k , e $t_1 + k$; estacionário de segunda ordem se $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ para qualquer inteiro $t_1, t_2, k, t_1 + k$ e $t_2 + k$; e estacionário de enésima ordem se

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = F_{Z_{t_1+k}, \dots, Z_{t_n+k}}(x_1, \dots, x_n)$$

para qualquer t_1, \dots, t_n e qualquer inteiro k .

Ruído Branco

Um processo x_t é chamado de ruído branco se for uma sequência de variáveis aleatórias não correlacionadas de uma distribuição fixa com:

- i) $E[x_t] = \mu, \forall t$
- ii) $Var[x_t] = \sigma^2, \forall t$
- iii) $Cov(x_t, x_{t+k}) = 0, \forall t \text{ e } \forall k \neq 0$

Função de Autocorrelação (ACF)

A função de autocorrelação mede a previsibilidade linear de x_t utilizando apenas o valor de x_s e é definida por:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

onde $-1 \leq \rho(s, t) \leq 1$.

Teste de Dickey-Fuller

Testamos a estacionariedade da série através do teste da raiz unitária de Dickey-Fuller (ver mais detalhes em Shumway e Stoffer (2011), página 252). Nos casos em que a suposição de estacionariedade foi rejeitada, obtemos a primeira diferença das séries, como sugerido em Shumway e Stoffer (2011), por exemplo. Após, checamos novamente esta mesma suposição através do teste de raiz unitária de Dickey-Fuller.

Teste de Ljung-Box

Segundo Hyndman e Athanasopoulos (2018), o teste de Ljung-Box foi proposto por Ljung and Box e é baseado em

$$Q^* = T(T + 2) \sum_{k=1}^h (T - k)^{-1} r_k^2$$

onde T é o tamanho da série temporal, r_k é o k th coeficiente de autocorrelação dos resíduos, e h é o número de defasagens a serem testadas. Grandes valores de Q^* indicam que existem autocorrelações significativas nas séries de resíduos.

QQ-Plot

Em um gráfico QQ-Plot, os quantis empíricos são plotados contra os quantis de uma distribuição normal padrão. Se os resíduos forem de uma distribuição normal com média 0, os pontos tendem a cair ao longo da linha de referência que possui um intercepto de 0 e uma inclinação igual ao desvio padrão estimado (SAS INSTITUTE INC., 2004).

APÊNDICE B

Nesta seção apresentamos o detalhamento da escolha dos modelos SARIMA para as regiões Centro-Oeste, Norte, Sudeste e Sul.

Região Centro-Oeste

Como método de avaliação inicial das séries de taxa de ocupação hoteleira da Região Centro-Oeste mostramos, na Figura 8 as taxas de ocupação dos dois grupos de hotéis estudados: 1 e 2, durante 639 dias correntes.

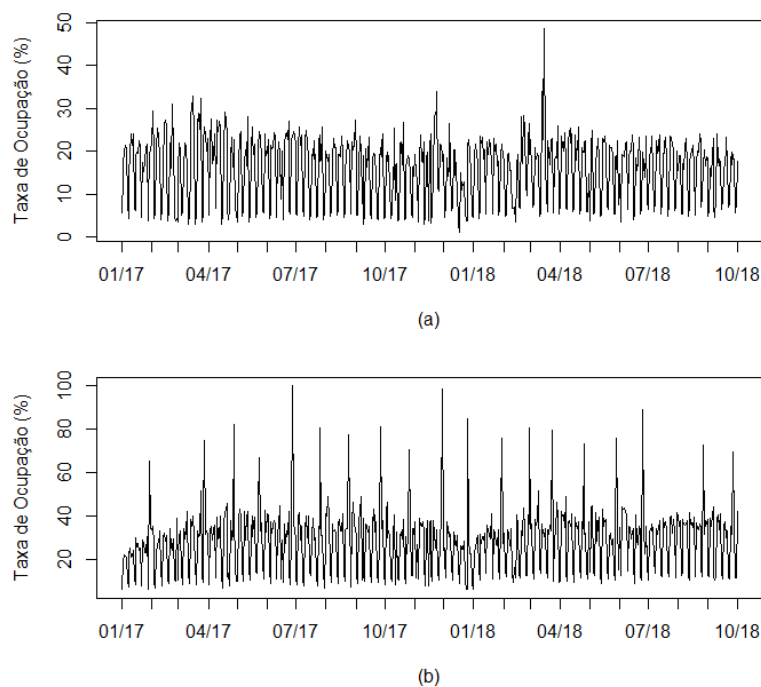


Figura 8 – Séries temporais da taxa de ocupação (%) da Região Centro-Oeste
(a) Tipo 1. (b) Tipo 2

Realizamos o teste da raiz unitária de Dickey-Fuller. Para ambas as séries obtivemos um p-valor < 0.01 e rejeitamos a hipótese de presença de raiz unitária, assumindo as séries das taxas de ocupação dos dois tipos de hotéis como estacionárias.

No Apêndice C, a Figura 12 mostra as funções de autocorrelação e autocorrelação parcial da série (a), enquanto a Figura 13 mostra as funções de autocorrelação e autocorrelação parcial da série (b). Atendidas as suposições de estacionariedade, diversos modelos ARIMA foram testados para ambas as séries.

Para os hotéis do tipo 1, a Tabela 21 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(3,1,3) \times (1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 21 – AIC dos modelos SARIMA sugeridos, Centro-Oeste, hotéis tipo 1

Modelo Sugerido	AIC
SARIMA(3,1,3)x(3,1,3) ₇	3569.46
SARIMA(3,1,3)x(1,1,1)₇	3567.91
SARIMA(3,1,3)x(2,1,2) ₇	3570.35

Para os hotéis do tipo 2, a Tabela 22 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(5,1,1)x(1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 22 – AIC dos modelos SARIMA sugeridos, Centro-Oeste, hotéis tipo 2

Modelo Sugerido	AIC
SARIMA(5,1,1)x(4,1,1) ₇	4721.12
SARIMA(5,1,1)x(1,1,1)₇	4719.38
SARIMA(5,1,1)x(1,2,1) ₇	4953.08

Região Norte

Como método de avaliação inicial das séries de taxa de ocupação hoteleira da Região Nordeste mostramos, na Figura 9, as taxas de ocupação dos dois grupos de hotéis estudados: 1 e 2. Hotéis do tipo 1 foram observados durante 303 dias, enquanto hotéis do tipo 2 possuem 639 observações.

