



PREVISÃO DE DEMANDA E DIMENSIONAMENTO DE LEITOS DE UTI  
POR MEIO DE TEORIA DE FILAS

Simone Aldrey Angelo

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Produção.

Orientador: Edilson Fernandes de Arruda

Rio de Janeiro

Maio de 2014

PREVISÃO DE DEMANDA E DIMENSIONAMENTO DE LEITOS DE UTI  
POR MEIO DE TEORIA DE FILAS

Simone Aldrey Angelo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA DE PRODUÇÃO.

Examinada por:

---

Prof. Edilson Fernandes de Arruda, D.Sc.

---

Prof. Basílio de Bragança Pereira, Ph.D.

---

Prof. André Assis Salles, D.Sc.

RIO DE JANEIRO, RJ - BRASIL  
JUNHO DE 2014

Angelo, Simone Aldrey

Previsão de demanda e dimensionamento de leitos de UTI  
por meio de teoria de filas/ Simone Aldrey Angelo – Rio de Janeiro:  
UFRJ/COPPE, 2014.

XI, 67 p.: il.; 29,7 cm.

Orientador: Edilson Fernandes de Arruda

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de  
Engenharia de Produção, 2014.

Referências Bibliográficas: p. 72-78.

1. Leitos de UTI. 2. Previsão de Demanda 3. Teoria de  
Filas.I.Arruda, Edilson Fernandes de. II. Universidade Federal do Rio  
de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

## AGRADECIMENTOS

A Deus e meu anjo da guarda por me iluminar e dar força para prosseguir nos momentos de dificuldade.

Ao Marcos, pelo apoio, paciência e dedicação. Você foi um grande e essencial companheiro nesta batalha.

Ao meu pai, Sebastião, pelo incentivo, força e por ser sempre um exemplo na busca de conhecimentos.

A minha mãe, Hulda, pelo apoio, otimismo e por muitas vezes me mostrar que eu era capaz.

Aos meus irmãos, Marco Antônio e Paulo Henrique, que estão sempre ao meu lado.

Aos professores Edilson, Basílio e André pelo apoio, competência e disponibilidade. E também os conhecimentos valiosos transmitidos que levarei sempre comigo.

Agradeço aos funcionários Roberta, Pedro e Pedrinho, por toda a ajuda prestada durante o mestrado. E também a Andreia que além de muito competente, esta sempre disposta a ajudar a todos. Que você seja sempre esta pessoa iluminada!

Aos amigos e grandes companheiros que o mestrado me trouxe, em especial: Raphael, Musbah, Ana, Marcão, Miranda, Caio, Juliana, Ivail, Marcelo, Maria Luiza e Luis Felipe. Todos sempre estiveram dispostos a ajudar, além de proporcionarem um ambiente agradável e bons momentos.

As amigas Giselle, Michele, Melina e Marina por todo apoio, carinho e paciência.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## PREVISÃO DE DEMANDA E DIMENSIONAMENTO DE LEITOS DE UTI POR MEIO DE TEORIA DE FILAS

Simone Aldrey Angelo

Maio/2014

Orientador: Edilson Fernandes de Arruda

Programa: Engenharia de Produção

A alocação de pacientes em uma Unidade de Terapia Intensiva (UTI) é um problema crítico para os administradores de redes de saúde. Portanto, é essencial determinar o número ótimo de leitos de UTI em uma dada região devido ao custo destes leitos e ao aumento de demanda por estes. Assim, este estudo procura determinar para seus administradores, principalmente, o número ótimo de leitos de UTI para um conjunto de hospitais da região metropolitana do Rio de Janeiro, utilizando dados referentes às solicitações diárias de leitos de UTI em 2010 e 2011. As informações coletadas foram utilizadas para se obter a previsão de demanda por meio de modelos de amortecimento exponencial e ARIMA, com essas previsões obteve-se a entrada de um modelo de teoria de filas. O resultado, ou saída, do modelo de teoria de filas utilizado é o número ótimo de leitos de UTI dado um critério definido a priori.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DEMAND FORECAST AND OPTIMAL ALLOCATION OF ICU BEDS THROUGH  
QUEUEING THEORY

Simone Aldrey Angelo

May/2014

Advisor: Edilson Fernandes de Arruda

Department: Production Engineering

The allocation of patients to a Intensive Care Unit (ICU) is a critical issue for managers of health networks. Therefore, it is essential to determine the optimal number of ICU beds in a given neighborhood due to both the expensive cost of these beds and the the ever increasing demand for them. Thus, this study is concerned with finding the optimal number of ICU beds in the metropolitan region of Rio de Janeiro, making use of information on the daily requests for ICU beds in a set of selected hospitals in 2010 and 2011. This information is used to obtain a demand forecast by means of exponential smoothing and ARIMA models, which is the input of queueing model. The result, or output, the queue theory model used is the optimal number of ICU beds, given an optimization criterion defined a priori.

# SUMÁRIO

<b>1. Introdução .....</b>	<b>1</b>
1.1 Considerações iniciais .....	1
1.2 Noções preliminares.....	3
1.3 Objetivos .....	7
1.4 Justificativa .....	7
1.5 Metodologia.....	8
1.6 Estrutura do trabalho.....	8
<b>2. Revisão Bibliográfica.....</b>	<b>10</b>
<b>3. Previsões de Séries Temporais .....</b>	<b>13</b>
3.1 Séries Temporais .....	13
3.2 Métodos de Previsão de Séries Temporais.....	14
3.3 Amortecimento exponencial.....	15
3.3.1 Amortecimento Exponencial Simples (método N,N).....	17
3.3.2 Método Linear de Holt (A,N) .....	18
3.3.3 Método de tendência amortecida (Ad,A).....	18
3.3.4 Método de Holt-Winters (A,A) e (A,M) .....	18
3.3.5 Previsões pontuais.....	20
3.3.6 Modelos de Espaço de Estado.....	22
3.4 Metodologia de Box-Jenkins .....	25
3.4.1 Modelos da Metodologia de Box-Jenkins.....	25
3.4.2 Função de Autocorrelação e Autocorrelação Parcial .....	27
3.4.3 Etapas da metodologia de Box-Jenkins .....	29
3.5 Medidas de desempenho .....	30
<b>4. Teoria de Filas.....</b>	<b>31</b>
4.1 Modelo M/M/s.....	31
<b>5. Aplicação dos Modelos de Previsão e Análises dos Resultados .....</b>	<b>34</b>
5.1 Análise 1.....	34

5.2 Análise 2.....	39
5.3 Conclusões das análises .....	41
<b>6. Aplicação de Teoria de Filas e Análises dos Resultados .....</b>	<b>43</b>
6.1 Cenário I.....	44
6.1.1 Distribuições .....	44
6.1.2 Modelo .....	45
6.1.3 Resultados .....	45
6.2 Cenário II.....	49
6.2.1 Distribuições .....	49
6.2.2 Modelo .....	50
6.2.3 Resultados .....	50
6.3 Cenário III.....	54
6.3.1 Distribuições .....	55
6.3.2 Modelo .....	56
6.3.3 Resultados .....	56
6.4 Cenário IV.....	59
6.4.1 Distribuições .....	59
6.4.2 Modelo .....	59
6.4.3 Resultados .....	60
6.5 Comparações dos resultados obtidos nos cenários propostos.....	63
6.5.1 Comparação dos cenários propostos com o atual - Previsão de 30 dias.....	64
6.5.2 Comparação dos cenários propostos com o atual - Previsão de 1 ano .....	65
6.5.3 Comparação do número de leitos/1000 habitantes dos cenários propostos (previsão de 30 dias e 1 ano) e atual, conforme Portaria GM/MS nº 1.101 .....	65
6.5.4 Comparação do número de leitos previsto x real .....	68
<b>7. Conclusão e Comentários Finais.....</b>	<b>70</b>



## LISTA DE FIGURAS

Figura 1- Perfis de previsão.....	17
Figura 2- Função de autocorrelação teórica.....	28
Figura 3 - Série Temporal 1.....	34
Figura 4- Decomposição da Série Temporal 1.....	35
Figura 5 - Autocorrelação da Série Temporal 1.....	35
Figura 6 – Autocorrelação Parcial da Série Temporal 1.....	35
Figura 7 - Modelo de Tendência Amortecida Multiplicativa ( $MM_dM$ ).....	36
Figura 8 - Previsão de 30 dias do modelo de Tendência Amortecida Multiplicativa ( $M,M_d,M$ ).....	37
Figura 9 - Previsão de 1 ano do modelo de Tendência Amortecida Multiplicativa ( $M,M_d,M$ ).....	37
Figura 10 - Previsão de 1 ano do modelo SARIMA.....	38
Figura 11 - Previsão de 1 ano do modelo Holt-Winters multiplicativo.....	38
Figura 12 - Previsão de 1 ano do modelo Holt-Winters aditivo.....	38
Figura 13 - Série Temporal 2.....	39
Figura 14- Previsão do modelo Tendência Amortecida Multiplicativa x dados reais.....	40
Figura 15 -Previsão do modelo SARIMA x dados reais.....	40
Figura 16- Previsão do modelo Holt-Winters aditivo x dados reais.....	41
Figura 17 - Previsão do modelo Holt-Winters multiplicativo x dados reais.....	41
Figura 18 - Histograma do tempo entre chegadas (2010 – 2011).....	44
Figura 19- Histograma do tempo de serviço (Hospitais CER).....	45
Figura 20 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário I (previsão de 30 dias).....	46
Figura 21 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário I (previsão de 1 ano).....	47
Figura 22 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário I (previsão fora da amostra - 30%).....	48
Figura 23 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário I (dados reais da parte fora da amostra - 30%).....	49
Figura 24 - Histograma do tempo de serviço (Todos hospitais do estado do RJ).....	50
Figura 25 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário II (previsão de 30 dias).....	51
Figura 26 – Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente - Cenário II (previsão de 1 ano).....	52
Figura 27 - Leitos de UTI com probabilidade $p$ de tempo de espera máximo por paciente – Cenário II (previsão fora da amostra - 30%).....	53

Figura 28 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (dados reais da parte fora da amostra - 30%) .....	54
Figura 29 - Saída das solicitações .....	55
Figura 30 - Análise das desistências .....	55
Figura 31 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão de 30 dias).....	56
Figura 32 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão de 1 ano) .....	57
Figura 33 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão fora da amostra - 30%) .....	58
Figura 34 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (dados reais da parte fora da amostra - 30%) .....	59
Figura 35 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (Previsão de 30 dias).....	60
Figura 36 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (previsão de 1 ano) .....	61
Figura 37 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (previsão fora da amostra - 30%) .....	62
Figura 38 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (dados reais da parte fora da amostra - 30%) .....	63
Figura 39 - Nº de leitos propostos x Nº de leitos atual (previsão de 30 dias) .....	64
Figura 40 - Nº de leitos propostos x Nº de leitos atual (previsão de 1 ano).....	65
Figura 41 - Leitos de UTI adulto (SUS)/1000 habitantes para cenários com previsão de 30 dias .....	67
Figura 42 - Leitos de UTI adulto (SUS)/1000 habitantes para cenários com previsão de 1 ano.....	67
Figura 43 -Nº de leitos: Previsto x Real - Sistema Estável .....	68
Figura 44 - Nº de leitos: Previsto x Real - (p=95% e t≤30 min) .....	68
Figura 45- Nº de leitos: Previsto x Real - (p=95% e t≤30 min) .....	69

## LISTA DE TABELAS

Tabela 1 - Hospitais x Nº de leitos .....	6
Tabela 2 - Taxonomia dos métodos de amortecimento exponencial .....	16
Tabela 3 - Fórmulas para cálculos recursivos e previsões pontuais .....	21
Tabela 4 - Equações de espaço de estado dos modelos de erro aditivo .....	23
Tabela 5 - Equações de espaço de estado dos modelos de erro multiplicativo .....	24
Tabela 6 - Comparação entre os modelos estimados na série 1 .....	36
Tabela 7 - Medidas de desempenho (dados estimados x dados reais) .....	39
Tabela 8 - Cenários propostos .....	63

# 1. Introdução

## 1.1 Considerações iniciais

Conforme dados do IBGE (2014), a população brasileira quase triplicou nos últimos 50 anos, partindo de aproximadamente 70 milhões de habitantes na década de 1960 para mais de 200 milhões, em 2014. Além deste crescimento expressivo, a expectativa de vida também aumentou consideravelmente, acarretando com isso um maior número de idosos. Estes fatores, que intrinsecamente são dinâmicos, implicam diretamente para o aumento da demanda por unidades hospitalares, em particular, Unidades de Tratamento Intensivo, o que dificulta a previsão adequada da demanda por essas unidades.

Ainda, soma-se às dificuldades dessa previsão, o fato de que a construção e a manutenção de leitos de uma Unidade de Tratamento Intensivo (UTI) são complexas e dispendiosas, devido à alta tecnologia e alto investimento estrutural e humano necessários para instrumentalizar uma UTI, de modo a oferecer segurança à vida do paciente. Ressalta-se, ademais, que um leito hospitalar não é apenas uma cama instalada num hospital para acomodar mais um paciente, isto é, um leito deve trazer consigo tudo de que necessita para ser funcional. Assim, diante da grande demanda por leitos de UTI e da dificuldade de sua previsão aliada à complexidade e ao custo inerentes, é comum que a oferta desses leitos não acompanhe sua demanda, ocasionando déficit de leitos de UTI.

De igual modo, o desequilíbrio entre a oferta e a demanda é influenciado pelo fato de que pacientes de UTI dificilmente agendam atendimento e têm estadias altamente variáveis, sendo assim um evento com características estocásticas. Como a natureza e o tempo de atendimento não são planejados com antecedência adequada, a disponibilidade de leitos de UTI muda dinamicamente. Assim, a admissão de um novo paciente de UTI provoca, pelo menos temporariamente, a redução da taxa de atendimento a novos pacientes.

No Brasil, o déficit de leitos provoca grandes filas. De acordo com Santos *et al.*(2010), muitos pacientes graves, ainda na fila por atendimento, evoluem para um estágio de maior gravidade, tornando-se impossível a restauração de sua saúde.

A crescente demanda por leitos UTI sem o correspondente crescimento de sua oferta aumenta os desafios para concretizar o preceito do SUS de universalidade e equidade, princípios regidos pela Constituição Federal, que por meio de seu Art. 196, garante aos cidadãos o direito à saúde, bem como o acesso universal e igualitário às ações e serviços para promoção e recuperação de sua saúde.

Segundo Barbosa (2004), no Brasil, na área de terapia intensiva não há equidade na distribuição dos leitos, existindo desigualdades nacionais e regionais, sendo o acesso limitado e

a qualidade dos serviços extremamente contrastante, variando de unidades altamente sofisticadas a outras sem a estrutura mínima necessária.

Se por um lado tem-se a necessidade de cumprir os princípios da universalidade e equidade de forma a atender toda demanda, por outro, não se deve ter excesso de oferta de UTI, pois o custo é alto. Sendo assim, é imprescindível determinar a quantidade ótima de leitos de UTI.

No Brasil, a Portaria nº 1.101 do Ministério da Saúde (2002) estabelece que a quantidade de leitos de UTI necessários é de 4% a 10% do total de leitos hospitalares, sendo que para os leitos hospitalares totais a definição é de 2,5 a 3 leitos para cada 1.000 habitantes. Entretanto, questiona-se se estas quantidades são suficientes para a realidade e demanda atual, pois a metodologia utilizada para estabelecê-las pode não ter sido a ideal ou ter se baseado em metodologias inadequadas à realidade atual.

A Pesquisa Operacional (PO) parece ser a metodologia adequada para apoiar as necessidades atuais de atendimento à demanda de UTI's. De acordo com Andrade (2000), a PO consiste numa metodologia administrativa que agrega, em sua teoria, economia, matemática, estatística e informática para subsidiar o processo de preparação, análise e tomada de decisão, por meio de um conjunto de métodos e modelos.

Papageorgiou (1978) apresentou um dos primeiros levantamentos sobre a aplicação da PO nos problemas de saúde e destacou sua contribuição para o planejamento, operação e gerenciamento dos hospitais. Desde então, a Pesquisa Operacional tem sido aplicada com grande sucesso na área da saúde. Royston (2009) afirma que, cada vez mais, a PO tem sido utilizada para enfrentar: o dia a dia da gestão hospitalar, operações com recursos limitados ou aspectos de planejamento de tratamento em saúde. Rais e Viana (2011) fizeram um levantamento de trabalhos que a PO foi aplicada exclusivamente na área de saúde. Eles citam mais de duzentas publicações, as quais abordam muitos trabalhos recentes, relevantes e que não foram abordados em outros artigos. Para os autores, o grande número de publicações recentes indica a importância da Pesquisa Operacional na saúde. Semelhante a Rais e Viana (2011), Cardoen *et al.* (2010) fizeram uma vasta revisão na literatura sobre Pesquisa Operacional no planejamento e agendamento de salas de operação, e desta forma, também demonstraram a contribuição da PO na área da saúde.

As técnicas de Pesquisa Operacional são implementadas na resolução de uma variedade de problemas na área da saúde. Para problemas de sistemas de agendamento Bard e Purnomo (2005) utilizaram uma combinação de métodos de programação inteira e heurística, Dobson *et al.* (2011) utilizaram modelos matemáticos estocásticos e Persson e Persson (2009) empregaram uma abordagem híbrida de simulação e otimização. Em abordagens de doação de órgãos e transplantes, Alagoz *et al.* (2004) empregaram o processo de decisão de Markov, já van den Hout *et al.* (2003) utilizaram simulação de eventos discretos. Para seleção de localização Ndiaye

e Alfares (2008) utilizaram um modelo de programação inteira binária, enquanto Smith *et al.* (2009) utilizaram programação inteira mista. E, para planejamento de capacidade Mayhew e Smith (2007) e Green e Savin (2008) utilizaram teoria das filas, enquanto Oliveira e Bevan(2008) desenvolveram um modelo multinível estocástico.

Os trabalhos comentados anteriormente são apenas uma pequena amostra dos problemas de saúde que utilizaram técnicas de Pesquisa Operacional, ressaltando que existem outras técnicas de PO aplicadas a problemas de saúde e também outros tipos de problemas de saúde solucionados com a PO. Porém, de acordo Rais e Viana (2011), apesar do grande número de problemas na área da saúde solucionados com a Pesquisa Operacional, ainda existem uma gama enorme a serem identificados e combatidos.

Neste sentido, é que se pretende encontrar o número ótimo de leitos de UTI-adulto do SUS regidos pela Central Estadual de Regulação (CER) do estado do Rio de Janeiro. Assim, conforme detalhado no capítulo 2, serão utilizados os modelos de previsão de séries temporais como amortecimento exponencial e Box-Jenkins, bem como a teoria de filas. Aos quais serão submetidos os dados das solicitações diárias por leitos de UTI-adulto do SUS dos hospitais regidos pela CER do estado do Rio de Janeiro nos anos de 2010 e 2011. Com o modelo de séries temporais que tiver o melhor desempenho obtém-se a previsão da taxa média de demanda, a qual será a entrada do modelo de teoria de filas que tem como saída o número ótimo de leitos de UTI, de acordo com critérios pré estabelecidos. Cabe ressaltar que a chegada de pacientes dá-se de acordo com um processo estocástico, cuja taxa média é determinada por meio dos referidos modelos de séries temporais.

## **1.2 Noções preliminares**

### **Unidade de Tratamento Intensivo (UTI)**

Segundo a Resolução nº 7 do Ministério da Saúde (2010), a UTI é uma área crítica destinada à internação de pacientes graves, que requerem atenção profissional especializada de forma contínua, materiais específicos e tecnologias necessárias ao diagnóstico, monitorização e terapia.

De acordo com Nishide *et al.*(2005), as UTI's surgiram durante a guerra da Criméia no século XIX, quando a enfermeira e estatística Florence Nightingale resolveu deixar os pacientes mais graves próximos a área onde as enfermeiras desenvolviam seus trabalhos, para que eles pudessem ter um melhor atendimento e uma maior vigilância. Na década de 1920 foram criadas as primeiras salas de recuperação para pacientes do pós-operatório neurocirúrgico, já em meados da década de 1950 começaram a ser desenvolvidas as primeiras técnicas e equipamentos modernos, e aos poucos foram sendo criadas outras unidades para atender

pacientes críticos e no pós-operatório. No Brasil, as primeiras UTI's surgiram no início da década de 70.

Embora as UTI's tenham surgido no Brasil desde os anos 1970, apenas em 1998 o Ministério da Saúde, através da portaria nº. 3432/98 do Ministério da Saúde (1998), definiu a forma de atendimento na UTI e estabeleceu padrões para: materiais e equipamentos necessários, composição da equipe, ambiente físico, dentre outros.

Para Barbosa (2004), o tratamento em Terapia Intensiva é uma modalidade avançada e indispensável da medicina atual e deve estar acessível a todo paciente grave e devidamente avaliado por um médico especialista, sendo o tempo um fator decisivo para garantir a sobrevivência desses pacientes. O objetivo da UTI é prover suporte médico e tecnológico de alta complexidade, de modo temporário, para pacientes graves potencialmente recuperáveis, conforme critérios pré-estabelecidos. Assim, Nogueira *et. al*(2012) argumentam que as UTI's são unidades de alto custo, por necessitarem de espaço físico diferenciado, equipamento de alta tecnologia e uma equipe multidisciplinar qualificada.

Segundo a Portaria nº 466 do Ministério da Saúde (1998), a UTI faz parte dos Serviços de Terapia Intensiva que, além dela, são compostos por Unidade de Tratamento Semi-Intensivo e o Serviço de Tratamento Intensivo Móvel.

- a) **Unidade de Tratamento Intensivo (UTI)** - constitui-se de um conjunto de elementos funcionalmente agrupados, destinado ao atendimento de pacientes graves ou de risco que exijam assistência médica e de enfermagem ininterruptas, além de equipamento e recursos humanos especializados. A UTI pode estar ligada a uma Unidade de Tratamento Semi-Intensivo.
- b) **Unidade de Tratamento Semi-Intensivo (Unidade Semi-Intensiva ou UTI de cuidados intermediários)** - constitui-se de um conjunto de elementos funcionalmente agrupados, destinado ao atendimento de pacientes, preferencialmente oriundos da UTI, que requeiram cuidados de enfermagem intensivos e observação contínua, sob supervisão e acompanhamento médico, este último não necessariamente contínuo, porém linear.
- c) **Serviço de Tratamento Intensivo Móvel** - constitui-se de um conjunto de elementos funcionalmente agrupados e uma frota de veículos devidamente projetados e equipados, destinados a garantir suporte avançado de vida durante o transporte de pacientes graves ou de risco, no atendimento de emergência pré-hospitalar e no transporte inter-hospitalar. Este Serviço pode ser parte integrante do serviço do hospital ou constituir-se em um prestador autônomo de Serviço de Tratamento Intensivo Móvel.

Os Serviços de Tratamento Intensivo dividem-se de acordo com a faixa etária dos pacientes atendidos, nas seguintes modalidades:

- a) **Neonatal** - destinado ao atendimento de pacientes com idade de 0 a 28 dias
- b) **Pediátrico** - destinado ao atendimento de pacientes com idade de 29 dias a 18 anos incompletos
- c) **Adulto** – destinado ao atendimento de pacientes com idade acima de 14 anos

Pacientes na faixa etária de 14 a 18 anos incompletos podem ser atendidos nos Serviços de Tratamento Intensivo Adulto ou Pediátrico, de acordo com o manual de rotinas do Serviço.

De acordo com a Portaria 466/98 do Ministério da Saúde (1998), denomina-se:

**UTI Especializada** - aquela destinada ao atendimento de pacientes em uma especialidade médica ou selecionados por grupos de patologias, podendo compreender: cardiológica, coronariana, neurológica, respiratória, trauma, queimados, dentre outras.

**Centro de Tratamento Intensivo (CTI)** - o agrupamento, numa mesma área física, de duas ou mais UTI's, incluindo-se, quando existentes, as Unidades de Tratamento Semi-Intensivo.

### **Sistema Único de Saúde (SUS)**

De acordo com o Portal da Saúde (2013), o SUS é um dos maiores sistemas públicos de saúde do mundo. Este sistema foi previsto na Constituição Federal de 1988, tendo como finalidade garantir aos cidadãos o direito à saúde, bem como o acesso universal e igualitário às ações e serviços para promoção e recuperação de sua saúde. Para tanto o SUS foi regulamentado pelas Leis 8.080/90 BRASIL (1990) também conhecida como Lei Orgânica da Saúde e 8.142/90 BRASIL (1990).

Este sistema abrange aproximadamente 6.1 mil hospitais credenciados, 45 mil unidades de atenção primária e 30.3 mil Equipes de Saúde da Família (ESF). Além disso, estão na composição do SUS os centros e postos de saúde, hospitais (inclusive universitários), laboratórios, hemocentros, bancos de sangue, fundações e institutos de pesquisa.

O SUS destina-se a todos os cidadãos. Pois além do amparo constitucional, o sistema único de saúde é financiado por recursos provenientes de tributos, especialmente impostos e contribuições sociais, que são pagos direta ou indiretamente pela população. Pelo SUS, a população tem acesso a consultas, exames, internações e tratamentos nos estabelecimentos de saúde, sejam eles federais, estaduais ou municipais, podendo ser públicos ou privados, sendo estes últimos conveniados ou contratados pelo gestor de saúde, atuando em regime complementar aos estabelecimentos públicos.



## Central Estadual de Regulação (CER)

De acordo com a SES (2013), a Central Estadual de Regulação (CER) foi implementada em 2005 pela Secretaria de Estado de Saúde e tem como finalidade executar a regulação estadual dos leitos vinculados ao SUS para os casos de alta complexidade.

O Estado do Rio de Janeiro tem hoje nove centrais reguladoras de leitos e procedimentos. São elas: Baixada Litorânea (Cabo Frio), Centro-Sul Fluminense (Três Rios), Médio Paraíba (Volta Redonda), Metropolitana II (Niterói), Noroeste Fluminense (Itaperuna), Norte Fluminense (Campos), Serrana (Teresópolis), Central Estadual de Regulação – Polo Estadual Rio de Janeiro e Metropolitana I, as duas últimas funcionando na cidade do Rio.

Na Tabela 1, seguem apenas os hospitais regidos pela CER entre 2010 e 2011, que continham leitos de UTI-adulto e UTI Intermediária-adulto integrantes do SUS. Entretanto, como nos anos de 2010 e 2011 a quantidade destes leitos não é conhecida, optou-se pelos respectivos dados de 2013 extraídos do CNES (2014).

Hospitais regidos pela CER em 2010 e 2011	Nº DE LEITOS EXISTENTES NO SUS - 2013	
	UTI - adulto	Cuidados intermediários-adulto
Hospital Estadual Adão Pereira Nunes	0	8
Hospital Estadual Albert Schweitzer	0	18
Hospital Estadual Alberto Torres	0	6
Hospital Estadual Azevedo Lima	8	0
Hospital Estadual Carlos Chagas	6	20
Hospital Estadual Getúlio Vargas	37	10
Hospital Estadual Prefeito João Batista Caffaro	0	2
Hospital Estadual Roberto Chabo	9	7
Hospital Estadual Rocha Faria	16	0
Hospital Estadual Vereador Melchiades Calazans	6	4
Hospital do Andaraí	16	0
Hospital Cardoso Fontes	8	0
Hospital Geral de Bonsucesso	17	0
Hospital de Ipanema	10	0
Hospital da Lagoa	25	0
Hospital dos Servidores do Estado	22	13
<b>TOTAL</b>	<b>180</b>	<b>88</b>

Tabela 1 – Hospitais x Nº de leitos

## **1.3 Objetivos**

### **Objetivo Geral**

Este estudo visa facilitar o planejamento da disponibilidade dos referidos leitos de UTI-adulto com a previsão da taxa média de chegada de pacientes, a partir da qual será dimensionado o número ótimo dos respectivos leitos. A facilitação aqui mencionada refere-se ao fornecimento de uma metodologia adequada e de aplicação simples, a qual possibilite a qualquer gestor encontrar uma previsão de demanda de acordo com o cenário atual e assim permita um planejamento que viabilize solucionar situações-problema similares de filas, inclusive indo além da área de saúde pública.

### **Objetivos Específicos**

- Identificar o modelo de previsão mais adequado para os dados do estudo.
- Determinar previsões de 30 dias e um ano para os dados do estudo.
- Determinar o número ótimo dos leitos de UTI para diversos cenários.
- Comparar o número atual dos referidos leitos com o número ótimo dos diversos cenários previstos.
- Comparar o número ótimo de leitos dos diversos cenários previstos com o número de leitos determinado pela portaria GM/MS nº 1.101.
- Comparar o número ótimo de leitos obtidos dos dados reais da parte fora da amostra com o número ótimo de leitos obtidos através da previsão fora da amostra.

## **1.4 Justificativa**

O Brasil possui cerca de 200 milhões de habitantes, dos quais aproximadamente 75% dependem exclusivamente do SUS. No estado do Rio de Janeiro a situação é parecida, cerca de 63% de seus aproximadamente 16,5 milhões habitantes também dependem exclusivamente do SUS. Entretanto, o quadro piora ao passo que nem todos os planos de saúde privados cobrem leitos de UTI. Isto posto conforme dados da Agência Nacional de Saúde Suplementar ANS (2013).

Portanto, é imprescindível determinar o número ótimo de leitos de UTI do SUS, uma vez que deve-se conciliar o cumprimento do direito constitucional à saúde e a implantação de novos leitos que são de elevado custo. Para tanto, é necessário uma metodologia prática para determinar a demanda que possibilite o devido gerenciamento e criação de novos de leitos de UTI, pois é notório o grande diferencial entre oferta e demanda de UTI, em especial na saúde pública.

## **1.5 Metodologia**

Para este estudo, que é caracterizado como quantitativo, realizou-se uma coorte retrospectiva do banco de dados da Central Estadual de Regulação, referentes aos pacientes com indicação médica para internação em leitos de UTI adulto do SUS dos hospitais regidos pela CER do estado do Rio de Janeiro, no período de 2010 a 2011.

Com posse deste banco de dados, iniciou-se uma pesquisa bibliográfica em busca de técnicas adequadas para se estimar o número de leitos de UTI, bem como a demanda por estes leitos. Assim, após a pesquisa bibliográfica, optou-se por utilizar a teoria das filas para estimar o número de leitos e os modelos de amortecimento exponencial e Box Jenkins para prever a futura demanda de pacientes. Estes métodos foram escolhidos por serem simples e apresentarem melhor desempenho que muitos outros mais sofisticados.

As previsões obtidas serão comparadas por meio das medidas de desempenho de precisão de modo a identificar o melhor método de previsão, o qual, por sua vez, será a entrada do modelo identificado de teoria das filas.

Como esta teoria abrange vários modelos, é necessário detectar o mais apropriado para o problema. E, para tanto, deve-se identificar a distribuição do tempo entre chegadas, a distribuição do tempo de serviço e o número de servidores.

A distribuição do tempo entre chegadas será obtida a partir da distribuição probabilística do intervalo entre chegadas dos pacientes, conforme informações extraídas do banco de dados da CER. Já a distribuição do tempo de serviço será obtida a partir da distribuição probabilística do intervalo de ocupação dos leitos, conforme informações extraídas do Departamento de Informática do SUS, DATASUS (2013). E, finalmente, o número de servidores refere-se ao número de leitos atualmente disponíveis, sendo este último obtido do Cadastro Nacional dos Estabelecimentos de Saúde, CNES (2013).

Desta forma, com o modelo de teoria das filas devidamente identificado, obtêm-se o número ótimo de leitos de UTI ao submeter os dados constantes deste estudo ao modelo escolhido. E de posse do número atual de leitos disponíveis, pode-se comparar o desempenho do sistema atual com o desempenho a ser obtido com a instalação do número ótimo de leitos obtido pelo modelo.

## **1.6 Estrutura do trabalho**

Além dessa introdução, este estudo conta com mais outros capítulos, estruturados da seguinte forma:

O Capítulo 2 apresenta uma revisão bibliográfica do problema estudado, na qual aborda-se aspectos específicos do problema relacionados tanto à previsão de demanda quanto ao dimensionamento dos leitos de UTI. No Capítulo 3 são abordados os modelos de previsão de

amortecimento exponencial e de Box-jenkins, bem como medidas de desempenho relativas a esses métodos, enquanto o Capítulo 4 apresenta uma breve introdução à teoria das filas, bem como o modelo a ser utilizado neste estudo. No Capítulo 5 são apresentados os resultados obtidos com os modelos de previsão, bem como a comparação entre eles com base nas medidas de desempenho das previsões e o Capítulo 6 mostra a aplicação do modelo de teoria de filas para quatro cenários, bem como a comparação do número ótimo de leitos destes cenários com o número de leitos determinado pela portaria GM/MS nº 1.101. Por fim, as conclusões e sugestões para trabalhos futuros estão no último capítulo, seguindo-se as referências bibliográficas.

## 2. Revisão Bibliográfica

Existem diversos estudos quantitativos e qualitativos acerca de leitos de UTI. Souza (2004) e Barbosa (2004) dimensionam o número de leitos de UTI de acordo com a distribuição geográfica, enquanto Gonçalves e Pontes, (2012) propõem otimizar o número de leitos de UTI através do método de gestão *PDCA (Plan-Do-Check-Act)* que consiste em quatro passos: planejar, executar, verificar e agir. Rhodes *et al.* (2012) identificaram o número total de leitos de UTI para cada país da Europa no período de julho de 2010 a junho de 2011 e utilizaram regressão linear para avaliar as possíveis associações entre o número de leitos de UTI e as seguintes variáveis: tamanho da população, produto interno bruto (PIB), PIB gasto com saúde e idade da população.

Muitas abordagens correlacionam o número de leitos com o número de habitantes e fazem uso da distribuição de Poisson para estimar o número de leitos necessários. Com este propósito Milne e Whitty (1995) descrevem os requisitos de número de leitos por milhões de habitantes, assumindo implicitamente uma relação linear entre a exigência de leitos e a população atendida. Milne e Whitty (1995) observam que sendo a terapia intensiva um evento raro numa grande população, o modelo teórico mais apropriado de utilização de leitos é o de uma distribuição de Poisson, tendo como parâmetro o número médio de leitos ocupados por dia. Entretanto, os mesmos ressaltam que duas premissas básicas devem ser observadas na utilização desse modelo, a taxa de ocupação do leito deve ser constante e os desvios na ocupação do leito devem ser estatisticamente independentes.

Pearson *et al.* (2012), objetivando derivar o número de leitos de UTI necessários para satisfazer a demanda, utilizaram dois modelos. O primeiro utilizou-se a demanda esperada e o tempo de internação baseados na distribuição de Poisson, já o segundo modelo utilizou simulação de Monte Carlo. Estes modelos apresentaram resultados diferentes quanto ao número de leitos: com o primeiro modelo obteve-se um número fixo e com o segundo que acompanhou a variação sazonal com o método Monte Carlo obtiveram-se números diferentes de leitos para o inverno e para o verão. No último modelo, os autores destacam a vantagem de não ser, inicialmente, essencial assumir um parâmetro de distribuição. Tornando-o, portanto, de grande utilidade na resolução de problemas determinísticos onde há incerteza significativa na entrada e nas condições de partida.

A simulação de eventos discretos também é uma abordagem muito utilizada na estimativa de leitos de UTI. Caminham neste sentido, Zhu *et al.* (2010) ao afirmar que a simulação de eventos discretos fornece mais flexibilidade para descrever a complexidade do sistema de UTI que as outras técnicas. Argumentam ainda que esta abordagem descreve a situação real com maior precisão e flexibilidade suficiente para testar diferentes cenários.

Semelhante a Zhu *et al.* (2010), Griffiths *et al.* (2006), também ressaltam a vantagem da simulação de eventos discretos de possibilitar a modelagem individual dos pacientes. E, assim, podendo conter certas características, tais como “fonte de admissão” ou “tipo de cirurgia” que influenciam o seu progresso através do sistema. Outro aspecto positivo desse modelo de simulação é o uso de uma interface visual amigável, que apresenta claramente o progresso dos pacientes. Entretanto, o mesmo autor resalta que devido à natureza aleatória de um modelo de simulação, os resultados obtidos podem não ser exatos. Tornando-se, desta forma, necessária a utilização de várias rodadas para obtenção de resultados razoavelmente precisos, o que pode causar demora no processamento deste modelo. Além disso, simulação propicia uma boa interface para uma análise de alternativas de gerenciamento, mas não permite diretamente o uso de técnicas de otimização, as quais realizam a comparação implícita de um número muito grande de alternativas.

Outra abordagem de grande importância na questão de leitos de UTI é a Teoria de Filas. (Shmueli *et al.* 2003) aplicaram a Teoria de Filas para otimizar admissões numa UTI e consequentemente aumentar o número de vidas salvas. Para tal, consideraram três políticas de admissão, comparando-as em termos de aumento de sobrevivência fornecido pela internação em leitos e determinaram a melhor dentre as 3 políticas de internação propostas, baseando-se num modelo de filas. Mcmanus *et al.* (2004) também utilizaram a Teoria de Filas para construir um modelo matemático de fluxo de pacientes de uma UTI, compararam previsões do modelo com o desempenho da unidade observada e exploraram a sensibilidade do modelo ao alterarem o tamanho da unidade. Por fim, concluíram que a teoria das filas é muito precisa para modelar a utilização e quantificação de leitos de UTI. Ressalta-se ainda que os modelos de Teoria de Filas propiciam o uso de otimização no planejamento de sistemas, o que constitui uma vantagem em relação simulação. A teoria de filas também se destaca por permitir que o sistema evolua ao longo do tempo, diferentemente dos sistemas estáticos, cujo comportamento depende exclusivamente de sua situação no instante considerado.

Assim, neste estudo, optou-se por abordar a teoria de filas a fim de determinar o número ótimo de leitos de UTI adulto, no entanto, este estudo possui um diferencial em relação aos outros aqui relatados que também utilizaram esta teoria. Este diferencial se identifica pela combinação de teoria das filas com técnicas de previsão de demanda, isto é, a taxa média de chegada do modelo de filas será obtida a partir da previsão de demanda por séries temporais, buscando desta forma um modelo ainda mais preciso.

De acordo com Rais e Viana (2011), a previsão de demanda é essencial em planejamentos na área da saúde, uma vez que seus resultados fornecem a entrada de muitos problemas de otimização. Porém, de acordo com Jones *et al.* (2008), apesar de sua importância, os métodos de previsão ainda não despertaram muito interesse na área de saúde, haja vista que existem poucos estudos que exploram o uso de métodos de previsão para prever a demanda de

pacientes. Jones *et al.* (2008) buscaram a previsão de demanda de pacientes diários para os serviços de emergência de três hospitais distintos. E para tal, utilizaram: regressão de séries temporais, amortecimento exponencial, modelos ARIMA (especificamente o modelo SARIMA) e modelos de redes neurais. Para avaliar a precisão, os autores compararam as previsões obtidas com o método de referência utilizado. E concluíram que a metodologia existente que é baseada em modelos de regressão linear múltipla com base em calendário é razoável, pois apesar dos métodos propostos terem apresentado melhores resultados em relação a metodologia existente, esta melhora foi pequena. O estudo de Jones *et al.* (2008) também confirmou uma crença muito difundida de que a demanda diária por serviços de emergência é caracterizada por padrões sazonais e semanais. No mesmo seguimento, Xue *et al.* (2001) buscaram a previsão do número de pacientes com insuficiência renal terminal em 2010 através de dados históricos de 1982-1997. Para tal, utilizaram modelos de amortecimento exponencial e um modelo que combina regressão de tendência temporal com um modelo autorregressivo.

Assim, como Jones *et al.* (2008) e Xue *et al.* (2001), neste estudo buscou-se a previsão de demanda de pacientes através de modelos de previsão, porém diferentemente destes autores, este estudo não focou apenas modelos de previsão de demanda, isto é, utilizou-se da previsão destes modelos para compor o modelo de teoria de filas. E assim, buscou o número de leitos ótimo para um determinado tempo máximo de espera. Ao trabalhar com tempo máximo de espera, este estudo também se diferencia de muitos, como Siddharthan *et al.* (1996) e Marinho(2006) que também empregam teoria de filas na área de saúde, porém trabalham com tempo médio de espera ao invés de tempo máximo de espera.

## 3. Previsões de Séries Temporais

### 3.1 Séries Temporais

Segundo Chatfield (2000), uma série temporal é um conjunto de observações ordenadas cronologicamente. Uma das finalidades de estudo de séries temporais é a previsão, que consiste em prever valores futuros de uma série de dados baseados em valores presentes e passados de uma ou mais séries. Maiores detalhes e discussões sobre séries temporais podem ser encontrados em Box e Jenkins (1970), Brockwell e Davis (1991), Kendall e Ord (1990), Montgomery *et. al* (1990), dentre outros.

De acordo com Chatfield (2000), se valores futuros podem ser previstos exatamente a partir de valores passados, a série é considerada determinística. Entretanto, a maioria das séries é estocástica ou aleatória, nas quais o futuro é apenas parcialmente determinado pelos valores passados.

Assim, as séries passada e futura sofrem influência dos mesmos fatores, nos quais incluem-se as componentes: irregular, tendência, sazonalidade e cíclica. Estas componentes são descritas a seguir, de acordo com Salles (2007).

A componente irregular, ou ruído branco da série, se refere às variações aleatórias ou irregulares da série, em geral, provocadas por eventos imprevisíveis e sem correspondência com intervalos de tempo. A componente tendência corresponde a um movimento sistemático, crescente ou decrescente, dos valores da série temporal.

A sazonalidade se refere às variações periódicas no tempo, ciclos que se repetem regularmente na série histórica provocados por fatores climáticos, por eventos associados a determinado período do tempo ou outros ligados a estações do tempo. E, por último, a componente cíclica, que se assemelha à componente sazonal, apresenta como diferença a amplitude do ciclo, ou dos ciclos, que apresentam um maior espaçamento, isto é, se referem a um prazo maior ao passo que variação sazonal se aplica, ou se reporta, ao curto prazo.

O mesmo autor finaliza que a análise de séries temporais divide-se em três etapas, sendo a primeira correspondente à decomposição da série em cada um dos componentes explicitados. A segunda etapa é a própria modelagem dessas componentes. E na terceira etapa modela-se a série com os modelos de cada componente da variável de interesse.