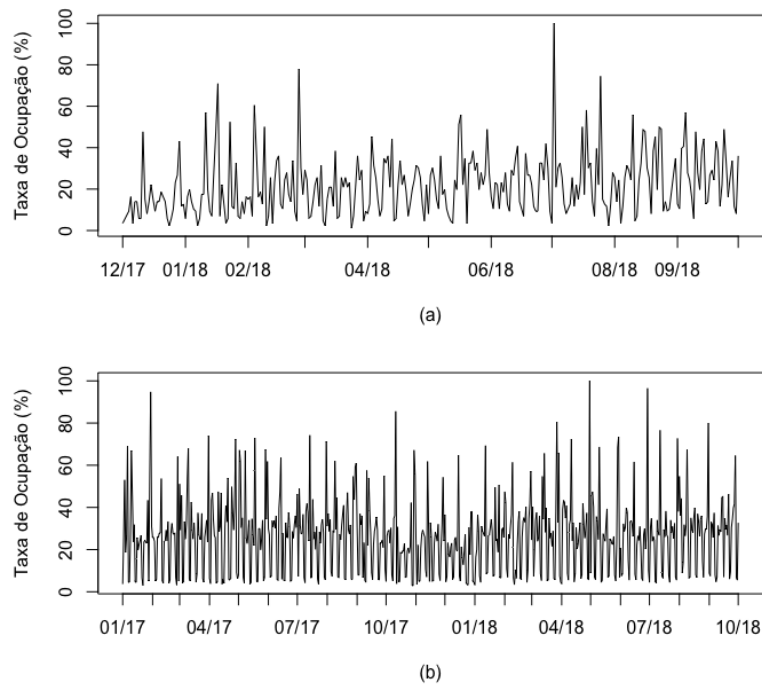


Figura 9 – Séries temporais da taxa de ocupação (%) da Região Norte
(a)Tipo 1. (b) Tipo 2

Realizamos o teste da raiz unitária de Dickey-Fuller. Para ambas as séries obtivemos um p-valor < 0.01 e rejeitamos a hipótese de presença de raiz unitária, assumindo as séries das taxas de ocupação dos dois tipos de hotéis como estacionárias.

No Apêndice C, a Figura 16 mostra as funções de autocorrelação e autocorrelação parcial da série (a), enquanto a Figura 17 mostra as funções de autocorrelação e autocorrelação parcial da série (b). Atendidas as suposições de estacionariedade, diversos modelos ARIMA foram testados para ambas as séries.

Para os hotéis do tipo 1, a Tabela 23 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo $SARIMA(1,1,1) \times (0,1,1)_7$ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 23 – AIC dos modelos SARIMA sugeridos, Norte, hotéis tipo 1

Modelo Sugerido	AIC
$SARIMA(1,1,1) \times (1,1,1)_7$	2364.37
$SARIMA(1,1,1) \times (0,1,1)_7$	2362.42
$SARIMA(1,1,1) \times (1,0,1)_7$	2409.06

Para os hotéis do tipo 2, a Tabela 24 mostra o valor do AIC dos modelos sugeridos.

dos. Percebemos que o modelo SARIMA(2,1,1)x(1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 24 – AIC dos modelos SARIMA sugeridos, Norte, hotéis tipo 2

Modelo Sugerido	AIC
SARIMA(2,1,1)x(1,1,1)₇	5056.37
SARIMA(2,0,1)x(1,0,1) ₇	5112.24
SARIMA(2,1,1)x(2,1,2) ₇	5060.36

REGIÃO SUDESTE

Como método de avaliação inicial das séries de taxa de ocupação hoteleira da Região Nordeste mostramos, na Figura 10, as taxas de ocupação dos dois grupos de hotéis estudados: 1 e 2, durante 639 dias correntes.

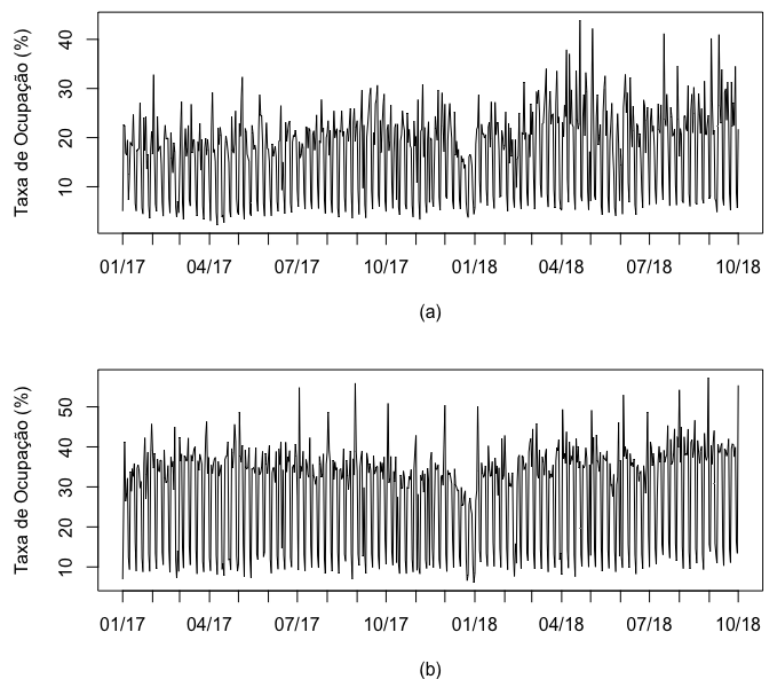


Figura 10 – Séries temporais da taxa de ocupação (%) da Região Sudeste
(a) Tipo 1. (b) Tipo 2

Realizamos o teste da raiz unitária de Dickey-Fuller. Para ambas as séries obtivemos um p-valor < 0.01 e rejeitamos a hipótese de presença de raiz unitária, assumindo as séries das taxas de ocupação dos dois tipos de hotéis como estacionárias.

No Apêndice C, a Figura 18 mostra as funções de autocorrelação e autocorrelação parcial da série (a), enquanto a Figura 19 mostra as funções de autocorrelação e autocorrelação parcial da série (b). Atendidas as suposições de estacionariedade, diversos modelos ARIMA foram testados para ambas as séries.

Para os hotéis do tipo 1, a Tabela 25 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(2,1,3)x(0,1,2)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 25 – AIC dos modelos SARIMA sugeridos, Sudeste, hotéis tipo 1

Modelo Sugerido	AIC
SARIMA(2,1,3)x(2,1,3) ₇	3733.86
SARIMA(2,1,3)x(0,1,2)₇	3732.42
SARIMA(2,1,3)x(1,1,2) ₇	3734.3

Para os hotéis do tipo 2, a Tabela 26 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(2,1,2)x(1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 26 – AIC dos modelos SARIMA sugeridos, Sudeste, hotéis tipo 2

Modelo Sugerido	AIC
SARIMA(2,1,2)x(0,1,1) ₇	4011.75
SARIMA(2,1,2)x(1,1,0) ₇	4221.88
SARIMA(2,1,2)x(1,1,1)₇	4010.65

REGIÃO SUL

Como método de avaliação inicial das séries de taxa de ocupação hoteleira da Região Nordeste mostramos, na Figura 11, as taxas de ocupação dos dois grupos de hotéis estudados: 1 e 2, durante 639 dias correntes.

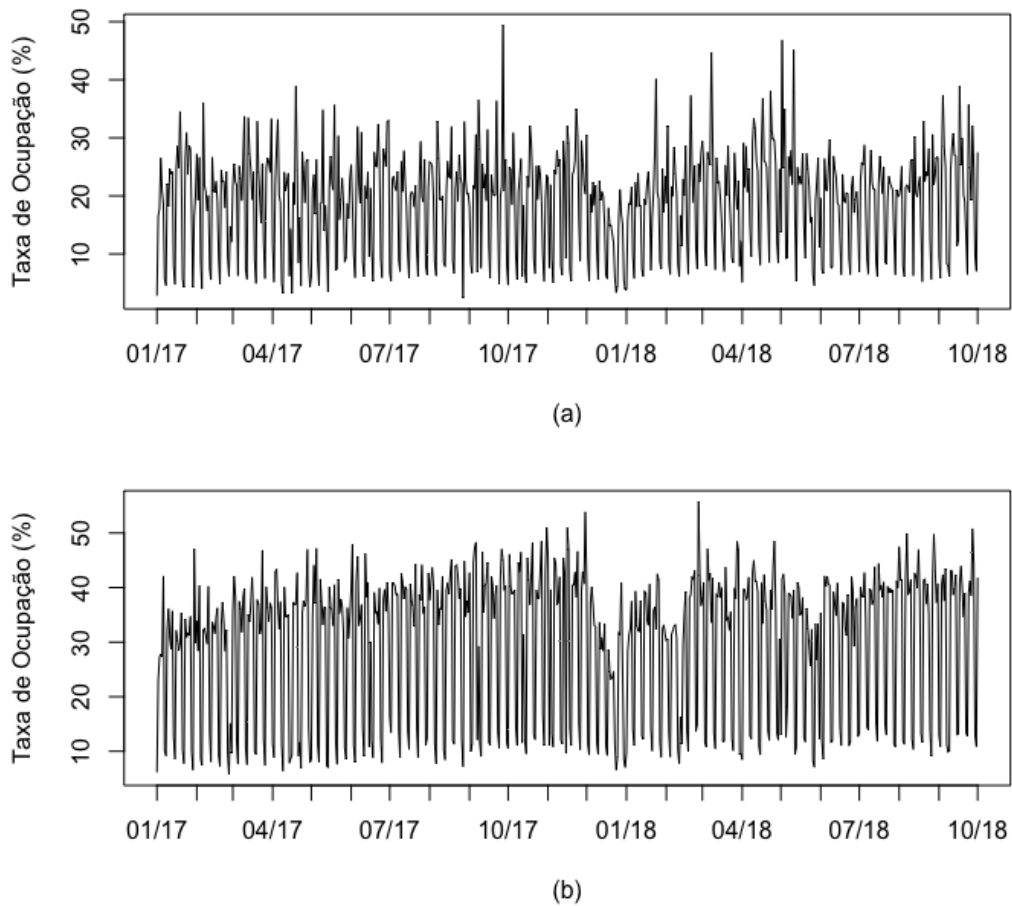


Figura 11 – Séries temporais da taxa de ocupação (%) da Região Sul
(a) Tipo 1. (b) Tipo 2

Realizamos o teste da raiz unitária de Dickey-Fuller. Para ambas as séries obtivemos um p-valor < 0.01 e rejeitamos a hipótese de presença de raiz unitária, assumindo as séries das taxas de ocupação dos dois tipos de hotéis como estacionárias.

No Apêndice C, a Figura 20 mostra as funções de autocorrelação e autocorrelação parcial da série (a), enquanto a Figura 21 mostra as funções de autocorrelação e autocorrelação parcial da série (b). Atendidas as suposições de estacionariedade, diversos modelos ARIMA foram testados para ambas as séries.

Para os hotéis do tipo 1, a Tabela 27 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo $SARIMA(5,0,1) \times (1,1,1)_7$ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 27 – AIC dos modelos SARIMA sugeridos, Sul, hotéis tipo 1

Modelo Sugerido	AIC
SARIMA(5,0,1)x(1,0,1) ₇	3954.01
SARIMA(5,0,1)x(2,0,2) ₇	3956.42
SARIMA(5,0,1)x(1,1,1)₇	3901.21

Para os hotéis do tipo 2, a Tabela 28 mostra o valor do AIC dos modelos sugeridos. Percebemos que o modelo SARIMA(3,1,3)x(1,1,1)₇ possui o menor valor de AIC, justificando a sua escolha como modelo final. Nota-se que, neste caso, foi detectada uma sazonalidade semanal nos dados. Verificadas as suposições, no Apêndice D estão os gráficos de autocorrelação e QQ-Plot dos resíduos do modelo.

Tabela 28 – AIC dos modelos SARIMA sugeridos, Sul, hotéis tipo 2

Modelo Sugerido	AIC
SARIMA(3,1,3)x(1,1,1)₇	4068.17
SARIMA(3,1,3)x(2,1,1) ₇	4077.11
SARIMA(3,1,3)x(0,1,1) ₇	4080.52

APÊNDICE C

A Figura 12 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 1, para a Região Centro-Oeste.

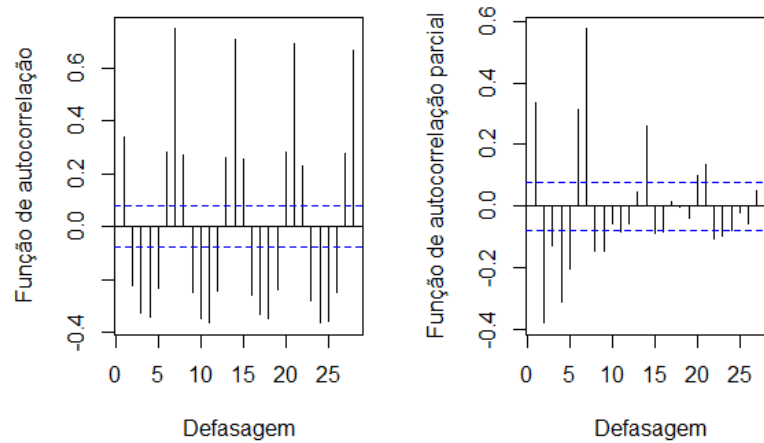


Figura 12 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Centro-Oeste.

A Figura 13 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 2, para a Região Centro-Oeste.