Também é importante mencionara classificação dos métodos de análise de séries temporais, as quais podem ser estacionárias e não estacionárias. Assim, é importante para o estudo de séries temporais obter o conceito de estacionariedade, que a seguir é apresentado por Pereira (1980):

*“Intuitivamente diz-se que a série temporal é estacionária se não existe uma mudança sistêmica na média (tendência), na variância e se as variações*



*cíclicas forem removidas. A maior parte da teoria de séries temporais trata de séries estacionárias, por esta razão, a análise é feita transformando as séries não estacionárias em estacionárias para usar a teoria. Por exemplo, é possível remover a tendência e a variação sazonal dos dados e então tentar modelar a variação dos resíduos por processo estocástico estacionário.”*

Existem testes para verificar a estacionariedade, neste estudo utilizou-se o teste de raiz unitário, designado por teste de estacionariedade de Dickey e Fuller Aumentado ou teste ADF. Para Clemente e Clemente (2011), o teste ADF consiste em estimar os parâmetros autoregressivos de várias ordens e testar a presença de raiz unitária a partir da comparação da estatística  $t$  com a estatística Dickey-Fuller.

### **3.2 Métodos de Previsão de Séries Temporais**

Existem vários métodos de previsão de séries temporais, tais como: médias móveis, decomposição clássica, amortecimento exponencial, metodologia de Box-Jenkins, análise de regressão, redes neurais, dentre outros. No entanto, para este estudo, os métodos escolhidos a ser abordados foram o amortecimento exponencial e a metodologia de Box-Jenkins, que também é conhecida como modelo auto-regressivo integrado de média móvel (ARIMA). Esses métodos foram escolhidos por serem simples, populares e apresentarem melhor desempenho que muitos outros mais sofisticados.

Gardner (1985) desenvolveu um artigo de review, desde o trabalho original de Brown e Holt em 1950, uma revisão crítica sobre amortecimento exponencial, na qual reconhece a dificuldade de identificar o método, devido à falta de procedimentos úteis para simplificar a identificação dos métodos de amortecimento exponencial. Assim, como comentários finais no próprio estudo de Gardner, McKenzie afirma que, ao contrário do amortecimento exponencial, o ARIMA é útil na seleção de modelos. Na mesma obra, Hillmer complementa que uma das maiores contribuições do ARIMA para a área de análises de séries temporais é a sua abordagem organizada para a modelagem destas séries, o que muito facilita o início da modelagem de séries temporais para pessoas com pouca experiência. E, por fim, Hillmer salienta a necessidade de se ter um modelo voltado para os novatos na área de amortecimento exponencial.

Gardner (2006) é um artigo de review atualizado, e nesta obra o autor comenta que quando surgiu seu review de 1985, muitos autores afirmaram que o amortecimento exponencial deveria ser desconsiderado, pois acreditavam que o mesmo era um caso especial de modelagem ARIMA ou um procedimento *ad hoc* sem lógica estatística. Porém, para Gardner o argumento de o amortecimento exponencial ser um caso especial de modelagem ARIMA foi amadurecido. E, atualmente, sabe-se que os métodos de amortecimento exponencial são ideais para uma classe muito geral de modelos de espaço de estado, que é de fato mais abrangente do que a

classe ARIMA. Esta abrangência maior se comprova pelo fato de todos os métodos de amortecimento exponencial lineares ter modelos ARIMA equivalentes, o que não ocorre quanto aos métodos de amortecimento exponencial não lineares.

Por outro lado, Hyndman e Khandakar (2008) afirmam que, para os dados sazonais, existem mais modelos ARIMA do que métodos de amortecimento exponencial. Entretanto, para os autores uma ampla classe de modelos não se mostra vantajosa, e para sustentar tal afirmação ele se baseia em Hyndman *et al.* (2002), o qual demonstrou que no geral os métodos de amortecimento exponencial tiveram melhor desempenho que os modelos ARIMA para a competição sazonal M3 que é a competição mais recente de Makridakis e em Hyndman *et al.* (2001) o qual especula que a ampla classe de modelos ARIMA realmente prejudica o desempenho de previsão, porque introduz incerteza adicional. Assim, concluem que a menor classe de amortecimento exponencial é suficientemente rica para capturar a dinâmica de quase todas as séries temporais.

Hyndman e Khandakar (2008) também acrescentam que há um mito muito difundido que os modelos ARIMA são mais gerais do que o amortecimento exponencial, o que para os autores é uma falácia. E como justificativa argumentam que de um lado têm-se os modelos ARIMA que não apresentam um modelo equivalente para os métodos de amortecimento exponencial não lineares, porém de outro lado o amortecimento exponencial também não apresenta outros modelos que o ARIMA contém. Assim sendo, os autores, admitem que os modelos se complementam.

### **3.3 Amortecimento exponencial**

Segundo Morettin e Toloi (2006), a maioria dos métodos de previsão partem do princípio que observações passadas contêm informações sobre o comportamento da série temporal. Sendo que estes métodos objetivam distinguir o padrão de qualquer ruído interno às observações, para então usar esse padrão de modo a prever valores futuros da série. Os autores vão além ao afirmar que:

*“Uma grande classe de métodos de previsão, que tenta tratar ambas as causas de flutuações em séries de tempo, é a das suavizações. Técnicas específicas desse tipo assumem que os valores extremos da série representam a aleatoriedade e, assim, por meio da suavização desses extremos, pode-se identificar o padrão básico.”*

Os autores ainda ressaltam que o modelo de suavização exponencial ou, em tradução livre amortecimento exponencial, é muito popular devido à sua simplicidade, eficiência computacional e razoável precisão.

De acordo com Hyndman *et. al* (2008), no amortecimento exponencial sempre se inicia pela componente tendência, que se caracteriza por ser uma combinação do termo nível ( $\ell$ ) e do termo tendência ( $b$ ). O nível e a tendência podem ser combinados de diversas maneiras, dando cinco tipos de tendências futuras. Se  $T_h$  denotar a tendência da previsão ao longo do período de tempo  $h$ , e  $\Phi$  denotar o parâmetro de amortecimento ( $0 < \Phi < 1$ ), têm-se os seguintes tipos de tendência ou padrão de crescimento:

$$\text{Nenhum: } T_h = \ell \tag{1}$$

$$\text{Aditivo: } T_h = \ell + bh \tag{2}$$

$$\text{Aditivo amortecido: } T_h = \ell + (\Phi + \Phi^2 + \dots + \Phi^h)b \tag{3}$$

$$\text{Multiplicativo: } T_h = \ell b^h \tag{4}$$

$$\text{Multiplicativo amortecido: } T_h = \ell b^{(\Phi + \Phi^2 + \dots + \Phi^h)} \tag{5}$$

Para os mesmos autores, o método da tendência amortecida é apropriado quando há uma tendência na série temporal, embora acreditem que no futuro, essa tendência se dá apenas num curto espaço de tempo. As equações da tendência amortecida, como o nome indica, amortecem a tendência conforme o aumento do horizonte temporal, melhorando a precisão da previsão, particularmente para situações de horizontes de previsões longos.

Hyndman *et. al* (2008) ainda afirmam que além do componente tendência, o método de amortecimento pode também ser constituído pela componente sazonalidade e pela inclusão do erro, tanto na forma aditiva como na multiplicativa. Porém este erro tem sido frequentemente ignorado, pois a distinção entre erros aditivos e multiplicativos não faz diferença para previsões pontuais. Com o erro ignorado, obtêm-se quinze métodos possíveis de amortecimento exponencial, aos quais estão representados sua taxonomia na Tabela 2 e seu perfil de previsão na Figura 1.

De acordo com Hyndman e Khandakar (2008), a taxonomia dos métodos de amortecimento exponencial originou-se da classificação de Pegels (1969). Tempo depois foi estendida por Gardner (1985), modificada por Hyndman *et al.* (2002) e estendida novamente por Taylor (2003).

Componente Tendência	Sazonalidade		
	N(Nenhum)	A(Aditivo)	M(Multiplicativo)
N (Nenhum)	N,N	N,A	N,M
A (Aditivo)	A,N	A,A	A,M
A <sub>d</sub> (Aditivo amortecido)	A <sub>d</sub> ,N	A <sub>d</sub> ,A	A <sub>d</sub> ,M
M (Multiplicativo)	M,N	M,A	M,M
M <sub>d</sub> ( Multiplicativo amortecido)	M <sub>d</sub> ,N	M <sub>d</sub> ,A	M <sub>d</sub> ,M

Tabela 2- Taxonomia dos métodos de amortecimento exponencial

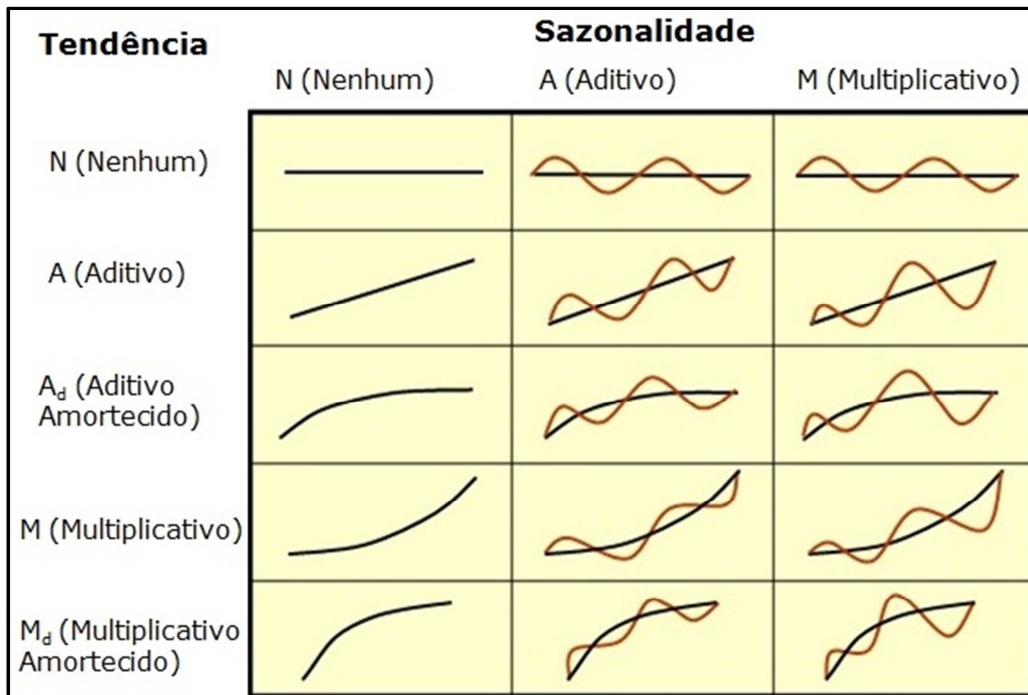


Figura 1- Perfis de previsão. Fonte: Adaptada de Gardner (1985)

Alguns destes métodos também são bem conhecidos por outro nome. Por exemplo, N-N descreve o método da amortização exponencial simples ou SES- single exponential smoothing, A-N descreve o método linear de Holt, A<sub>d</sub>N descreve o método de tendência amortecida, A-A descreve o método aditivo de Holt-Winters e A-M descreve o método multiplicativo de Holt-Winters. Tais métodos serão descritos a seguir, com base em Hyndman et al.(2008).

### 3.3.1 Amortecimento Exponencial Simples (método N,N)

Devido ao trabalho de Brown realizado em 1950 e publicado em Brown (1959), o método de amortecimento exponencial Simples ou SES (Single Exponential Smoothing), utiliza a previsão do período anterior, ajustando-a com o erro da própria previsão. Assim, esta previsão é representada pela seguinte equação:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha (y_t - \hat{y}_t), \quad (6)$$

sendo que  $\hat{y}_{t+1}$  é a previsão do período  $(t+1)$  que será calculado,  $y_t$  é o valor real atual da observação,  $\hat{y}_t$  é a previsão anterior, feita para o mesmo período de  $y_t$  e a sua subtração  $(y_t - \hat{y}_t)$  é o erro da previsão. O valor de  $\alpha$  é conhecido por parâmetro de amortecimento e está limitado a valores entre 0 e 1.

É possível observar que a nova previsão é simplesmente a velha previsão com o ajustamento do erro que ocorreu na última previsão. Sabe-se que, para valores de  $\alpha$  reduzidos (próximos de 0) o ajustamento do erro é muito pequeno e os pesos dos dados são quase uniformes, tornando a previsão mais insensível e muito próxima da média de todos os valores.

Se os valores de  $\alpha$  forem próximos de 1, os pesos decrescem muito rapidamente, implicando com que as observações mais recentes tenham influência efetiva no resultado da previsão.

### 3.3.2 Método Linear de Holt (A,N)

Holt(1957) estendeu o amortecimento exponencial simples para amortecimento exponencial linear, permitindo a previsão de dados com tendência. A previsão para o método desenvolvido é encontrada usando duas constantes de amortecimento,  $\alpha$  e  $\beta^*$  (ambas com valores entre 0 e 1), e três Equações:

$$\text{Nível:} \quad l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + b_{t-1}) \quad (7)$$

$$\text{Tendência:} \quad b_t = \beta(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1} \quad (8)$$

$$\text{Previsão:} \quad \hat{y}_{t+h/t} = l_t + b_{t-1}h \quad (9)$$

sendo que  $l_t$  denota a estimaco para o nvel da srie no perodo t e  $b_t$  a estimaco do declive (tendncia) da srie no perodo t.  possvel observar que  $b_t$   a mdia ponderada da tendncia do perodo anterior,  $b_{t-1}$ , e a estimaco da tendncia baseada na diferena entre nveis sucessivos.

### 3.3.3 Mtodo de tendncia amortecida (Ad,A)

Gardner (1985) props uma modificaco no mtodo linear de Holt para permitir a “amortizaco” de tendncia. As Equaces deste mtodo so:

$$\text{Nvel:} \quad l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + \phi b_{t-1}), \quad (10)$$

$$\text{Tendncia:} \quad b_t = \beta(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1} \quad (11)$$

$$\text{Previso:} \quad \hat{y}_{t+h/t} = l_t + (\phi + \phi^2 + \dots + \phi^h) b_t \quad (12)$$

Assim, a tendncia para a previso da etapa  $y_{t+1}$    $\phi b$ , ou seja, a tendncia  amortecida por um fator de  $\phi$  para cada perodo de tempo adicional futuro.

### 3.3.4 Mtodo de Holt-Winters (A,A) e (A,M)

Holt(1957) props um mtodo para dados sazonais. Seu mtodo foi estudado por Winters (1960), e agora  usualmente conhecido como “mtodo de Holt-Winters”.

O mtodo de Holt-Winters  baseado em trs equaces de amortecimento: uma para o nvel, outra para a tendncia e outra para a sazonalidade.  semelhante ao mtodo de Holt, porm com uma equaco adicional que lida com a sazonalidade. De fato, h dois diferentes mtodos de Holt-Winters, dependendo se a sazonalidade  modelada de forma aditiva ou multiplicativa.

## Modelo multiplicativo de Holt-Winters

As equações básicas do método multiplicativo de Holt-Winters são as seguintes:

$$\text{Nível:} \quad l_t = \alpha y_t / s_{t-m} + (1-\alpha) (l_{t-1} + b_{t-1}), \quad (13)$$

$$\text{Tendência:} \quad b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1} \quad (14)$$

$$\text{Sazonalidade:} \quad s_t = \gamma y_t / (l_{t-1} - b_{t-1}) + (1 - \gamma) s_{t-m} \quad (15)$$

$$\text{Previsão:} \quad \hat{y}_{t+h/t} = (l_t + b_t h) s_{t-m+h}^+ \quad (16)$$

sendo que  $m$  é o comprimento da sazonalidade (o número de meses ou trimestres num ano),  $l_t$  representa o nível da série,  $b_t$  denota a tendência,  $s_t$  é a componente sazonalidade e  $\hat{y}_{t+h/t}$  é a previsão para  $h$  períodos posteriores e  $h^+ = [(h-1) \bmod m] + 1$ .

A Equação (15) é ligeiramente diferente da Equação usual de Holt-Winters tal como em Makridakis *et. al* (1998) ou Bowerman *et. al* (2005). Estes autores substituem (15) por:

$$s_t = \gamma y_t / l_t + (1 - \gamma) s_{t-m} \quad (17)$$

## Modelo aditivo de Holt-Winters

A componente de sazonalidade do método Holt-Winters também pode ser tratada adicionalmente, embora na prática pareça ser bem menos utilizada. As equações básicas do método aditivo de Holt-Winters são as seguintes:

$$\text{Nível:} \quad l_t = \alpha (y_t - s_{t-m}) + (1 - \alpha) (l_{t-1} + b_{t-1}) \quad (18)$$

$$\text{Tendência:} \quad b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1} \quad (19)$$

$$\text{Sazonalidade:} \quad s_t = \gamma (y_t - l_{t-1} - b_{t-1}) + (1 - \gamma) s_{t-m} \quad (20)$$

$$\text{Previsão:} \quad \hat{y}_{t+h/t} = l_t + b_t h + s_{t-m+h}^+ \quad (21)$$

A Equação (19) é idêntica à Equação (14) do método, as únicas diferenças entre as outras equações são relativas aos índices sazonais que agora são adicionados e subtraídos ao invés de multiplicados e divididos.

Tal como acontece no modelo multiplicativo, a equação usual para o termo sazonal é ligeiramente diferente da Equação (20). A maioria dos livros utiliza

$$s_t = \gamma^* (y_t - l_t) + (1 - \gamma^*) s_{t-m} \quad (22)$$

Se  $l_t$  for substituído em (25), obtêm-se:

$$s_t = \gamma^*(1-\alpha) (y_t - l_{t-1} - b_{t-1}) + [1 - \gamma^* (1 - \alpha)] s_{t-m} \quad (23)$$

Assim, obtém-se a mesma previsão usando a abordagem de substituir  $\gamma$  em (20) por  $\gamma^*(1-\alpha)$ .

### 3.3.5 Previsões pontuais

Segundo Hyndman *et. al* (2008), na Tabela 3 encontra-se equações recorrentes para calcular as previsões pontuais  $h$  períodos posteriores de todos os métodos de amortecimento exponencial. Em cada caso  $l_t$  denota o nível de série no momento  $t$ ,  $b_t$  a inclinação no momento  $t$ ,  $s_t$  a componente de sazonalidade da série no tempo  $t$ , e  $m$  denota o número de estações de ano;  $\alpha$ ,  $\beta^*$ ,  $\gamma$  e  $\varphi$  são constantes, e  $\varphi_h = \varphi + \varphi^2 + \dots + \varphi^h$ .

Alguns casos especiais interessantes podem ser obtidos ajustando os parâmetros de amortecimento para valores extremos. Por exemplo, se  $\alpha = 0$ , o nível é constante ao longo do tempo, se  $\beta^* = 0$ , a inclinação é constante ao longo do tempo, e se  $\gamma = 0$ , a sazonalidade padrão é constante ao longo do tempo. Em outro extremo, previsões ingênuas (ou seja,  $\hat{y}_{t+h|t} = y_t$  para todos  $h$ ) são obtidos usando o método de (N, N) com  $\alpha = 1$ . Finalmente, os métodos de tendência aditiva e métodos tendência multiplicativa são casos especiais de seus homólogos amortecidos, obtidos deixando  $\varphi = 1$ .

TENDÊNCIA	SAZONALIDADE		
	N	A	M
N	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$ $\hat{y}_{t+h t} = \ell_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t s_{t-m+h_m^+}$
A	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $\hat{y}_{t+h t} = \ell_t + hb_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t + hb_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t-m+h_m^+}$
A <sub>d</sub>	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $\hat{y}_{t+h t} = \ell_t + \phi_h b_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$
M	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $\hat{y}_{t+h t} = \ell_t b_t^h$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t b_t^h + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t b_t^h s_{t-m+h_m^+}$
M <sub>d</sub>	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $\hat{y}_{t+h t} = \ell_t b_t^{\phi h}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^\phi) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t b_t^{\phi h} + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^\phi)) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = \ell_t b_t^{\phi h} s_{t-m+h_m^+}$

Tabela 3 - Fórmulas para cálculos recursivos e previsões pontuais. Fonte: Adaptado de Hyndman et. al. (2008)



### 3.3.6 Modelos de Espaço de Estado

Segundo Hyndman *et. al.* (2008), para cada método de amortecimento exponencial da Tabela 3 existem dois possíveis modelos de espaço de estado, nos quais um corresponde ao erro aditivo e o outro ao erro multiplicativo. Se os valores dos parâmetros utilizados forem os mesmos, os dois modelos dão previsões pontuais equivalentes, entretanto diferentes intervalos de previsão. Assim, pode-se dizer que o amortecimento exponencial fornece apenas previsões pontuais, enquanto os modelos de espaço de estado fornecem as mesmas previsões pontuais, porém com intervalos de previsões e outras propriedades.

Para distinguir os modelos com erros aditivos e multiplicativos adiciona-se uma letra extra na frente da notação do método. O trio (E,T,S) refere-se aos três componentes: erro, tendência e sazonalidade.

Existem modelos de espaço de estado semelhantes para todas as variações de amortecimento exponencial. O modelo geral envolve um vetor estado  $x_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$  e as equações de espaço de estado da seguinte forma:

$$y_t = w(x_{t-1}) + r(x_{t-1})\varepsilon_t \quad (24)$$

$$x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t \quad (25)$$

sendo que  $\{\varepsilon_t\}$  é um ruído branco Gaussiano com média zero e variância  $\sigma^2$ , e  $m_t = w(x_{t-1})$ . O modelo com erro aditivo tem  $r(x_{t-1})=1$ , assim  $y_t = m_t + \varepsilon_t$ . Já o modelo com erro multiplicativo tem  $r(x_{t-1})=m_t$ , assim  $y_t = m_t(1+\varepsilon_t)$ , logo  $\varepsilon_t = (y_t - m_t)/m_t$  é o erro relativo do modelo multiplicativo.