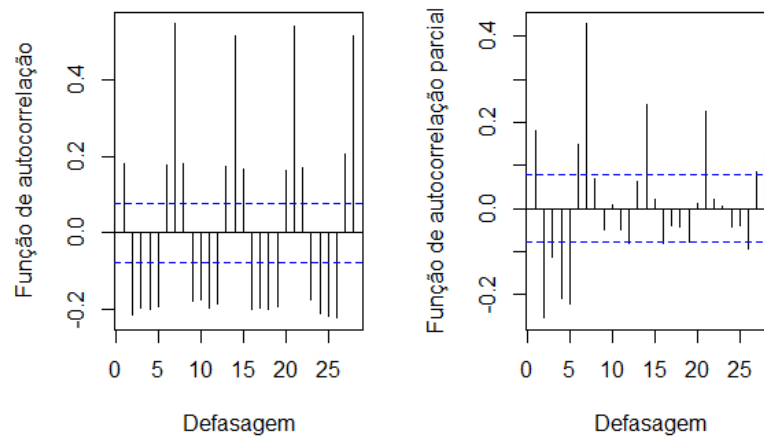


Figura 13 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Centro-Oeste.

A Figura 14 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 1, para a Região Nordeste.

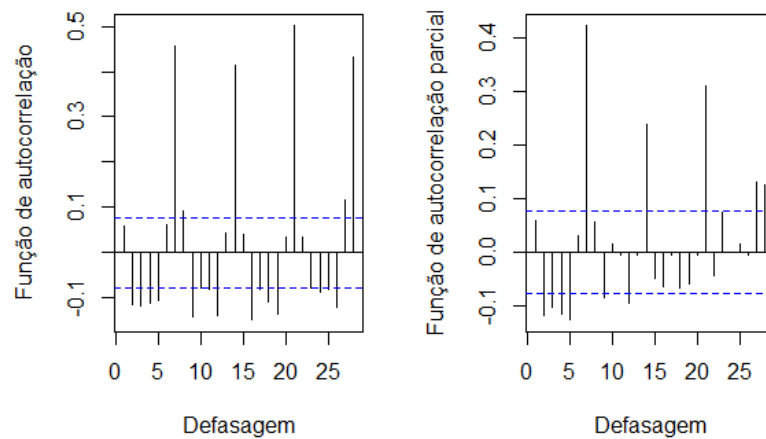


Figura 14 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Nordeste.

A Figura 15 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 2, para a Região Nordeste.

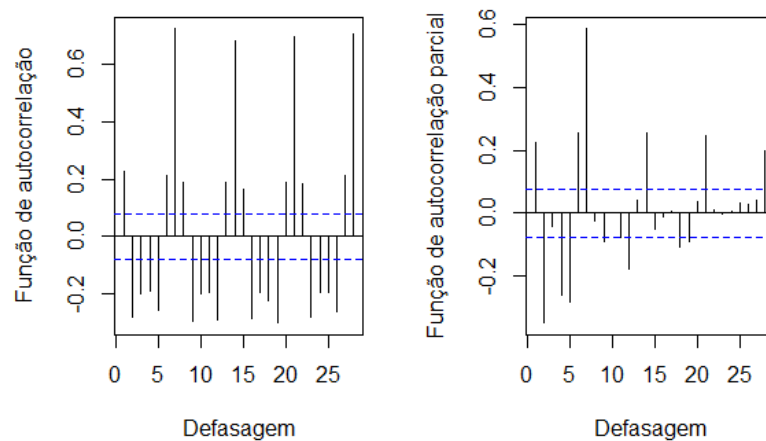


Figura 15 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Nordeste.

A Figura 16 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 1, para a Região Norte.

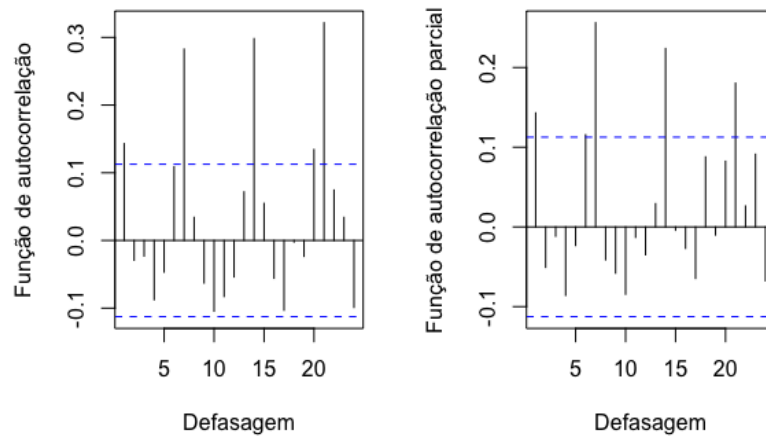


Figura 16 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Norte.

A Figura 17 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 2, para a Região Norte.

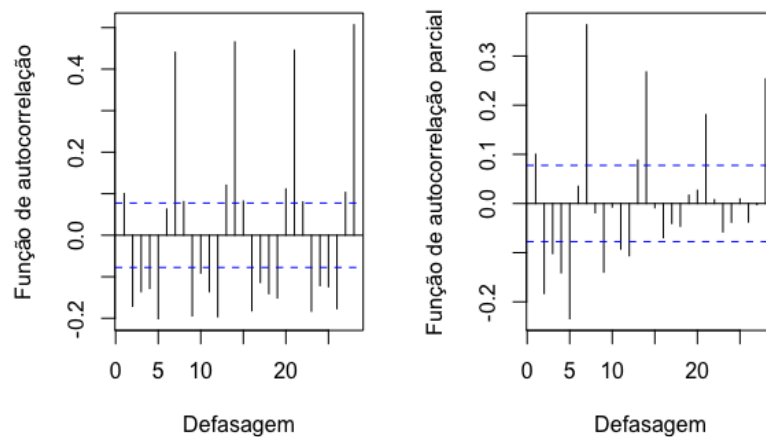


Figura 17 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Norte.

A Figura 18 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 1, para a Região Sudeste.

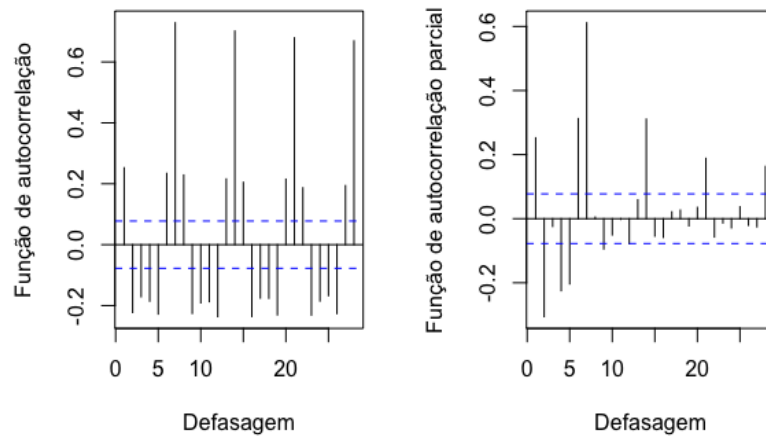


Figura 18 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Sudeste.

A Figura 19 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 2, para a Região Sudeste.

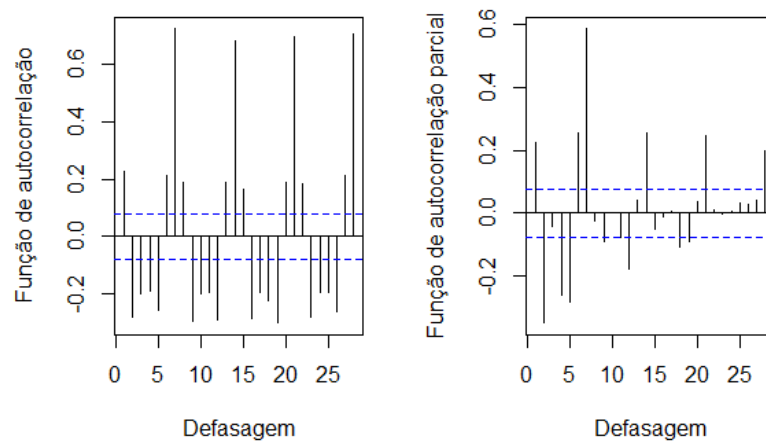


Figura 19 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Sudeste.

A Figura 20 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 1, para a Região Sul.

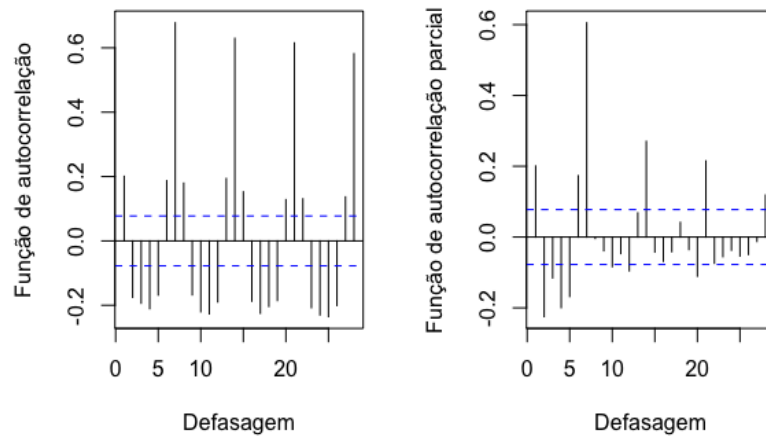


Figura 20 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 1, para a Região Sul.

A Figura 21 mostra as funções de autocorrelação e autocorrelação parcial dos hotéis do tipo 2, para a Região Sul.

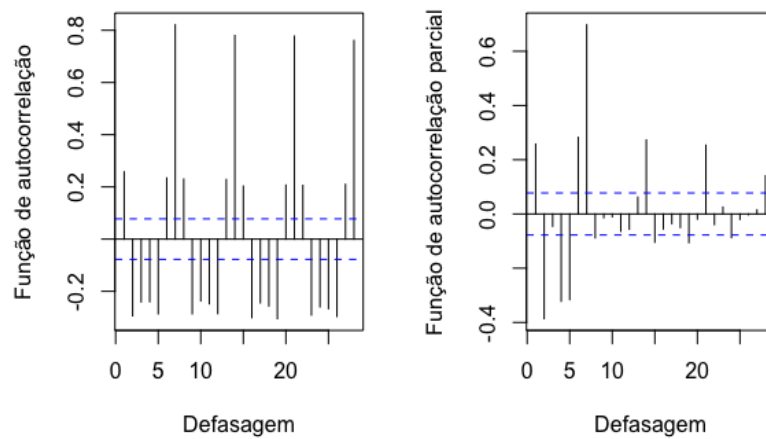


Figura 21 – Funções de autocorrelação e autocorrelação parcial da taxa de ocupação dos hotéis do tipo 2, para a Região Sul.

APÊNDICE D

A Figura 22 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo SARIMA(3,1,3)x(1,1,1)₇ para os hotéis do tipo 1, na Região Centro-Oeste.

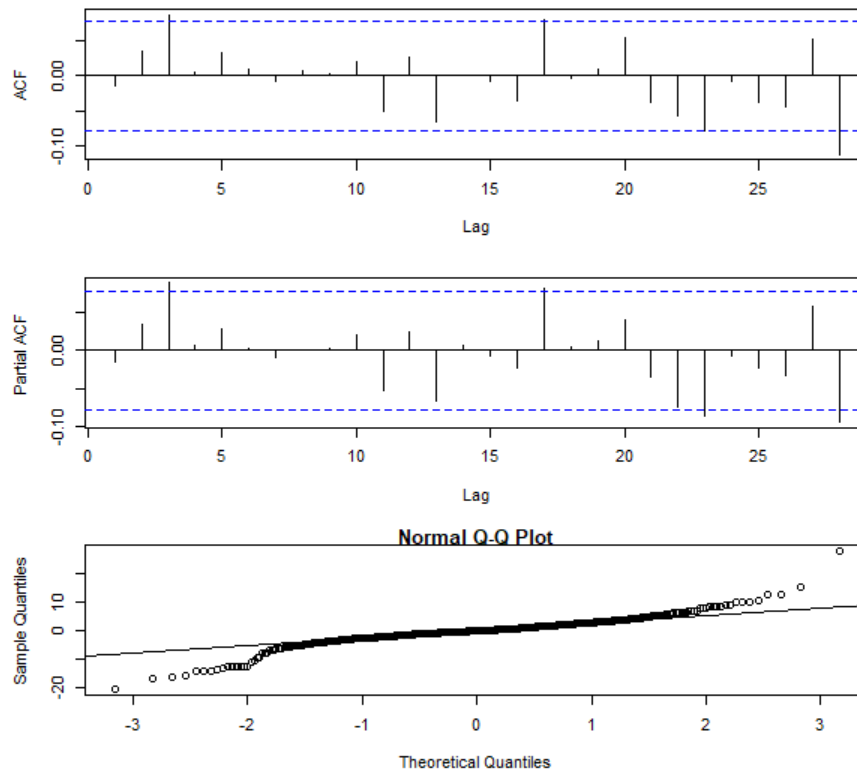


Figura 22 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Centro-Oeste

A Figura 23 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo SARIMA(5,1,1)x(1,1,1)₇ para os hotéis do tipo 2, na Região Centro-Oeste.