Cada método da Tabela 3 pode ser escrito da forma dada em (24) e (25). As equações para os modelos de erro aditivo e multiplicativo são apresentadas, respectivamente nas Tabelas 4 e 5.

	SAZONALIDADE		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A <sub>d</sub>	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M <sub>d</sub>	$\mu_t = \ell_{t-1}b_{t-1}^\phi$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

Tabela 4—Equações de espaço de estado dos modelos de erro aditivo. Fonte: Adaptado de Hyndman et. al. (2008)

TENDÊNCIA	SAZONALIDADE		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A <sub>d</sub>	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M <sub>d</sub>	$\mu_t = \ell_{t-1}b_{t-1}^\phi$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Tabela 5 - Equações de espaço de estado dos modelos de erro multiplicativo. Fonte: Adaptado de Hyndman et. al. (2008)

### 3.4 Metodologia de Box-Jenkins

A metodologia de Box-Jenkins, aqui tratada, baseia-se em Mabert (1975), sendo a mesma aplicável a séries estocásticas estacionárias. Caso a série seja não estacionária, é possível transformá-la em estacionária e então aplicar a referida metodologia.

Se uma série não estacionária envolver mudanças apenas no valor médio, ela pode se tornar estacionária por meio de diferenciações apropriadas da série original. Porém, se nesta série existir o fator sazonalidade, deve ser requerida uma diferenciação sazonal de comprimento  $s$ , isto é, cria-se uma nova série com diferenciações entre os períodos sazonais.

Outra causa comum de não estacionariedade é a magnitude da variação ao redor da média, neste caso para induzir a homogeneidade e tornar a série estacionária, pode-se transformar a série calculando o logaritmo natural dos dados brutos e depois prosseguir com a diferenciação dos valores deste logaritmo, mas geralmente utiliza-se a transformação de Box-Cox (1964).

#### 3.4.1 Modelos da Metodologia de Box-Jenkins

##### Modelo Auto-regressivo (AR)

Neste modelo a observação presente da série  $Y_t$  pode ser explicada como uma combinação linear das  $p$  observações passadas. A estrutura auto-regressiva geral é expressa por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t, \quad (26)$$

sendo que:

$p$  – determina o número de passos entre a observação passada e a previsão da próxima observação,

$\phi_i (1 \leq i \leq p < \infty)$  – representa um coeficiente de ponderação para  $p^{\text{th}}$  períodos anteriores,

$a_t$  – ruído branco com média zero e variância  $\sigma_a^2$ .

A razão para que este modelo seja referido como um modelo auto-regressivo decorre da análise de regressão, isto é, existe a relação da variável dependente  $Y_t$  com um conjunto de variáveis independentes ( $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ ) que são os valores passados da série, então o modelo é auto-regressivo.

##### Modelo Médias Móveis (MA)

Os modelos médias móveis são formados por combinação linear dos erros de previsão anteriores ( $a_t$ ). A estrutura médias móveis geral é expressa por::

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}, \quad (27)$$

sendo:

$\theta_i (1 \leq i \leq q < \infty)$  – representa um coeficiente de ponderação para  $p^{\text{th}}$  períodos anteriores

$a_t$  – ruído branco com média zero e variância  $\sigma_a^2$

Este modelo é similar ao de amortecimento exponencial, isto é, ele implica que é possível obter informações valiosas para previsões futuras, considerando a soma ponderada dos erros de previsão anteriores.

### Modelo Auto-Regressivo Médias Móveis (ARMA)

Este modelo é uma combinação de ambos modelos, auto-regressivo e médias móveis, onde  $Y_t$  é descrito por seus valores passados e pelos erros de previsão anteriores ( $a_t$ ). A estrutura geral ARMA é expressa por::

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (28)$$

sendo que:

$\phi_i (1 \leq i \leq p)$  – representa um coeficiente de ponderação para  $p^{\text{th}}$  períodos anteriores auto-regressivo

$\theta_i (1 \leq i \leq q)$  – representa um coeficiente de ponderação para  $p^{\text{th}}$  períodos anteriores médias móveis

$a_t$  – ruído branco no instante  $t$ ,  $t \in \mathbb{N}$

Usando o operador de defasagem  $B$ , tem-se:

$$\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (29)$$

$$\text{ou } \phi(B) Y_t = \theta(B) a_t \quad (30)$$

### Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA)

O modelo ARIMA  $(p, d, q)$  é composto pela integração do modelo ARMA e a técnica de diferenciação. Tal técnica permite a transformação de uma série não estacionária em estacionária.

O número necessário de diferenciações para tornar uma série estacionária é denominado ordem de integração  $(d)$ . A estrutura geral ARIMA  $(p, d, q)$  é expressa por:

$$\phi(B) \nabla^d Y_t = \theta(B) a_t \quad (31)$$

sendo que:

$\phi(B)$  – representa o operador auto-regressivo de ordem  $p$

$\theta(B)$  – representa o operador médias móveis de ordem  $q$

$a_t$  – ruído branco no instante  $t$ ,  $t \in \mathbb{N}$

$d$  – representa o número de diferenças

$\nabla = 1 - B$  representa o operador diferença

Este operador diferença é definido como:

$$Y_t - Y_{t-1} = Y_t - BY_t = (1-B)Y_t = \nabla Y_t, \quad (32)$$

$$\nabla^d = (1-B)^d. \quad (33)$$

### 3.4.2 Função de Autocorrelação e Autocorrelação Parcial

#### Função de Autocorrelação

A função de autocorrelação da amostra é utilizado para determinar qual dos dois modelos, auto-regressivo ou médias móveis, é apropriado para uma série temporal particular. A amostra de autocorrelação mede a correlação entre períodos na série temporal. A correlação ( $r_k$ ) para  $k$  períodos entre observações individuais é definida como:

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad k = 0, 1, \dots, k, \quad (34)$$

sendo,  $n$  o número de observações,  $y_t$  a série temporal observada e  $\bar{y}$  o desvio médio.

Pesquisas com séries temporais fornecem padrões da função autocorrelação que sugerem o uso dos modelos autoregressivo ou médias móveis. Esses padrões distintos representam a característica fundamental da metodologia de Box-Jenkins. Se a série temporal é melhor descrita por um modelo autoregressivo, a autocorrelação exibe um padrão de decaimento de uma função exponencial, oscilação ou em senoidal, conforme mostrado na Fig 2.A, 2.B e 2.C. Porém, se a série temporal é melhor descrita por um modelo médias móveis, a autocorrelação é marcada por grandes picos, conforme Fig 2.D e 2.E.

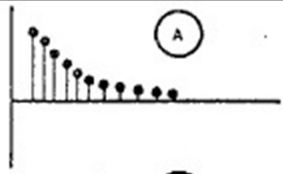
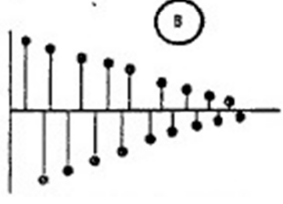
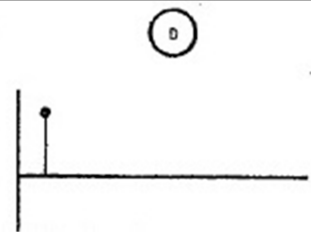

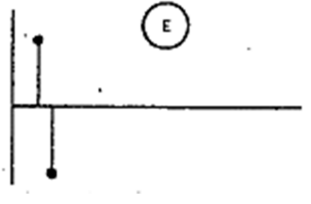
Modelos Autoregressivos	Modelos Médias Móveis
 	
AR1 - Primeira ordem tende a decair exponencialmente ou oscilando	MA 1 - Pico de primeira ordem no primeiro intervalo
	
AR 1 - Segunda ordem tende a onda senoidal amortecida	MA 2 - Pico de segunda ordem no primeiro e segundo intervalo

Figura 2- Função de autocorrelação teórica. Fonte: Adaptado de Mabert (1975)

Assim, para identificar o modelo de serie temporal mais apropriado, calcula-se recursivamente a amostra de autocorrelação conforme Eq.(34) e compara-se seu padrão com os padrões teóricos de autocorrelação.

### Função de Autocorrelação Parcial

A autocorrelação parcial da amostra mede a força do relacionamento entre períodos nas série. A autocorrelação da amostra ( $A_{kk}$ ) é definida como:

$$A_{kk} = \frac{r_k - \sum_{j=1}^{k-1} (A_{k-1,j} r_{k-j})}{1 - \sum_{j=1}^{k-1} (A_{k-1,j} r_j)}, k = 2, 3, \dots k \quad (35)$$

sendo,

$$A_{kj} = A_{k-1,j} - A_{kk} A_{k-1,k-j}, j = 1, 2, \dots k - 1 \quad (36)$$

A função de autocorrelação parcial tem um comportamento dual ao da função de autocorrelação e é muito útil na escolha do modelo mais adequado.

### 3.4.3 Etapas da metodologia de Box-Jenkins

A seguir, para modelar uma série temporal, serão descritas as três etapas da metodologia de Box-Jenkins.

#### Identificação do modelo

O processo de identificação do modelo envolve a tentativa de identificar o(s) modelo(s), utilizando o modelo geral ARIMA  $(p,d,q)$ , no qual denomina-se como:  $p$  os parâmetros auto-regressivos,  $d$  o(s) processos de diferenciação e  $q$  os parâmetros de médias móveis.

O modelo pode ser identificado com base nas seguintes características: conhecimento prévio do padrão dos dados, análise do gráfico das séries, análise da função de autocorrelação da amostra, análise da função de autocorrelação parcial da amostra. Sendo que, nesta combinação de fatores, as funções de autocorrelação e autocorrelação parcial usualmente são consideradas os principais indicadores.

Este processo envolve dois passos:

1-Realiza-se a diferenciação da série original até que a mesma se torne estacionária. Porém, caso a série contenha sazonalidade deve-se realizar a diferenciação sazonal ao invés da regular.

2- Seleciona-se o modelo ARIMA mais apropriado para descrever a série temporal. Para tal, identificam-se os parâmetros  $p$  e  $q$  através da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

#### Estimação

Uma vez que um modelo foi identificado, a próxima etapa da metodologia de Box - Jenkins é estimar os parâmetros desconhecidos do modelo (para qualquer forma AR,MA ou ARIMA). Para estimar tais parâmetros utiliza-se o método dos mínimos quadrados.

#### Verificação

O último passo da metodologia de Box -Jenkins consiste em verificar se o modelo representa adequadamente a série temporal observada. Esta verificação consiste em analisar a amostra de autocorrelação da série de resíduos. Para o modelo ser considerado adequado, a série de resíduos deve ser independente e aleatoriamente distribuída em torno de zero.

Se o modelo for considerado inadequado, deve-se reiniciar as três etapas novamente. Porém, se o modelo for considerado adequado, realizam-se as previsões bem como a obtenção de seu intervalo de confiança.



### 3.5 Medidas de desempenho

Existem várias medidas de avaliação do desempenho das séries temporais. As mais utilizadas serão aqui destacadas e, posteriormente, serão empregadas na comparação dos resultados das previsões deste trabalho.

#### RMSE - Raiz do Erro Médio Quadrático (Root Mean Square Error)

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2} \quad (37)$$

#### MAE - Erro Médio Absoluto (Mean Absolute Error)

$$\sum_{i=1}^n \frac{|Y_i - \widehat{Y}_i|}{n} \quad (38)$$

#### MPE – Erro Percentual Médio (Mean Percentage Error)

$$\frac{1}{n} \sum_{i=1}^n \frac{Y_i - \widehat{Y}_{i-1}}{Y_i} \quad (39)$$

#### MAPE – Erro Percentual Médio Absoluto (Mean Absolute Percentage Error)

$$\frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \widehat{Y}_{i-1}}{Y_i} \right| \quad (40)$$

## 4. Teoria de Filas

Para Abensur *et al.* (2003), fila é consequência de um descompasso entre a capacidade de atendimento do serviço oferecido e a demanda de seus usuários. Embora seja geralmente associada a um efeito maléfico, a fila representa um importante papel na gestão de operações de serviços, pois em muitos casos seria impraticável a oferta de uma infra-estrutura suficiente para atender a procura de todos os usuários.

A teoria das filas é um ramo da pesquisa operacional que estuda a formação de filas, através de análises matemáticas precisas e propriedades mensuráveis das filas. Ela provê modelos para demonstrar previamente o comportamento de um sistema, tornando possível dimensioná-lo de forma a satisfazer os clientes e ser viável economicamente para o provedor do serviço, evitando desperdícios e gargalos, como observa Gonçalves (2010).

Uma das notações usadas para a especificação de filas é a feita por Kendall, definida como  $(A/S/m/K/N/Z)$ , na qual  $A$  será a distribuição do tempo entre chegadas,  $S$  é a distribuição do tempo de serviço,  $m$  é o número de servidores,  $K$  é a capacidade do sistema,  $N$  é o tamanho da população e  $Z$  é a disciplina de atendimento.

Ao adaptar esta notação para o presente estudo,  $A$  representa a distribuição do tempo entre chegadas de paciente na Unidade de Terapia Intensiva,  $S$  será a distribuição do tempo de serviço (ocupação dos leitos) na Unidade de Terapia Intensiva,  $m$  será o número de leitos de UTI,  $K$  será a capacidade da fila de espera (infinito),  $N$  será o número potencial de pacientes que podem chegar (infinito) e  $Z$  obedecerá a ordem (FCFS), ou seja, o primeiro a chegar será primeiro a ser atendido. Após descrita a notação, pode-se dar sequência aos modelos de filas. O modelo a ser considerados neste estudo, será descrito a seguir baseado em Hillier & Lieberman (2010).

### 4.1 Modelo M/M/s

O modelo M/M/s parte dos seguintes pressupostos:

- Todos os tempos entre atendimento são distribuídos de forma independente e idêntica de acordo com uma distribuição exponencial, ou seja, o processo de entrada é de Poisson,
- Todos os tempos de atendimento também são distribuídos de forma independente e idêntica de acordo com uma distribuição exponencial,
- O número de atendentes de atendentes é igual a  $s$ .

Consequentemente, este modelo é o caso especial do processo de nascimento-e-morte na qual a taxa média de chegada e a taxa média de atendimento por atendente ocupado do sistema de filas são constantes,  $\lambda$  e  $\mu$  respectivamente, independente do estado do sistema. O processo de nascimento-e-morte é um tipo especial de cadeia de Markov em tempo contínuo.

No contexto da teoria de filas e especialmente deste estudo, o termo nascimento corresponde à chegada de um paciente no sistema de filas, e a morte corresponde à partida do paciente atendido.

Se o sistema possuir apenas um único atendente ( $s=1$ ), os parâmetros para o processo de nascimento-e-morte são  $\lambda_n=\lambda$  ( $n=0,1,2,3\dots$ ) e  $\mu_n=\mu$  ( $n=0,1,2,3\dots$ ). Entretanto, se o sistema for composto por vários atendentes ( $s > 1$ ) o  $\mu_n$  não pode ser expresso desta forma.

Sendo que  $\mu_n$  representa a taxa média de términos de atendimento para o sistema de filas global quando existem  $n$  clientes atualmente no sistema. Quando a taxa média de atendimento por atendente ocupado é  $\mu$ , a taxa média de términos de atendimento global para  $n$  atendentes ocupados é  $n\mu$ . Logo  $\mu_n = n\mu$ , quando  $n \leq s$ , ao passo que  $\mu_n = s\mu$  quando  $n \geq s$ , de modo que todos os  $s$  atendentes estejam ocupados.

Quando  $s\mu$  excede a taxa média de chegada  $\lambda$ , isto é, o fator de utilização  $\rho = \lambda/s\mu < 1$ , um sistema de filas que se ajusta a esse modelo atinge uma condição de estado estável, na qual a distribuição probabilística do estado do sistema permanece a mesma (a distribuição do estado estável ou estacionária) ao longo do tempo. Porém se  $\rho \geq 1$  o número médio de chegadas no sistema excede a taxa média de atendimento do sistema, de forma que a fila “explode” e cresce sem limites, impedindo qualquer situação de equilíbrio.

Os resultados de estado estável prontamente disponíveis para esse modelo geral são os seguintes:

$$P_0 = 1 / \left[ \sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} * \frac{1}{1-\lambda/(s\mu)} \right] \quad (41)$$

$$P_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} * P_0, & \text{se } 0 \leq n \leq s \\ \frac{(\lambda/\mu)^n}{s!s^{n-s}} * P_0, & \text{se } n \geq s \end{cases} \quad (42)$$

$$L_q = \frac{P_0(\lambda/\mu)^s \rho}{s!(1-\rho)^2} \quad (43)$$

$$W_q = \frac{L_q}{\lambda} \quad (44)$$

$$W = W_q + \frac{1}{\mu} \quad (45)$$

$$L = L_q + \frac{\lambda}{\mu} \quad (46)$$

$$P\{W_q > t\} = (1 - P\{W_q = 0\})e^{-s\mu(1-\rho)t} \quad (47)$$

$$P\{W_q = 0\} = \sum_{n=0}^{s-1} P_n \quad (48)$$

em que  $P_n$  é a probabilidade de exatamente  $n$  pacientes se encontrarem no sistema de filas;  $L_q$  é o comprimento esperado da fila (exclui pacientes que estão sendo atendidos);  $L$  é o número de pacientes esperado no sistema de filas;  $W_q$  é o tempo de espera na fila (exclui o tempo de atendimento) para cada paciente individual e  $W$  é o tempo (incluindo o tempo de atendimento) médio de espera no sistema para cada paciente individual.

Para maiores detalhes sobre teoria de filas e seus diferentes sistemas de entrada, saída e atendimento, recomenda-se Bose (2001), Kleinrock (1975) e Takagi (1991).

## 5. Aplicação dos Modelos de Previsão e Análises dos Resultados

Para a realização de todas as previsões deste estudo utilizou-se o *software* estatístico R (“R Development Core Team,” 2012), especificamente o pacote *forecast* e suas seguintes funções: *HoltWinters*, *ets*, *auto.arima*. Utilizou-se a função *HoltWinters* para obter previsões para os seguintes métodos de amortecimento exponencial: Holt-Winters aditivo e multiplicativo. A função *ets* foi utilizada para obter a previsão para o método de tendência amortecida do modelo de amortecimento exponencial. E a função *auto.arima* que seleciona automaticamente o melhor modelo da metodologia de Box-Jenkins foi utilizada para este fim.

Nas previsões deste estudo realizaram-se duas análises. Na primeira, considerou-se a série original e realizou a respectiva previsão para os próximos 30 dias e 12 meses. Para a segunda análise, dividiu-se a série original em duas partes, numa proporção aproximada de 70% e 30%. Consideraram-se as primeiras observações (70%) para estimar os modelos, e as restantes (30%) para analisar o desempenho de previsão de modelos.

### 5.1 Análise 1

Na Figura 3, estão representados os dados referentes ao número de solicitações diárias por leitos de UTI-adulto dos hospitais regidos pela CER do estado Rio de Janeiro, nos anos de 2010 e 2011. Para facilitar o entendimento do estudo esta série será chamada de Série Temporal 1.

Na Figura 4 segue a decomposição da Série Temporal 1, na Figura 5 a Autocorrelação da Série Temporal 1 e na Figura 6 a Autocorrelação Parcial da mesma série.

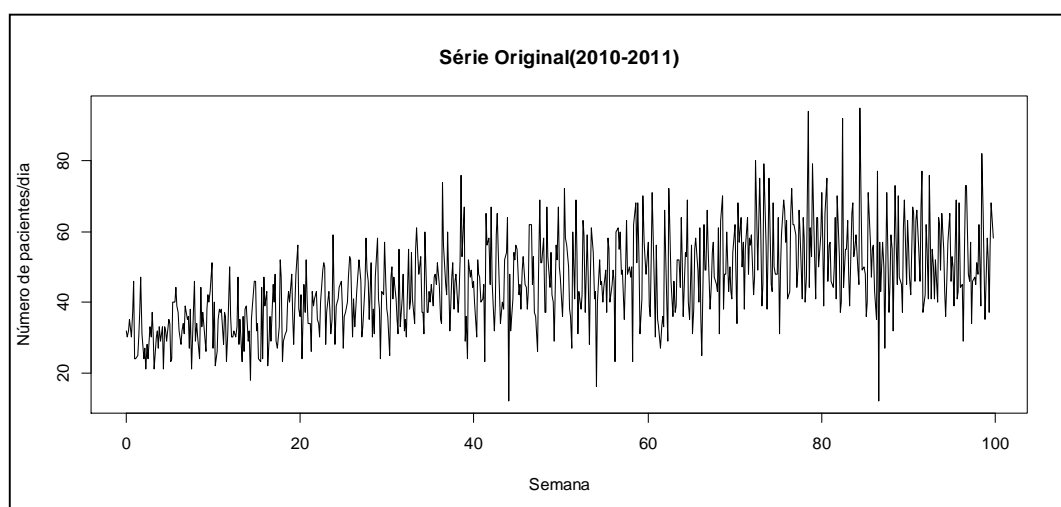


Figura 3 - Série temporal de solicitações diárias por leitos de UTI dos hospitais regidos pela CER do Estado do Rio de Janeiro, no período de 2010 e 2011 (Série Temporal 1)

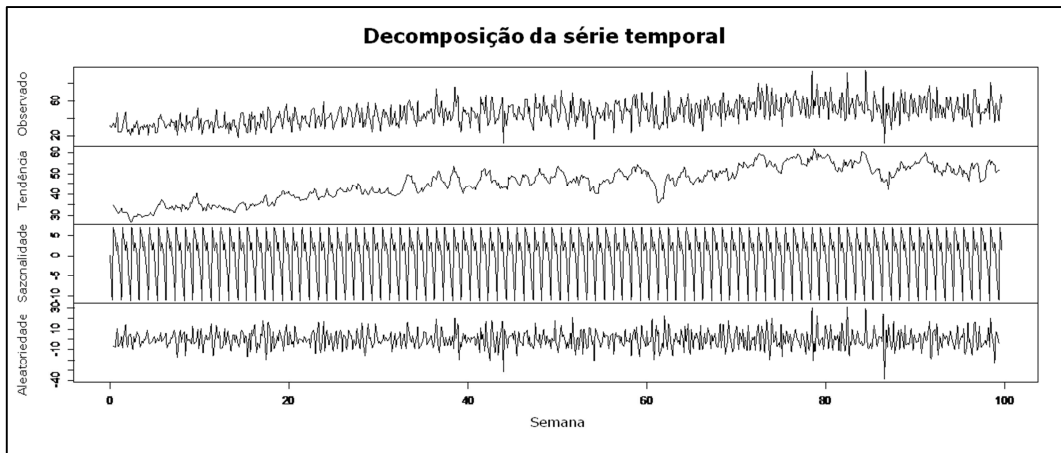


Figura 4- Decomposição da Série Temporal 1

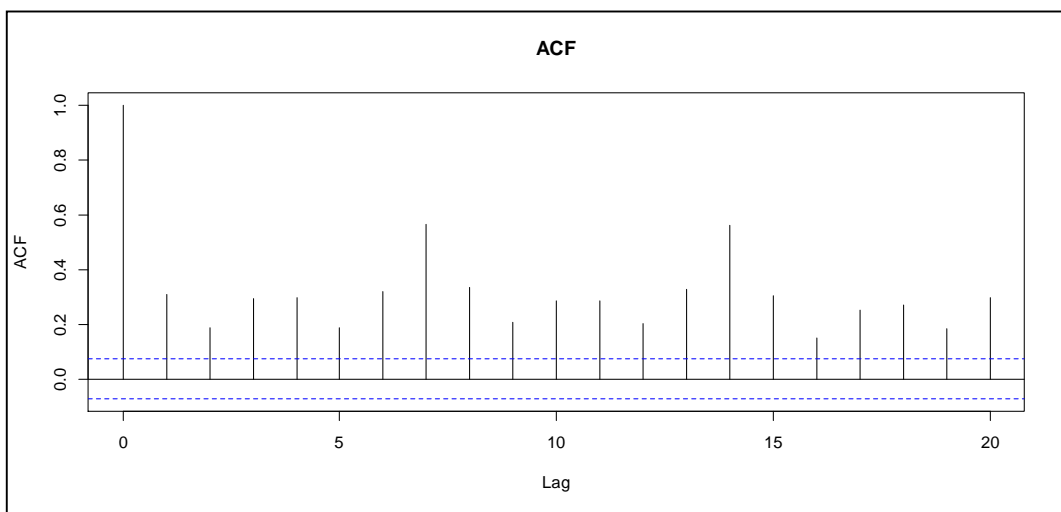


Figura 5 - Autocorrelação da Série Temporal 1

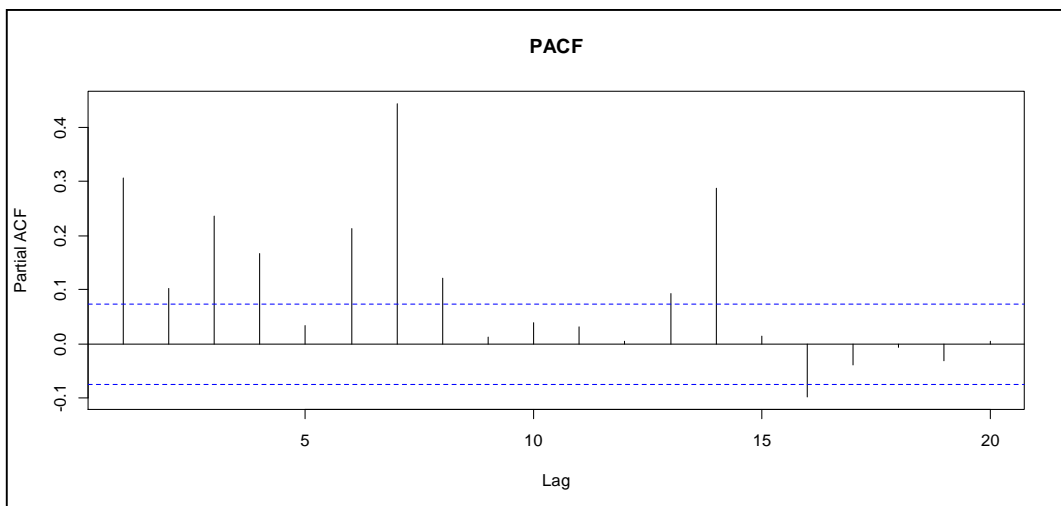


Figura 6 – Autocorrelação Parcial da Série Temporal 1

Para a série temporal 1, a função ets selecionou o método de tendência amortecida multiplicativo (M,Md,M) como o mais apropriado, enquanto a função auto.arima selecionou o

modelo SARIMA(0,1,1)(2,0,1)[7]. Com a função HoltWinters obteve-se previsões para o Holt-Winters aditivo e multiplicativo.

Os resultados das medidas de desempenho a partir dos modelos estimados estão representados na Tabela 6. Assim, é possível a comparação dos resultados e a seleção do melhor modelo para a série temporal 1.

		MEDIDAS DE DESEMPENHO			
		RMSE	MAE	MPE	MAPE
Modelos	Holt-Winters multiplicativo	9,22	7,12	-3,54	16,94
	Holt-Winters aditivo	9,26	7,16	-3,39	17,06
	Tendência Amortecida Multiplicativa - $MM_dM$	<b>8,97</b>	<b>6,83</b>	-2,88	<b>16,28</b>
	Sarima (0,1,1)(2,0,1)[7]	9,00	6,90	<b>-4,41</b>	16,74

Tabela 6–Comparação entre os modelos estimados na série 1

Após a análise da Tabela 6, pode-se concluir que, por suas medidas de desempenho apresentarem os menores valores, o modelo de Tendência Amortecida Multiplicativa foi o que apresentou os melhores resultados, seguido respectivamente pelos modelos: SARIMA, Holt-Winters Multiplicativo e Holt-Winters Aditivo.

A Figura 7 apresenta o modelo de melhor desempenho, isto é, o modelo de Tendência Amortecida Multiplicativa. A Figura 8 apresenta a previsão deste modelo para 30 dias e a Figura 9 a previsão deste modelo para 1 ano.

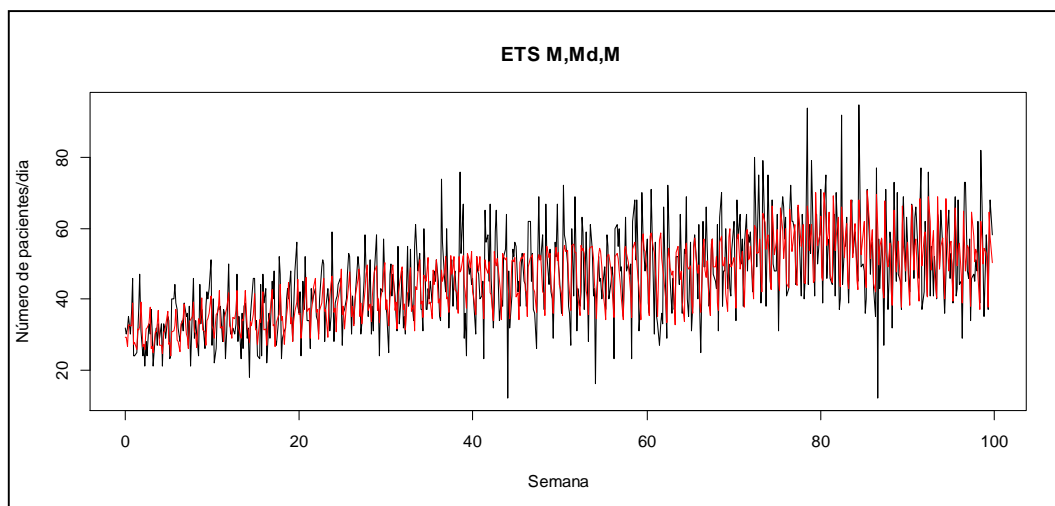


Figura 7 - Modelo de Tendência Amortecida Multiplicativa ( $MM_dM$ )

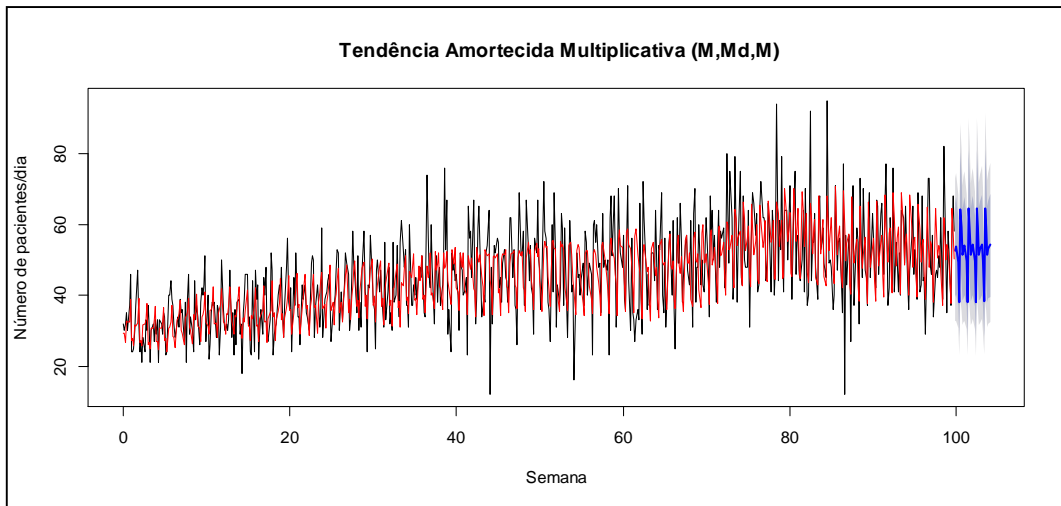


Figura 8 - Previsão de 30 dias do modelo de Tendência Amortecida Multiplicativa (M,Md,M)

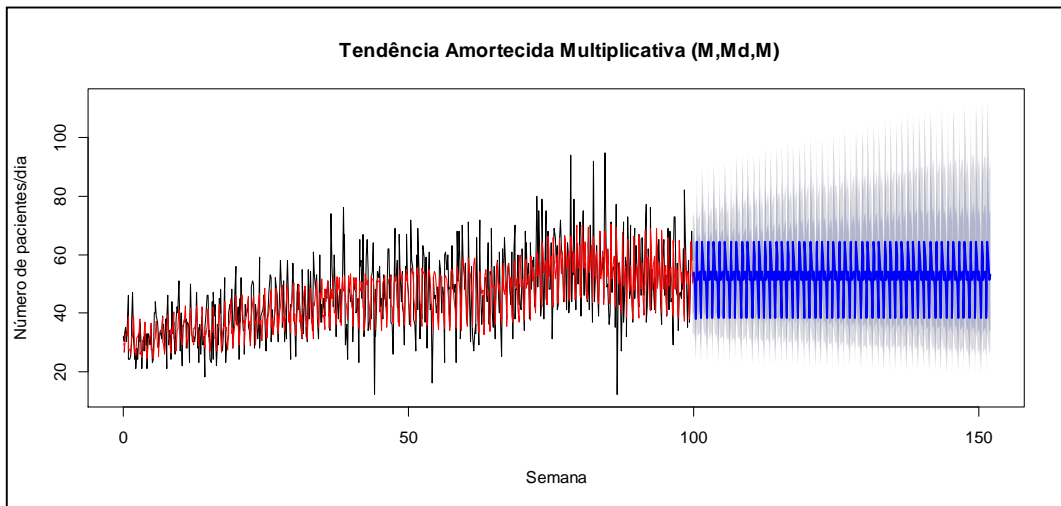


Figura 9 - Previsão de 1 ano do modelo de Tendência Amortecida Multiplicativa (M,Md,M)

Já as figuras 10,11 e 12 apresentam as previsões de 1 ano dos modelos restantes, conforme seqüência decrescente de desempenho apresentado.



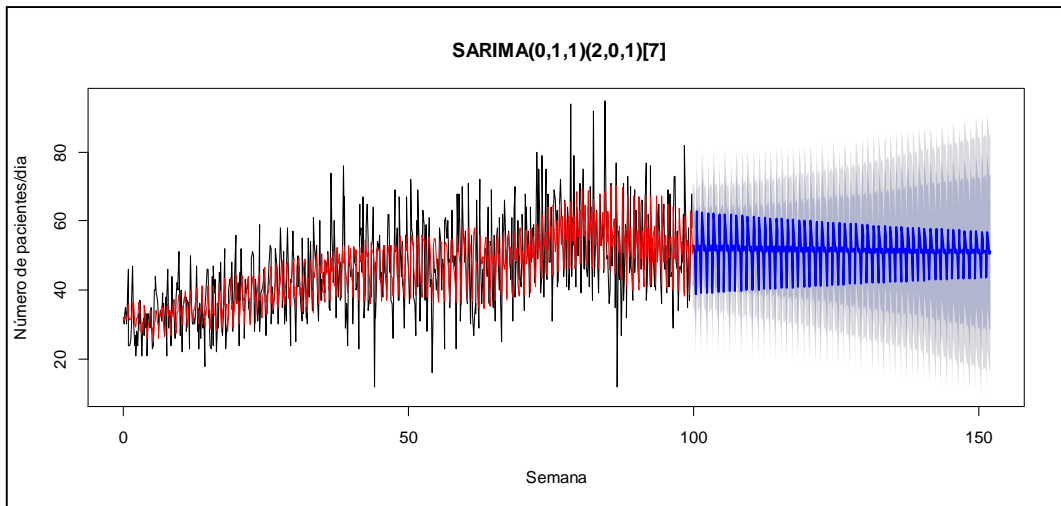


Figura 10 - Previsão de 1 ano do modelo SARIMA

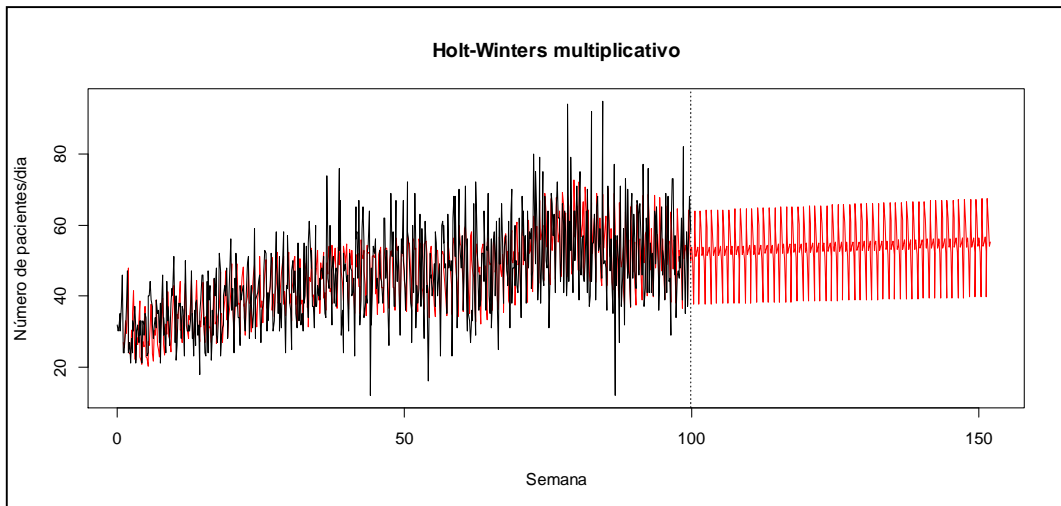


Figura 11 - Previsão de 1 ano do modelo Holt-Winters multiplicativo

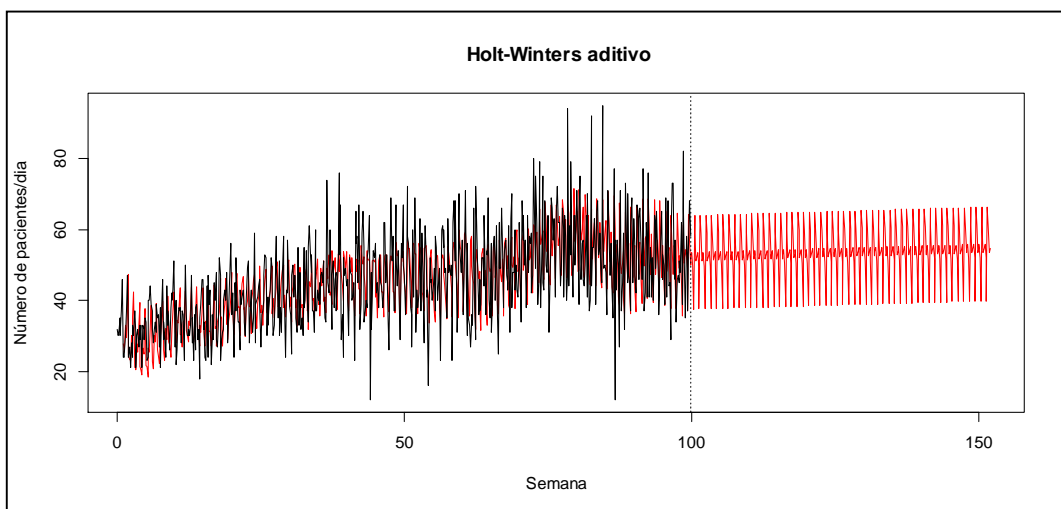


Figura 12 - Previsão de 1 ano do modelo Holt-Winters aditivo

## 5.2 Análise 2

Conforme mencionado acima, na segunda análise, dividiu-se a série original em duas partes, numa proporção aproximada de 70% e 30%. Na Figura 13, estão representados as primeiras observações (70%) que se referem aos dezesseis primeiros meses da série original. Para facilitar o entendimento do estudo esta série será chamada de Série Temporal 2.

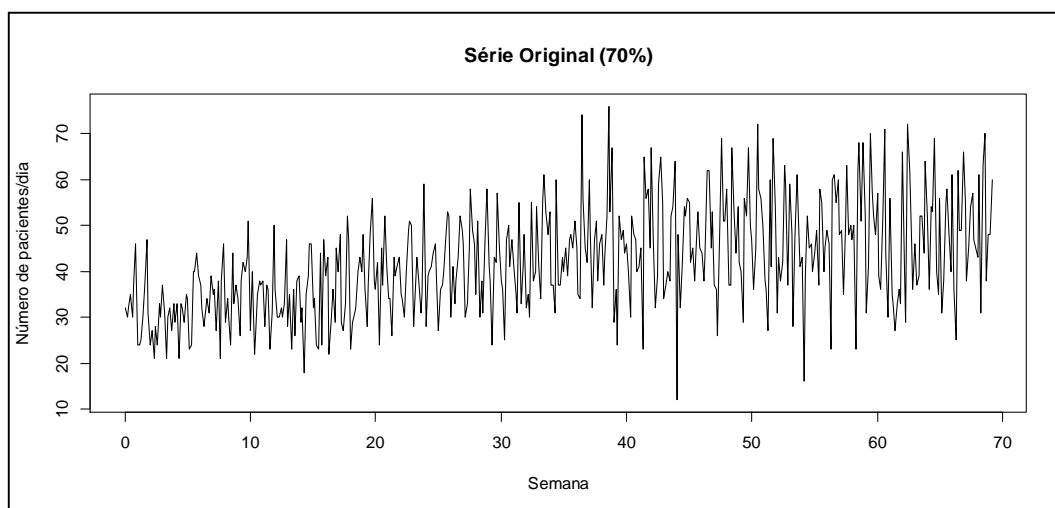


Figura 13 - Série Temporal de solicitações diárias por leitos de UTI dos hospitais regidos pela CER do Estado do Rio de Janeiro, no período de Janeiro de 2010 a Abril de 2011 (Série Temporal 2)

A partir da Série Temporal 2 realizou-se previsões para os próximos oito meses. Esses dados previstos foram comparados com os dados reais separados anteriormente (30%) e os resultados desta comparação foram obtidos por meio das medidas de desempenho utilizadas até então, as quais estão representadas na Tabela 7.

Para a série temporal 2, a função ets selecionou o método de tendência amortecida multiplicativo ( $M, A_d, A$ ) como o mais apropriado, enquanto a função auto.arima selecionou o modelo SARIMA(1,1,1)(2,0,1)[7]. Com a função HoltWinters obteve-se previsões para o amortecimento exponencial simples (SES), Holt-Winters aditivo e multiplicativo.