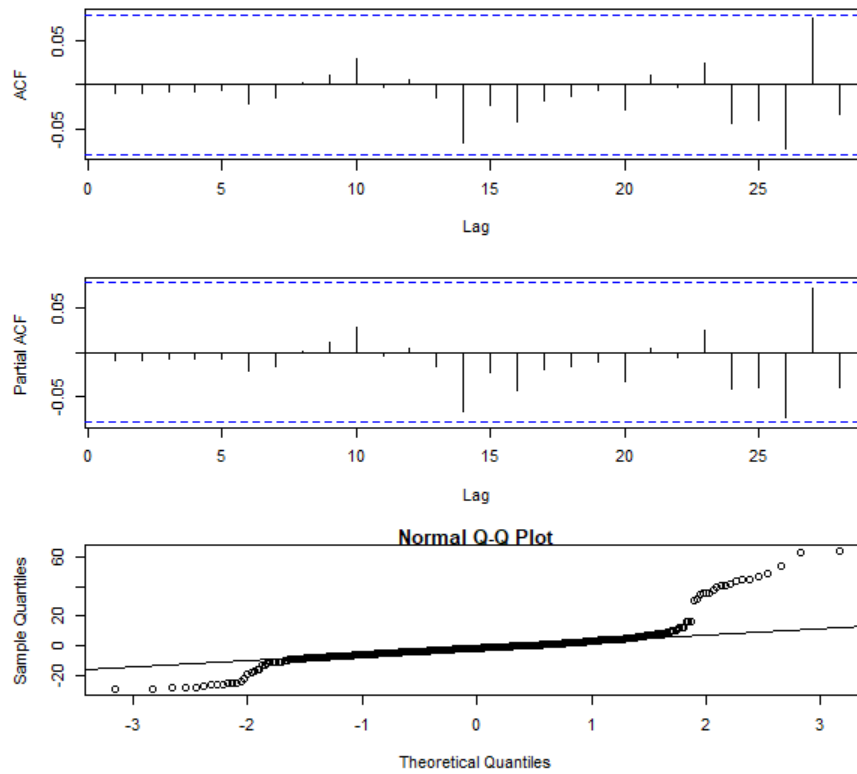


Figura 23 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Centro-Oeste

A Figura 24 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(5,1,1) \times (3,1,1)_7$ para os hotéis do tipo 1, na Região Nordeste.

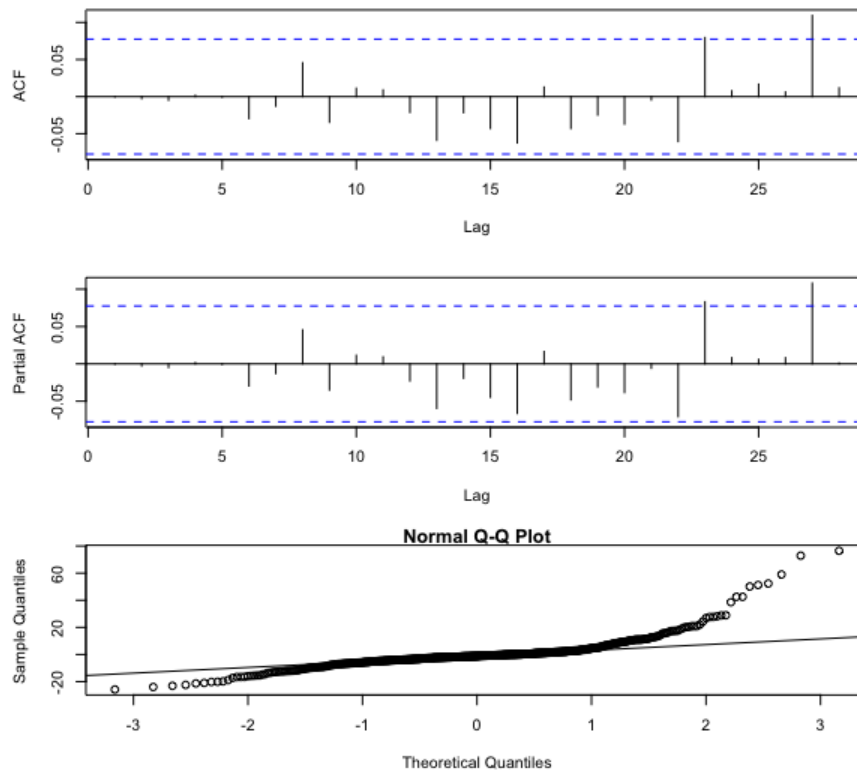


Figura 24 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Nordeste

A Figura 25 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo SARIMA(5,1,1)x(1,1,1)₇ para os hotéis do tipo 2, na Região Nordeste.

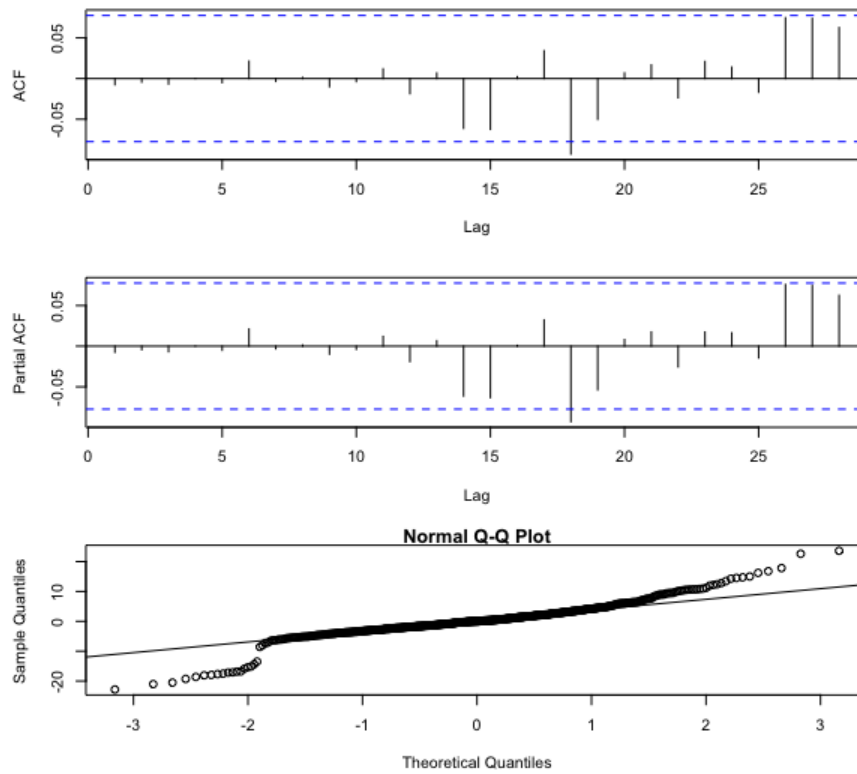


Figura 25 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Nordeste

A Figura 26 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(1,1,1) \times (0,1,1)_7$ para os hotéis do tipo 1, na Região Norte.