		MEDIDAS DE DESEMPENHO			
		RMSE	MAE	MPE	MAPE
Modelos	Holt-Winters multiplicativo	12,37	10,06	<b>-0,16</b>	22,36
	Holt-Winters aditivo	11,73	9,64	-0,14	21,56
	Tendência Amortecida Multiplicativa - $MM_dM$	<b>10,44</b>	<b>8,57</b>	0,03	<b>16,64</b>
	Sarima (0,1,1)(2,0,1)[7]	10,46	8,60	0,007	16,98

Tabela 7 - Medidas de desempenho (dados estimados x dados reais)

Após a análise da Tabela 7, pode-se concluir que na comparação dos dados reais com os estimados, o modelo de Tendência Amortecida Multiplicativa também apresentou os melhores

resultados, seguido respectivamente pelos modelos: SARIMA, Holt-Winters aditivo e Holt-Winters multiplicativo.

A Figura 14 apresenta a comparação dos dados reais com as previsões do modelo de melhor desempenho, isto é, o modelo de Tendência Amortecida Multiplicativa.

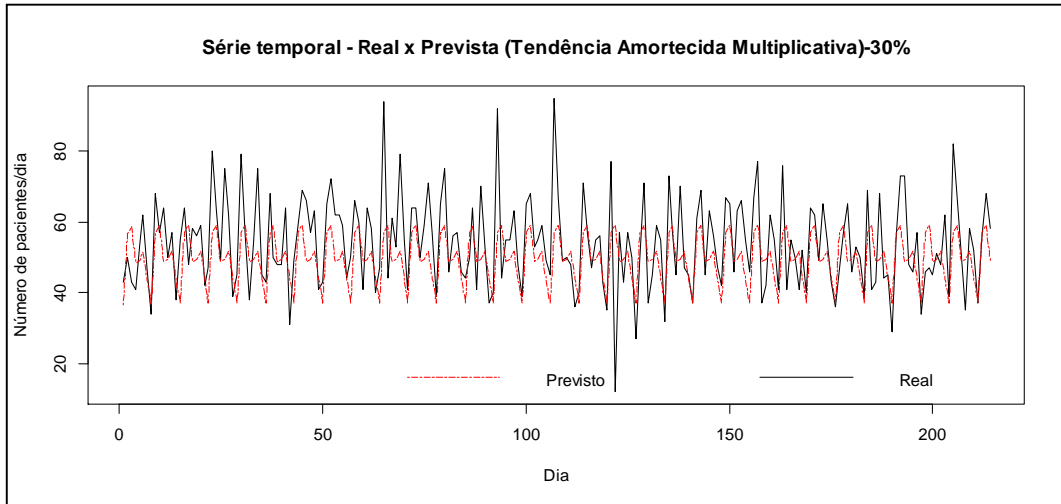


Figura 14- Previsão do modelo Tendência Amortecida Multiplicativa x dados reais

Já as figuras 15,16 e 17 apresentam a comparação dos dados reais com as previsões dos modelos restantes, conforme sequência decrescente de desempenho apresentado.

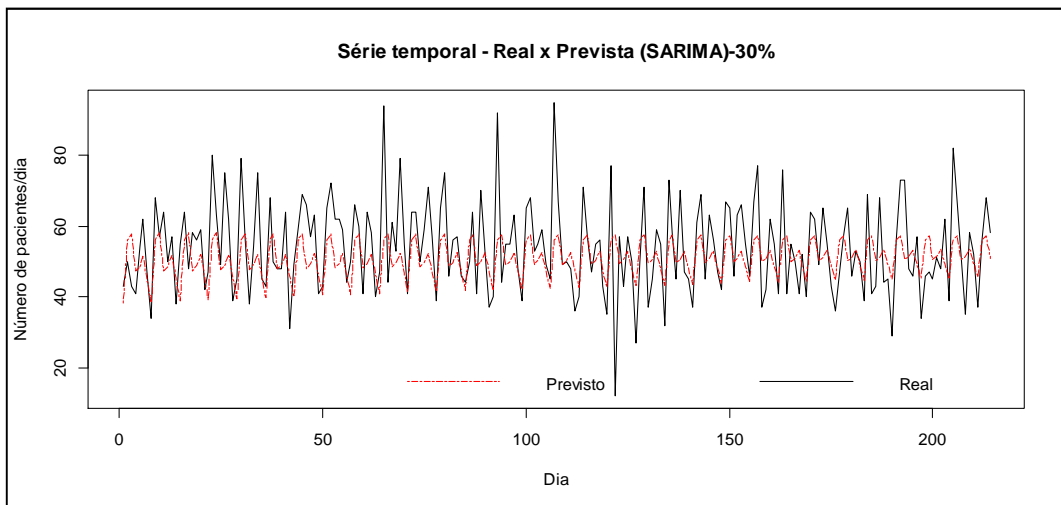


Figura 15 -Previsão do modelo SARIMA x dados reais

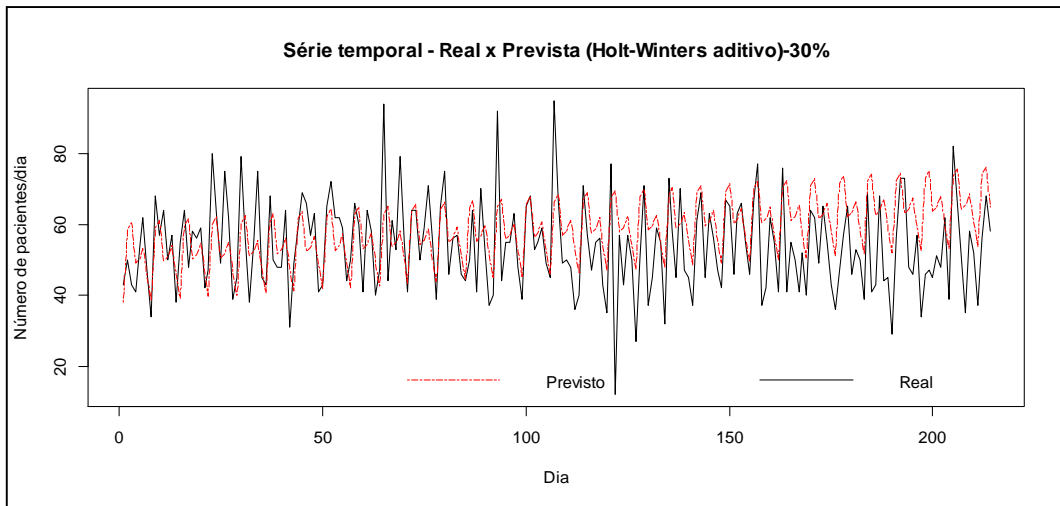


Figura 16- Previsão do modelo Holt-Winters aditivo x dados reais

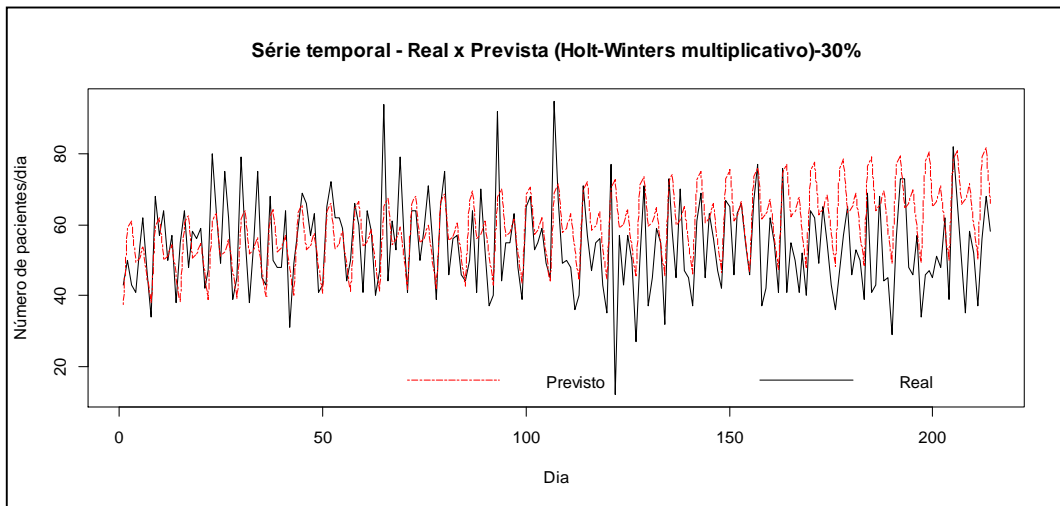


Figura 17 - Previsão do modelo Holt-Winters multiplicativo x dados reais

### 5.3 Conclusões das análises

Com a mesma série foram feitas duas análises, sendo que a segunda serviu para reafirmar a primeira, pois os resultados foram bem semelhantes. Nestas duas análises o modelo Tendência Amortecida Multiplicativa apresentou os melhores resultados, seguido pelo modelo SARIMA que também obteve um bom desempenho. Já os modelos de Holt-Winters se alternaram, ou seja, o modelo de Holt-Winters aditivo apresentou o pior resultado da análise 1 e o modelo de Holt-Winters multiplicativo apresentou o pior resultado da análise 2.

Percebe-se, portanto, que o modelo Tendência Amortecida Multiplicativa ao apresentar os melhores resultados é condizente a Gardner e Mckenzie (2010) que afirmam que o método de Tendência Amortecida se destaca por sua precisão nas previsões, tornando-o difícil de ser superado por outros métodos. Este método também se posiciona entre os cinco melhores em termos de erros em sub amostras das competições M's de Makridakis (1979, 1983, 1993,

2000). Assim, o modelo Tendência Amortecida Multiplicativa foi o selecionado neste estudo para realizar as previsões de demanda, as quais servirão de entrada para o modelo de teoria de filas.

## 6. Aplicação de Teoria de Filas e Análises dos Resultados

Conforme descrito anteriormente, para definir o modelo de teoria de filas mais adequado a um determinado problema é necessário identificar a distribuição do tempo entre chegadas e a do tempo de serviço que para este estudo são, respectivamente, a distribuição do tempo entre chegadas de pacientes na UTI - adulto e a distribuição do tempo de utilização destes leitos.

Assim, a seguir serão apresentados quatro cenários com características próprias, para os quais serão identificados as distribuições, os modelos e os resultados. Ressaltando que neste estudo admite-se a existência de uma central de regulação, assim os pacientes permanecem em um único ponto de atendimento e de lá são roteados para onde tem vaga.

Também é importante ressaltar que a taxa média de chegada será obtida através da previsão de demanda obtida pelo modelo Tendência Amortecimento Multiplicativa, logo como na análise 1 foram realizadas previsões para trinta dias e um ano, os quatro cenários serão compostos por taxa média de chegada igual a média dos últimos 7 dias da previsão de 30 dias e um ano. Considerou-se apenas os últimos 7 dias por serem os dados mais recentes da previsão e pelo fato da série ter sazonalidade de 7 dias.

Na análise 2 do capítulo 7 houve a divisão da série temporal dos dados reais em duas partes, uma de aproximadamente 70% e outra de 30%. Baseando-se apenas na parcial de 70% da série realizou-se a previsão para o seu restante de 30%. Ao final, comparou esta previsão com os dados reais correspondentes a sua parcial de 30% e constatou-se que o modelo Tendência Amortecimento Multiplicativa também apresentou o melhor resultado. Assim, os quatro cenários também serão compostos por taxa média de chegada baseada na média dos últimos 7 dias da previsão fora da amostra (30%) e na média dos dados reais da parte fora da amostra (30%).

Para cada cenário será apresentado um gráfico contendo o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades correspondentes aos intervalos de confiança de 95%, 96%, 97%, 98%, 99% e 99,9%. Assim, será possível relacionar graficamente a quantidade de leitos e o respectivo tempo de espera de acordo com o intervalo de confiança adotado para todos os cenários, bastando para tanto traçar as retas a partir da intersecção da quantidade de leito e tempo de espera na curva do intervalo de confiança. Entretanto, este estudo se aterá ao número mínimo de leitos que possibilite a estabilidade do sistema, uma espera na fila de no máximo 30 minutos, com 95% de probabilidade e uma espera na fila de no máximo 1 minuto, com 99,9% de probabilidade.

Os resultados obtidos nos cenários propostos para a previsão de 30 dias e 1 ano serão, primeiramente, comparados ao cenário atual. E, posteriormente, será realizada a comparação

entre número de leitos/1000 habitantes dos cenários propostos e atual com os parâmetros estabelecidos conforme a portaria nº 1.101 do Ministério da Saúde (2002).

O número de leitos obtidos nos cenários propostos para a previsão fora da amostra (30%) e a parte fora da amostra dos dados reais (30%) também serão comparados entre si.

## 6.1 Cenário I

O intervalo, em minutos, entre as solicitações por leitos de UTI-adulto formalizadas à CER será considerado o tempo entre chegadas, o qual está representado na Figura 18. Já o tempo de serviço, o qual esta representado na Figura 19, refere-se ao tempo diário de utilização dos leitos ora solicitados que foram obtidos através do DATASUS (2013).

### 6.1.1 Distribuições

As Figuras 18 e 19 sugerem que os dados sobre o tempo entre chegadas e o tempo de serviço de serviço provêm de uma distribuição exponencial, ou seja, a função de distribuição exponencial melhor se ajusta aos dados de interesse.

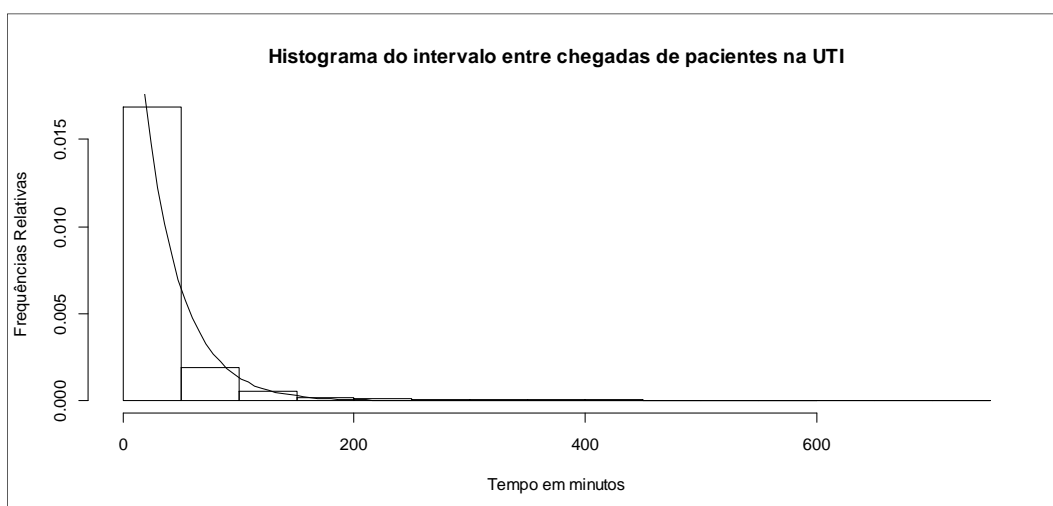


Figura 18 - Histograma do tempo entre chegadas (2010 – 2011)

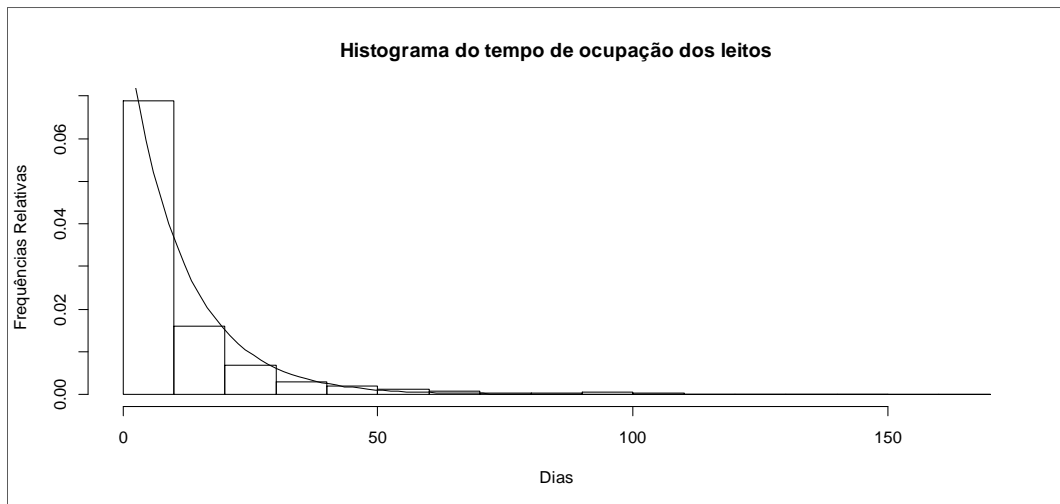


Figura 19- Histograma do tempo de serviço (Hospitais CER)

### 6.1.2 Modelo

Ao identificar que o tempo entre chegadas e o tempo de serviço são distribuídos de acordo com uma distribuição exponencial e que  $s > 1$ , o modelo M/M/s aparece como o mais apropriado. Assim, dentre as equações deste modelo, neste estudo serão aplicadas as equações 40,41,46,47, que foram descritas no capítulo 4.

### 6.1.3 Resultados

#### Resultados para previsão de 30 dias

Para a previsão de 30 dias, o número de solicitações diárias de pacientes por leitos de UTI-adulto dos hospitais regidos pela CER é de 52,9396, o intervalo médio entre estas solicitações é de 0,453347 hora e o  $\lambda = 2,205817$  pacientes/hora. E, de acordo com o DATASUS (2013), para os trinta primeiros dias de 2012, o tempo médio de utilização dos leitos dos hospitais regidos pela CER é de 10,01 dias, então  $\mu = 0,004163$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 20.



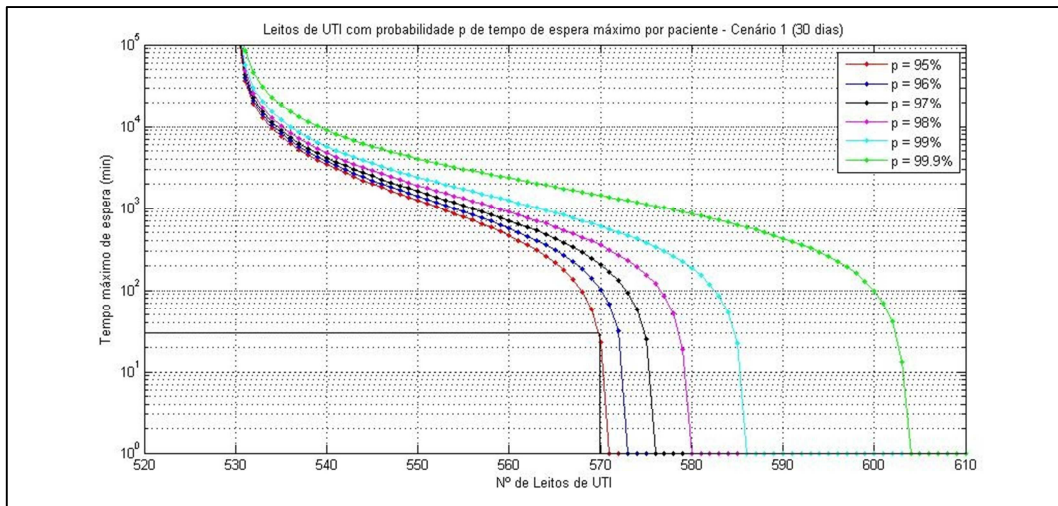


Figura 20 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário I (previsão de 30 dias)

Ao analisar a Figura 20, percebe-se que neste cenário são necessários:

- 530 leitos de UTI para o sistema ser estável.
- 569 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 604 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### Resultados para previsão de 1 ano

Para a previsão de 1 ano, o número de solicitações diárias de pacientes por leitos de UTI-adulto dos hospitais regidos pela CER é de 52,7569 o intervalo médio entre estas solicitações é de 0,454917 hora e o  $\lambda = 2,198204$  pacientes/hora. E, de acordo com o DATASUS(2013), para o ano de 2012, o tempo médio de utilização dos leitos dos hospitais regidos pela CER é de 11,254 dias, então  $\mu = 0,003702$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 21.