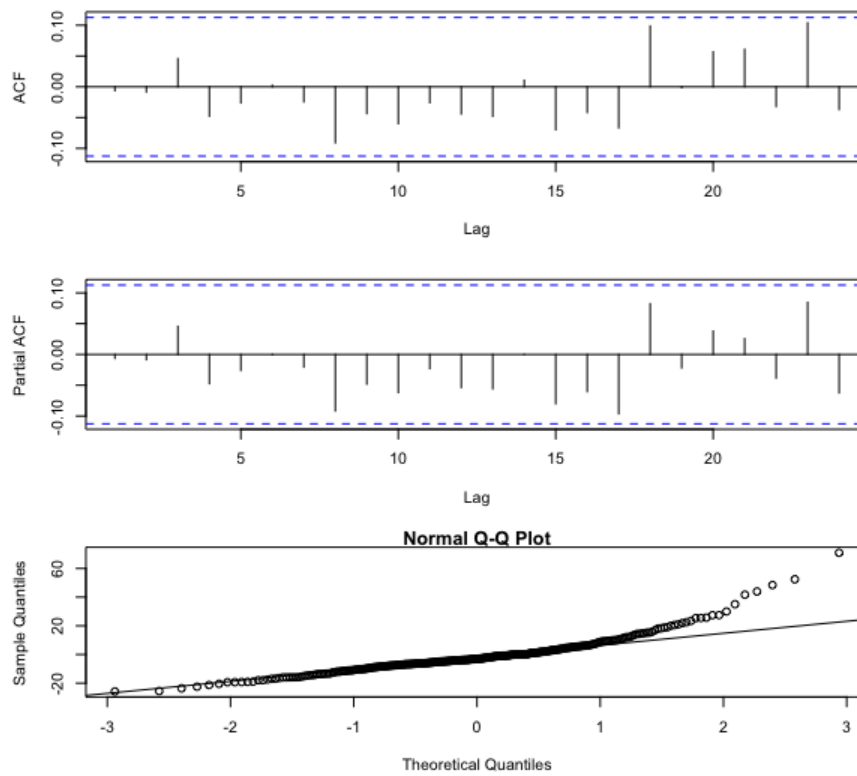


Figura 26 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Norte

A Figura 27 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(2,1,1) \times (1,1,1)_7$ para os hotéis do tipo 2, na Região Norte.

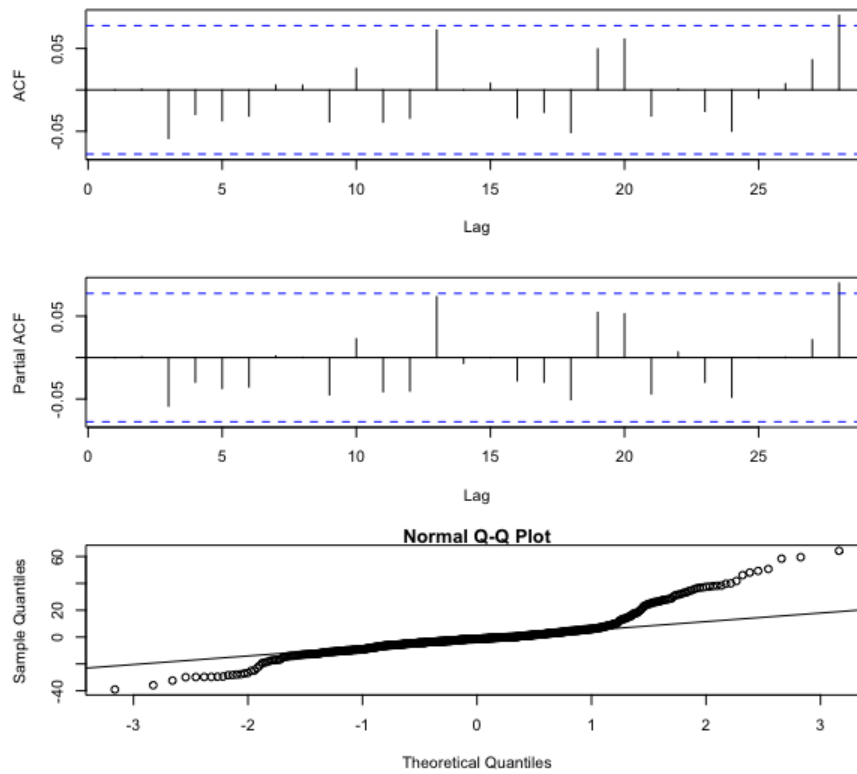


Figura 27 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Norte

A Figura 28 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(2,1,3) \times (0,1,2)_7$ para os hotéis do tipo 1, na Região Sudeste.

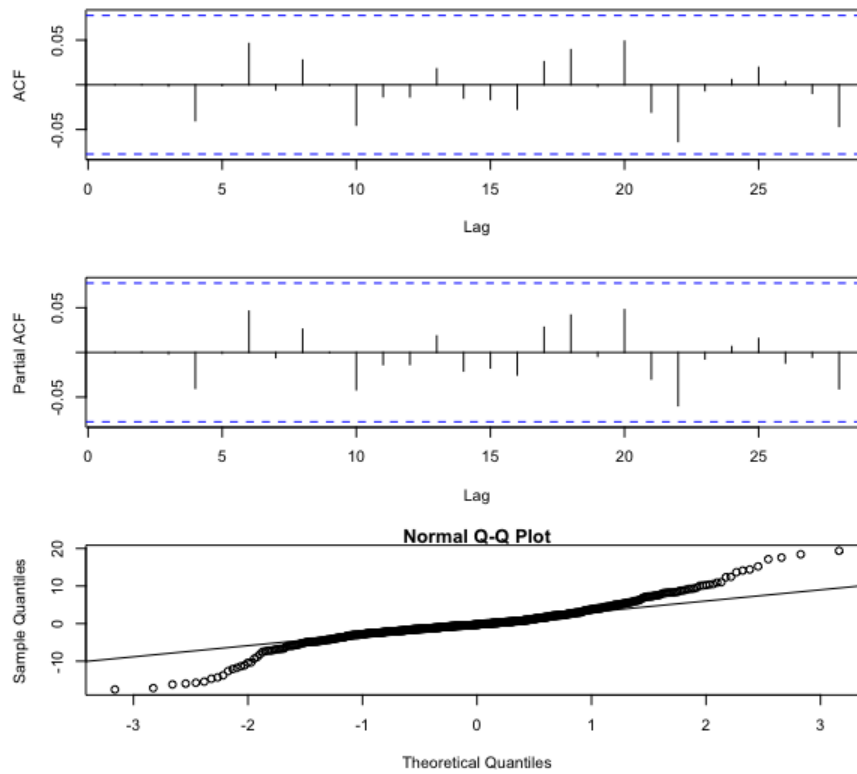


Figura 28 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Sudeste

A Figura 29 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(2,1,2) \times (1,1,1)_7$ para os hotéis do tipo 2, na Região Sudeste.

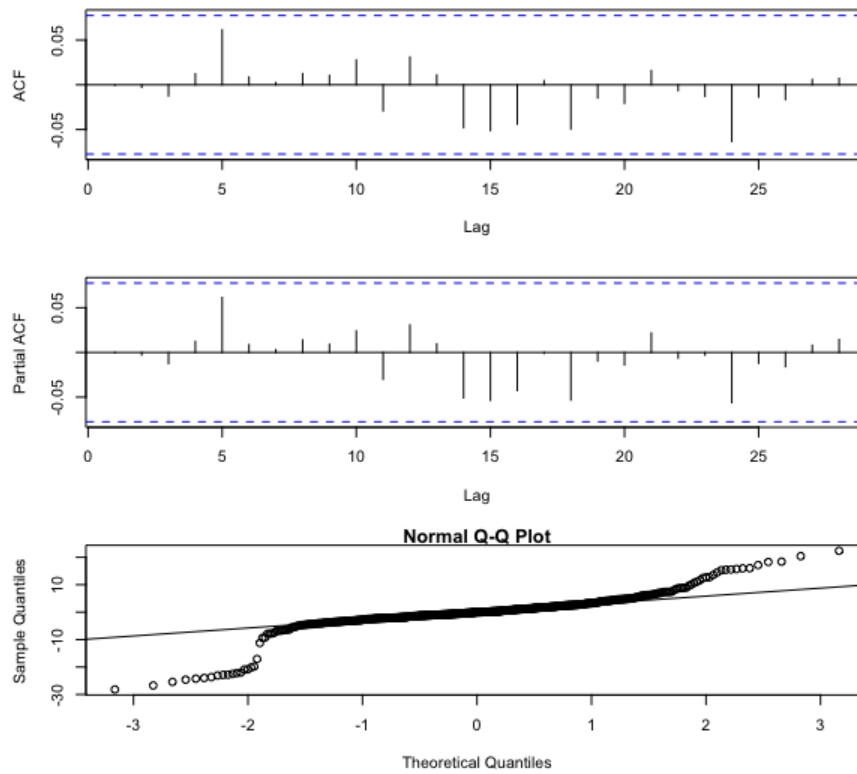


Figura 29 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Sudeste

A Figura 30 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo SARIMA(5,0,1)x(1,1,1)₇ para os hotéis do tipo 1, na Região Sul.

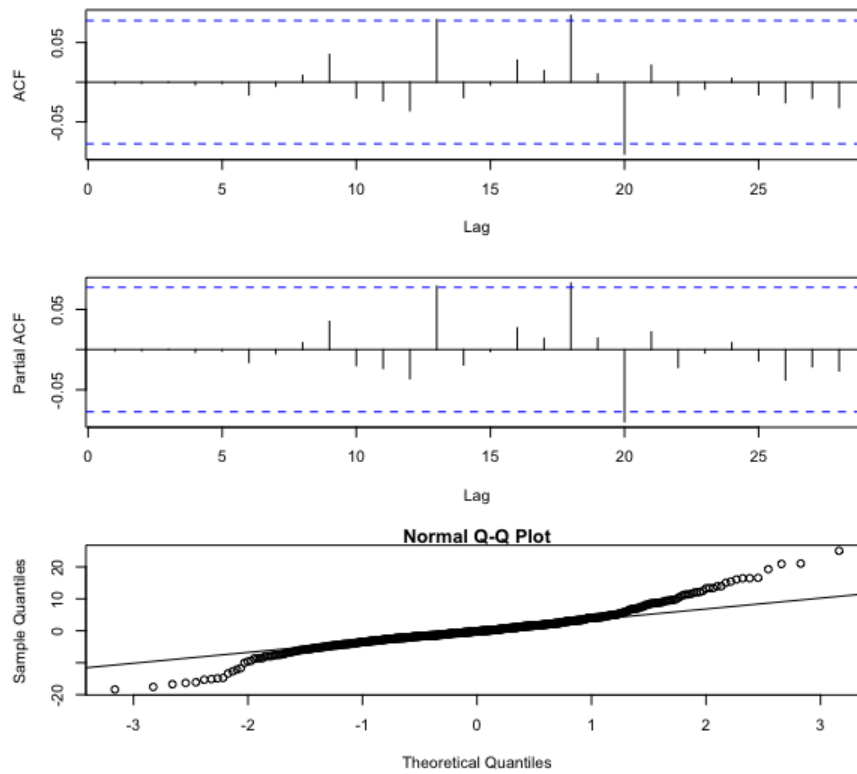


Figura 30 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 1, na Região Sul

A Figura 31 mostra os gráficos de autocorrelação, autocorrelação parcial e QQ-Plot dos resíduos do modelo $SARIMA(3,1,3) \times (1,1,1)_7$ para os hotéis do tipo 2, na Região Sul.

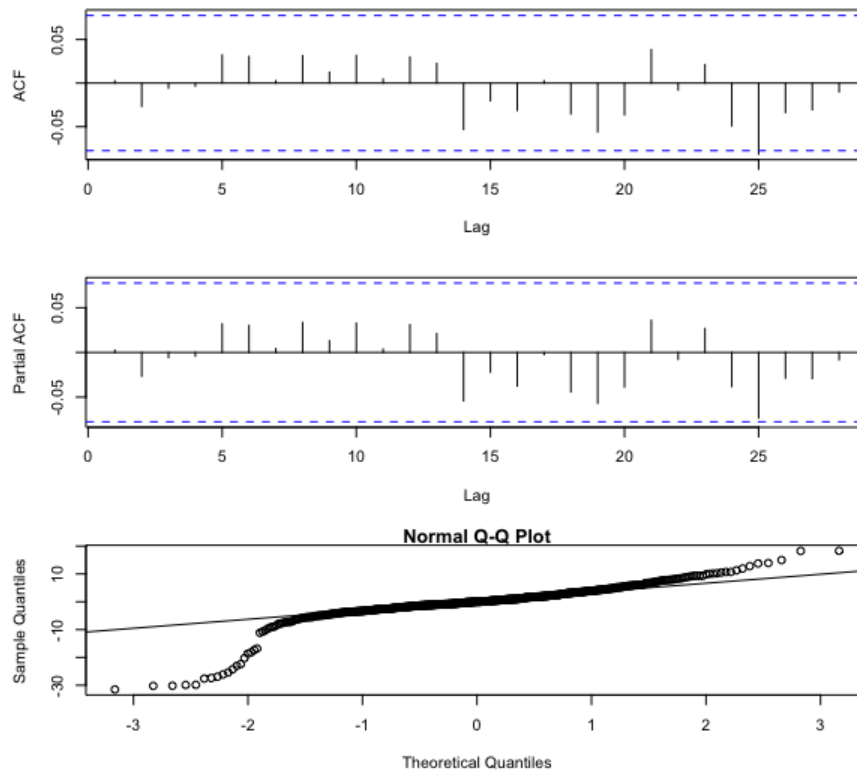


Figura 31 – ACF, PACF e QQ-Plot dos resíduos do modelo selecionado para os hotéis do tipo 2, na Região Sul