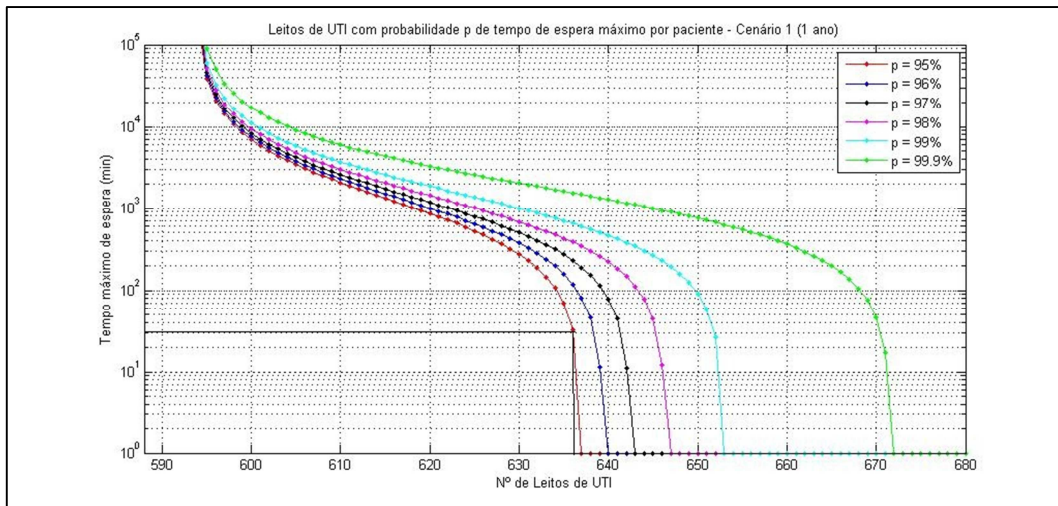


Figura 21 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário I (previsão de 1 ano)

Ao analisar a Figura 21, percebe-se que neste cenário são necessários:

- 594 leitos de UTI para o sistema ser estável.
- 636 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 672 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

**Resultados para previsão fora da amostra (30%)**

Para a previsão fora da amostra (30%), o número de solicitações diárias de pacientes por leitos de UTI-adulto dos hospitais regidos pela CER é de 49,56 o intervalo médio entre estas solicitações é de 0,484262 hora e o  $\lambda = 2,065$  pacientes/hora. E, de acordo com o DATASUS (2013), para os últimos 8 meses de 2011, o tempo médio de utilização dos leitos dos hospitais regidos pela CER é de 13,162 dias, então  $\mu = 0,003166$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 22.

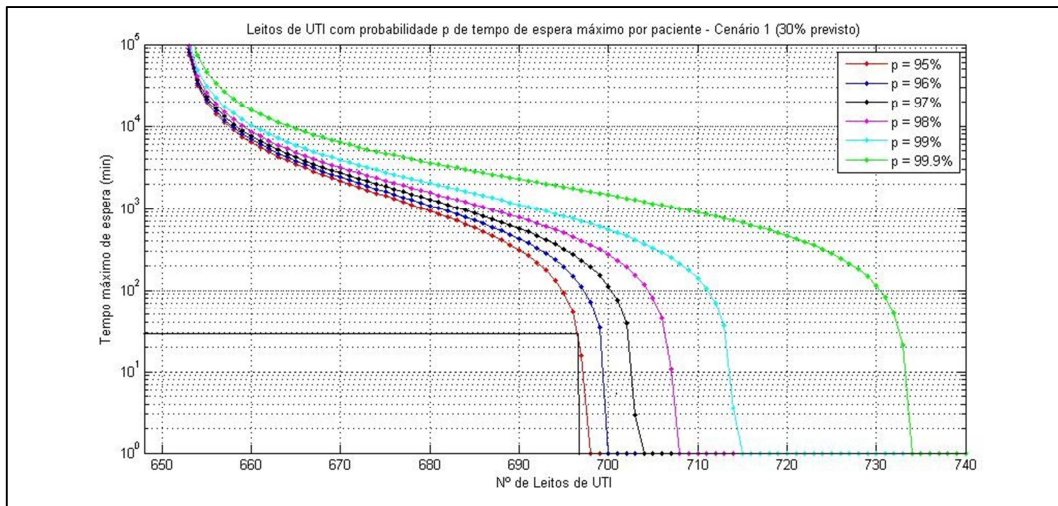


Figura 22 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário I (previsão fora da amostra - 30%)

Ao analisar a Figura 22, percebe-se que neste cenário são necessários:

- 653 leitos de UTI para o sistema ser estável.
- 696 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 734 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

**Resultados para os dados reais da parte fora da amostra (30%)**

Para os dados reais da parte fora da amostra (30%), o número de solicitações diárias de pacientes por leitos de UTI-adulto dos hospitais regidos pela CER é de 52,00 o intervalo médio entre estas solicitações é de 0,461538 hora e o  $\lambda = 2,1666$  pacientes/hora. E, de acordo com o DATASUS (2013), para os últimos 8 meses de 2011, o tempo médio de utilização dos leitos dos hospitais regidos pela CER é de 13,162 dias, então  $\mu = 0,003166$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 23.

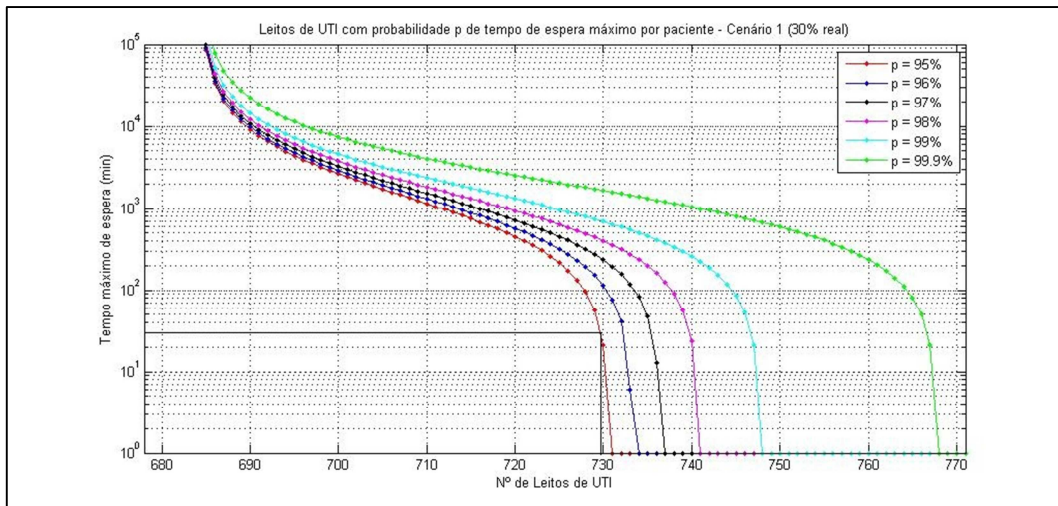


Figura 23 - Leitos de UTI com probabilidade  $p$  de tempo de espera máximo por paciente – Cenário I (dados reais da parte fora da amostra - 30%)

Ao analisar a Figura 23, percebe-se que neste cenário são necessários:

- 685 leitos de UTI para o sistema ser estável.
- 729 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 768 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

## 6.2 Cenário II

De acordo com o DATASUS (2013), o tempo médio de utilização dos leitos dos hospitais regidos pela CER é diferente do tempo médio de todos os hospitais do estado do Rio de Janeiro. Assim, neste cenário optou-se por utilizar apenas o tempo médio referente a todos os hospitais do estado do Rio de Janeiro.

Desta forma, o tempo entre chegadas será o mesmo do cenário I, o qual está representado na Figura 18. Já o tempo de serviço, refere-se ao tempo diário de utilização dos leitos de todos os hospitais do Rio de Janeiro, conforme Figura 24.

### 6.2.1 Distribuições

Já que o tempo entre chegadas é o mesmo do cenário I, a distribuição também é a mesma, ou seja, a distribuição é exponencial.

Para o tempo de serviço, analisa-se a Figura 24, a qual sugere que os dados sobre o tempo de serviço provêm de uma distribuição exponencial, ou seja, a função de distribuição exponencial melhor se ajusta aos dados de interesse.

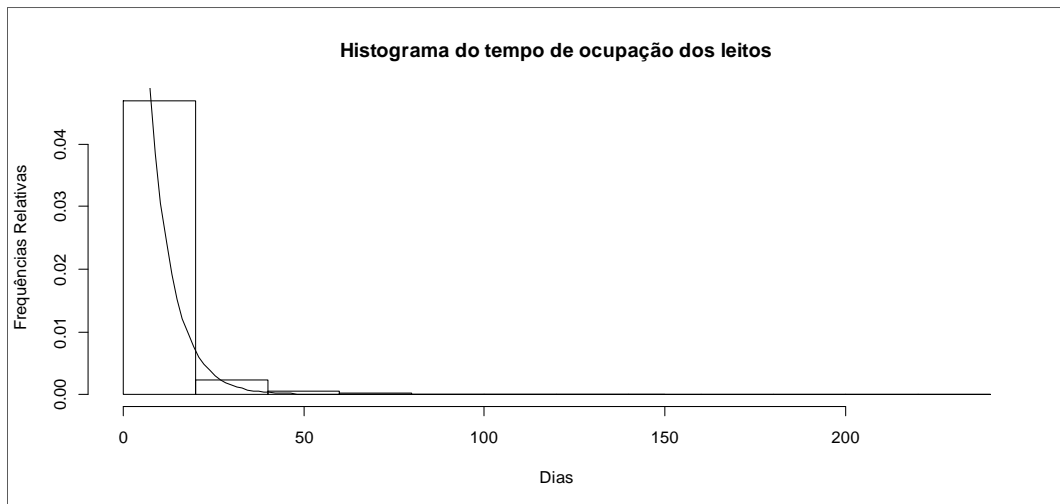


Figura 24 - Histograma do tempo de serviço (Todos hospitais do estado do RJ)

### 6.2.2 Modelo

Ao identificar que o tempo entre chegadas e o tempo de serviço são distribuídos de acordo com uma distribuição exponencial e que  $s > 1$ , o modelo M/M/s aparece novamente como o mais apropriado. Assim, os resultados posteriores também serão determinados com a utilização das equações 40,41,46,47 descritas no capítulo 4.

### 6.2.3 Resultados

#### Resultados para previsão de 30 dias

Como o tempo médio entre chegadas é o mesmo do cenário I – previsão de 30 dias, o  $\lambda$  também será, assim  $\lambda$  será igual a 2,2058 pacientes/hora. E, de acordo com o DATASUS (2013), para os trinta primeiros dias de 2012, o tempo médio de utilização dos leitos de todos os hospitais do estado do Rio de Janeiro é de 6,439 dias, então  $\mu = 0,006417$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 25.

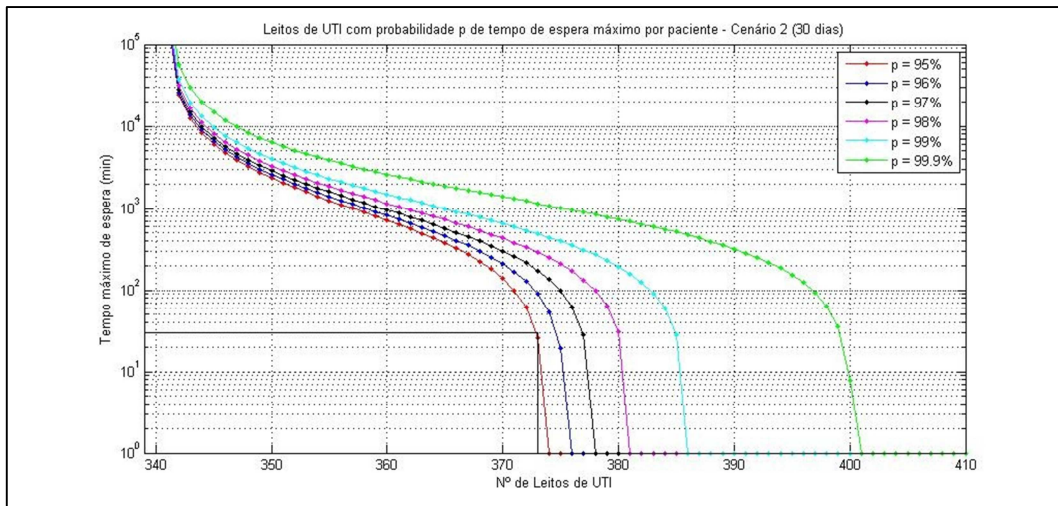


Figura 25 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (previsão de 30 dias)

Ao analisar a Figura 25, percebe-se que neste cenário são necessários:

- 341 leitos de UTI para o sistema ser estável.
- 372 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 401 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### Resultados para previsão de 1 ano

Como o tempo médio entre chegadas é o mesmo do cenário I – previsão de 1 ano, o  $\lambda$  também será, assim  $\lambda$  será igual a 2,1982 pacientes/hora. E, de acordo com o DATASUS (2013), para o ano de 2012, o tempo médio de utilização dos leitos de todos os hospitais do estado do Rio de Janeiro é de 6,468 dias, então  $\mu=0,006442$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 26.

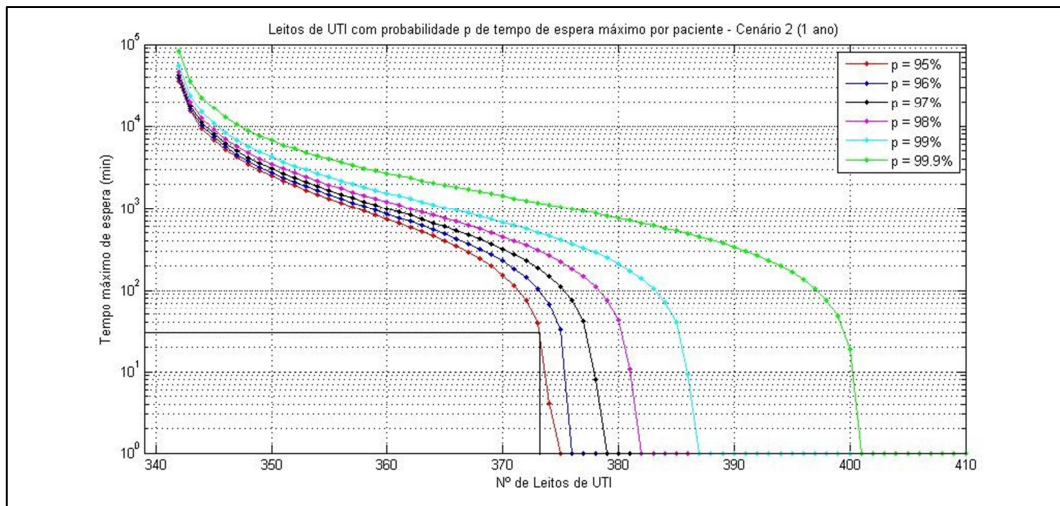


Figura 26 – Leitos de UTI com probabilidade p de tempo de espera máximo por paciente - Cenário II (previsão de 1 ano)

Ao analisar a Figura 26, percebe-se que neste cenário são necessários:

- 342 leitos de UTI para o sistema ser estável.
- 373 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 401 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### Resultados para previsão fora da amostra (30%)

Como o tempo médio entre chegadas é o mesmo do cenário I – previsão fora da amostra (30%), o  $\lambda$  também será, assim  $\lambda$  será igual a 2,065 pacientes/hora. E, de acordo com o DATASUS (2013), para os últimos 8 meses de 2011, o tempo médio de utilização dos leitos de todos os hospitais do estado do Rio de Janeiro é de 6,423 dias, então  $\mu=0,006487$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 27.

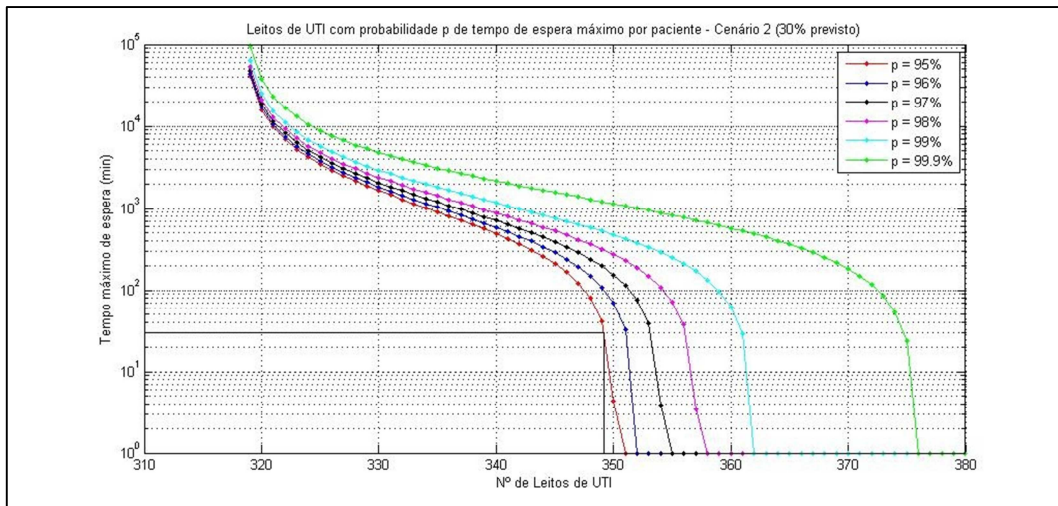


Figura 27 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (previsão fora da amostra - 30%)

Ao analisar a Figura 27, percebe-se que neste cenário são necessários:

- 319 leitos de UTI para o sistema ser estável.
- 349 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 376 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### Resultados para os dados reais da parte fora da amostra (30%)

Como o tempo médio entre chegadas é o mesmo do cenário I – dados reais da parte fora da amostra (30%), o  $\lambda$  também será, assim  $\lambda$  será igual a 2,166 pacientes/hora. E, de acordo com o DATASUS (2013), para os últimos 8 meses de 2011, o tempo médio de utilização dos leitos de todos os hospitais do estado do Rio de Janeiro é de 6,423 dias, então  $\mu=0,006487$  pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 28.



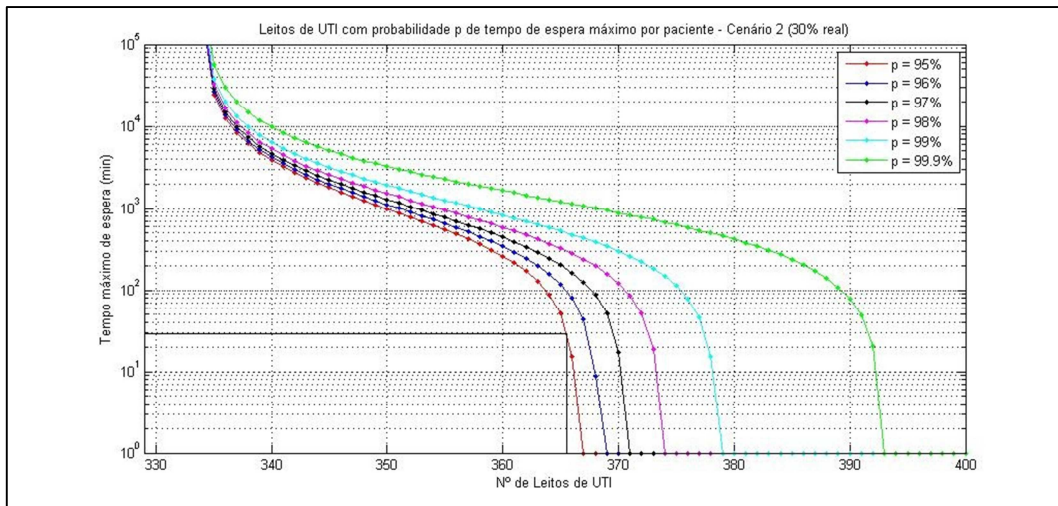


Figura 28 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (dados reais da parte fora da amostra - 30%)

Ao analisar a Figura 28, percebe-se que neste cenário são necessários:

- 334 leitos de UTI para o sistema ser estável.
- 365 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 393 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### 6.3 Cenário III

De acordo com a Figura 29, obtida através do banco de dados da CER, há 55% de desistência das solicitações por leitos de UTI-adulto formalizadas à CER. Segundo a SESDEC (2010), essa desistência decorre dos seguintes motivos: alta à revelia, alta ou melhora clínica, diagnóstico fora do perfil de regulação, falta de contato com a unidade após 48 horas, recusa de transferência pela família e pacientes de enfermaria. A proporção de cada motivo é representada pela Figura 30.

Assim, neste cenário, o tempo entre chegadas será o intervalo, em minutos, entre as solicitações por leitos de UTI – adulto formalizadas à CER, desconsiderando a porcentagem de desistência por motivos de alta ou melhora clínica. Já o tempo de serviço será o mesmo do cenário I, o qual está representado na Figura 19.

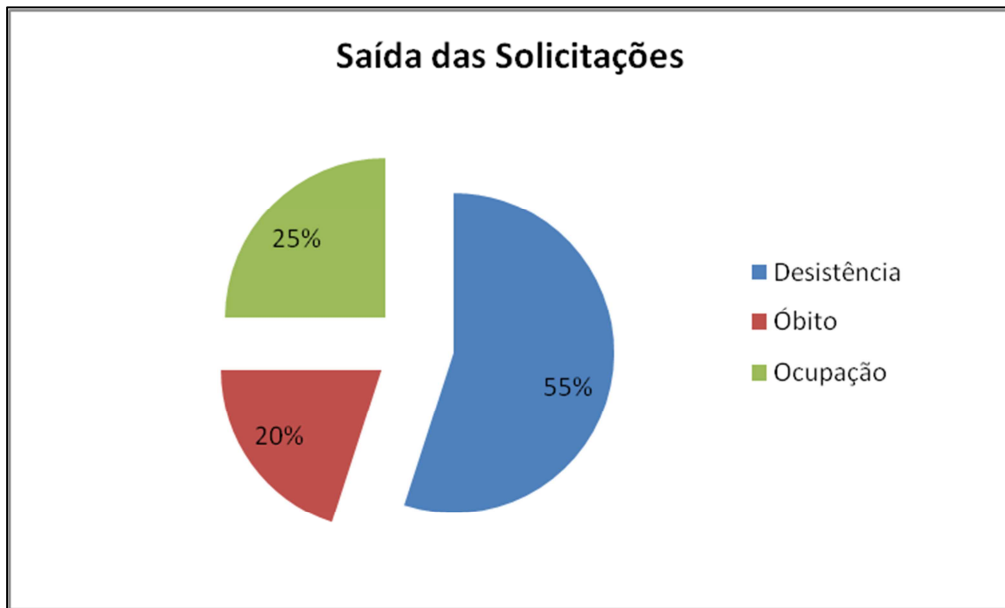


Figura 29 - Saída das solicitações

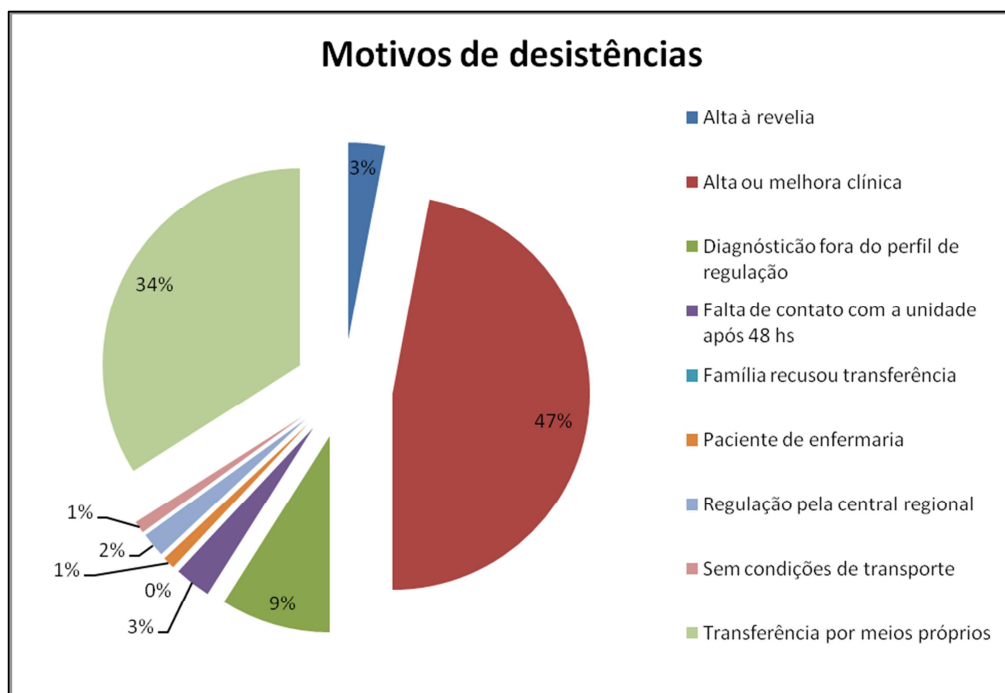


Figura 30 - Análise das desistências

### 6.3.1 Distribuições

Já que o tempo entre chegadas é o intervalo das solicitações por leitos de UTI – adulto formalizadas à CER descontando a porcentagem de desistência por alta ou melhora clínica, a distribuição será a mesma cenário I, isto é, uma exponencial.

O tempo de serviço é o mesmo do cenário I, assim a distribuição também é a mesma, ou seja, a distribuição é exponencial.

### 6.3.2 Modelo

Ao identificar que o tempo entre chegadas e o tempo de serviço são distribuídos de acordo com uma distribuição exponencial e que  $s > 1$ , o modelo M/M/s aparece novamente como o mais apropriado. Assim, os resultados posteriores também serão determinados com a utilização das equações 40,41,46,47 descritas no capítulo 4.

### 6.3.3 Resultados

#### Resultados para previsão de 30 dias

O tempo médio entre chegadas é o mesmo do cenário I – 30 dias, porém desconta-se a desistência por motivos de alta ou melhora clínica, assim  $\lambda = 2,205817 * 0,7415$ , resultando em  $\lambda = 1,635614$  pacientes/hora. Como o tempo médio de utilização dos leitos é o mesmo do cenário I – 30 dias, o  $\mu$  também será igual 0,004163 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 31.

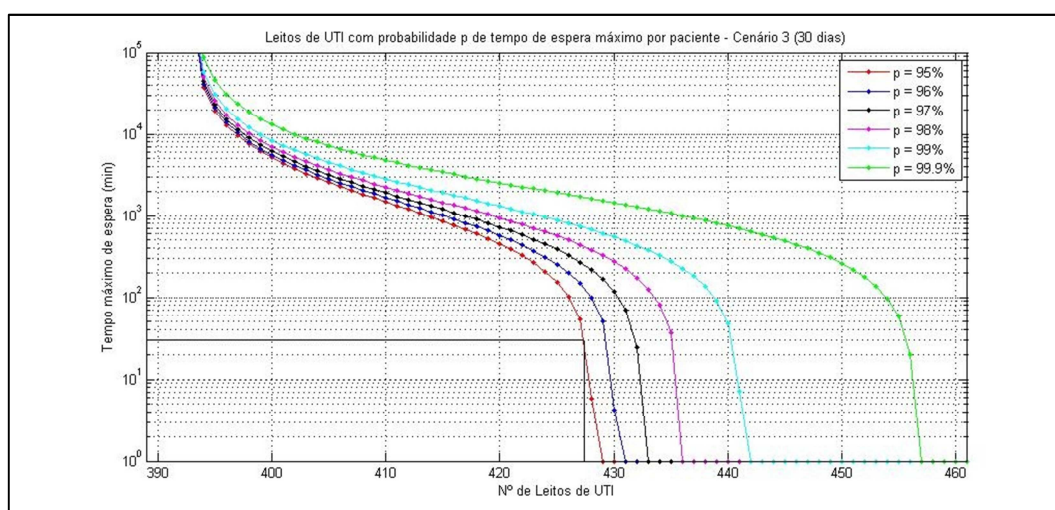


Figura 31 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão de 30 dias)

Ao analisar a Figura 31, percebe-se que neste cenário são necessários:

- 393 leitos de UTI para o sistema ser estável.
- 427 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 457 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

## Resultados para previsão de 1 ano

O tempo médio entre chegadas é o mesmo do cenário I – previsão de 1 ano, porém desconta-se a desistência por motivos de alta ou melhora clínica, assim  $\lambda = 2,198204 * 0,7415$ , resultando em  $\lambda = 1,629969$  pacientes/hora. Como o tempo médio de utilização dos leitos é o mesmo do cenário I - previsão de 1 ano, o  $\mu$  também será igual 0,003702 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 32.

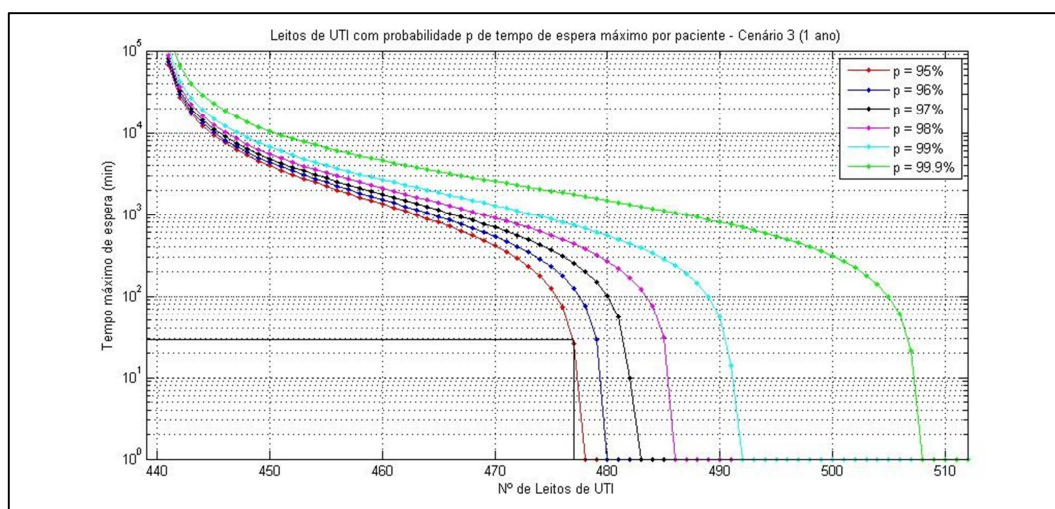


Figura 32 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão de 1 ano)

Ao analisar a Figura 32, percebe-se que neste cenário são necessários:

- 441 leitos de UTI para o sistema ser estável.
- 476 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 508 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

## Resultados para previsão fora da amostra (30%)

O tempo médio entre chegadas é o mesmo do cenário I– previsão fora da amostra (30%), porém desconta-se a desistência por motivos de alta ou melhora clínica, assim  $\lambda = 2,065 * 0,7415$ , resultando em  $\lambda = 1,531198$  pacientes/hora. Como o tempo médio de utilização dos leitos é o mesmo do cenário I - previsão fora da amostra (30%), o  $\mu$  também será igual 0,003166 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 33.

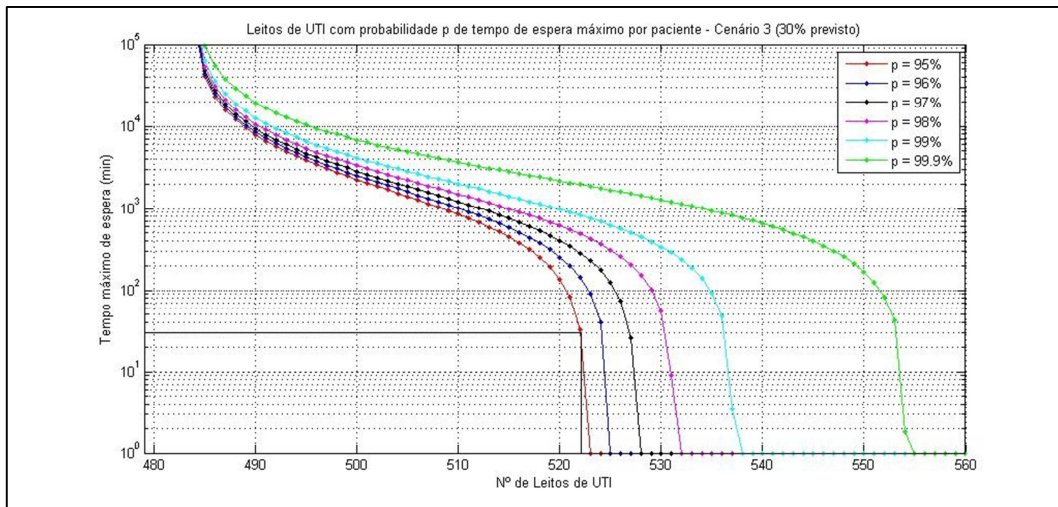


Figura 33 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário III (previsão fora da amostra - 30%)

Ao analisar a Figura 33, percebe-se que neste cenário são necessários:

- 484 leitos de UTI para o sistema ser estável.
- 522 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 555 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

#### Resultados para os dados reais da parte fora da amostra (30%)

O tempo médio entre chegadas é o mesmo do cenário I – dados reais da parte fora da amostra (30%), porém desconta-se a desistência por motivos de alta ou melhora clínica, assim  $\lambda = 2,166 * 0,7415$ , resultando em  $\lambda = 1,6065$  pacientes/hora. Como o tempo médio de utilização dos leitos é o mesmo do cenário I – dados reais da parte fora da amostra (30%), o  $\mu$  também será igual 0,003166 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 34.

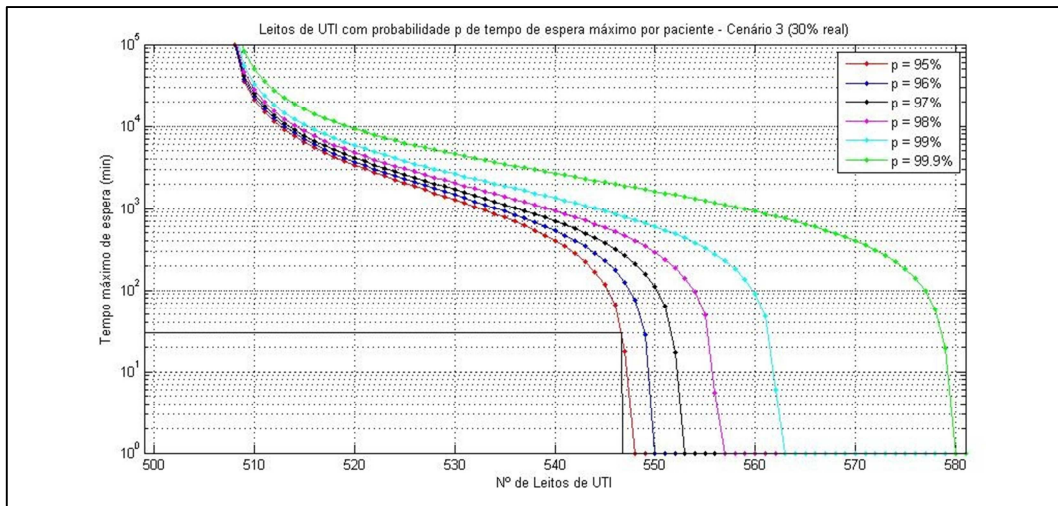


Figura 34 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário II (dados reais da parte fora da amostra - 30%)

Ao analisar a Figura 34, percebe-se que neste cenário são necessários:

- 508 leitos de UTI para o sistema ser estável.
- 546 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 580 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

#### 6.4 Cenário IV

O tempo entre chegadas será o mesmo do modelo III, isto é, o intervalo entre as solicitações por leitos de UTI – adulto formalizadas à CER, desconsiderando a desistência por motivos de alta ou melhora clínica. Já o tempo de serviço será o mesmo do cenário II, isto é, o tempo médio de todos os hospitais do Rio de Janeiro.

##### 6.4.1 Distribuições

Já que o tempo entre chegadas é o intervalo das solicitações por leitos de UTI – adulto formalizadas à CER descontado a porcentagem de desistência por alta ou melhora clínica, a distribuição será a mesma cenário I, isto é, uma exponencial.

O tempo de serviço é o mesmo do cenário II, assim a distribuição também é a mesma, ou seja, a distribuição é exponencial.

##### 6.4.2 Modelo

Ao identificar que o tempo entre chegadas e o tempo de serviço são distribuídos de acordo com uma distribuição exponencial e que  $s > 1$ , o modelo M/M/s aparece novamente como o mais apropriado. Assim, os resultados posteriores também serão determinados com a utilização das equações 40,41,46,47 descritas no capítulo 4.

### 6.4.3 Resultados

#### Resultados para previsão de 30 dias

Considerando o mesmo tempo médio entre chegadas do Cenário III-Previsão de 30 dias,  $\lambda$  também será igual a 1,635614pacientes/hora. E considerando o mesmo tempo médio de utilização dos leitos do cenário II - 30 dias, o  $\mu$  também será igual 0,006471 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 35.

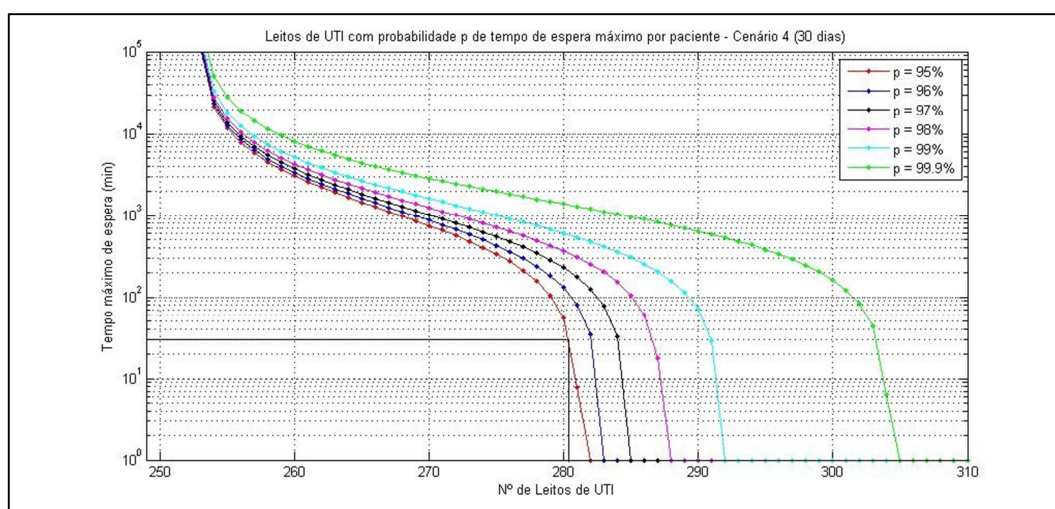


Figura 35 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (Previsão de 30 dias)

Ao analisar a Figura 35, percebe-se que neste cenário são necessários:

- 253 leitos de UTI para o sistema ser estável.
- 280 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 305 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

#### Resultados para previsão de 1 ano

Considerando o mesmo tempo médio entre chegadas do Cenário III- previsão de 1 ano,  $\lambda$  também será igual a 1,629969 pacientes/hora. E considerando o mesmo tempo médio de utilização dos leitos do Cenário II- previsão de 1 ano, o  $\mu$  também será igual 0,006442 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, também determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 36.

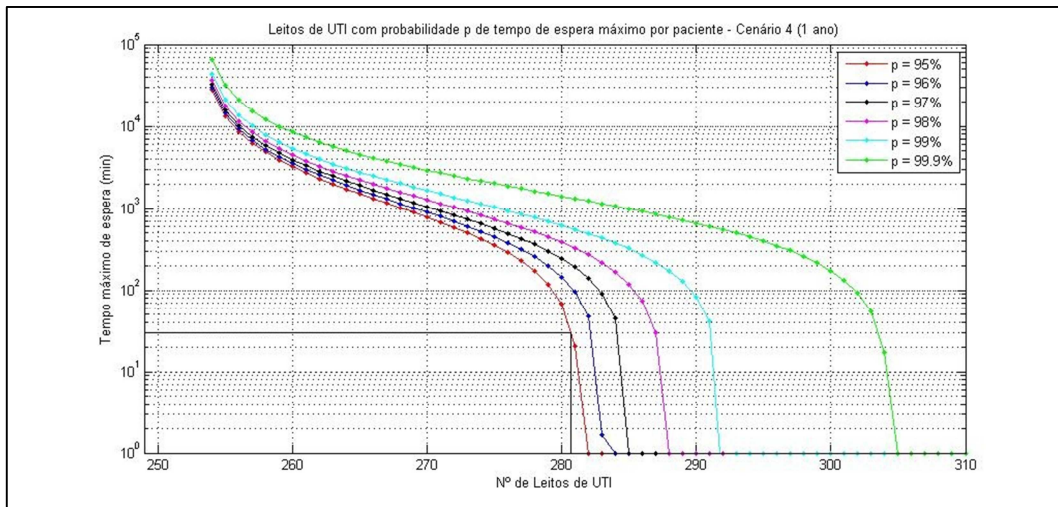


Figura 36 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (previsão de 1 ano)

Ao analisar a Figura 36, percebe-se que neste cenário são necessários:

- 254 leitos de UTI para o sistema ser estável.
- 280 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 305 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

#### Resultados para previsão fora da amostra (30%)

Considerando o mesmo tempo médio entre chegadas do Cenário III- previsão fora da amostra (30%),  $\lambda$  também será igual a 1,531198 pacientes/hora. E considerando o mesmo tempo médio de utilização dos leitos do Cenário II- previsão fora da amostra (30%), o  $\mu$  também será igual 0,006487 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 37.



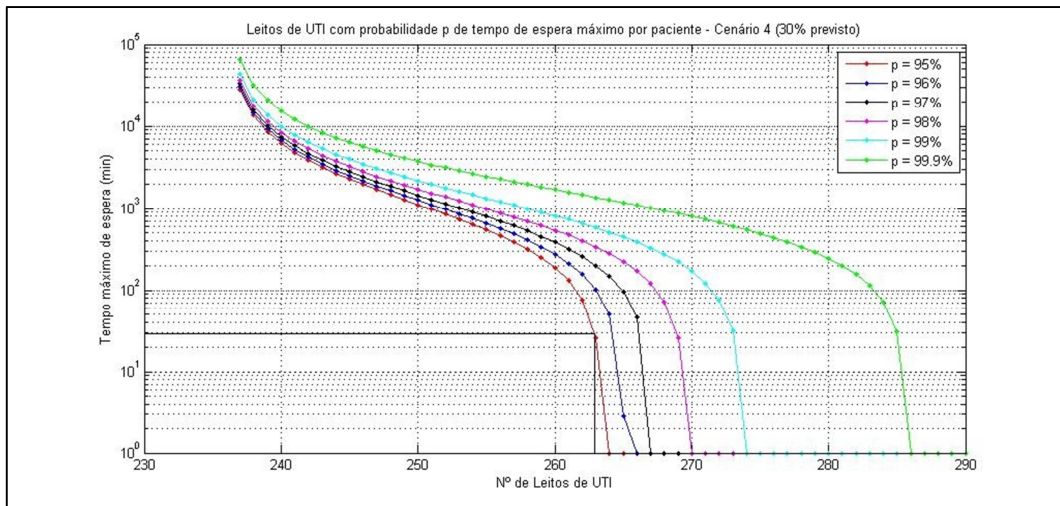


Figura 37 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (previsão fora da amostra - 30%)

Ao analisar a Figura 37, percebe-se que neste cenário são necessários:

- 237 leitos de UTI para o sistema ser estável.
- 262 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 286 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### Resultados para os dados reais da parte fora da amostra (30%)

Considerando o mesmo tempo médio entre chegadas do Cenário III- dados reais da parte fora da amostra (30%),  $\lambda$  também será igual a 1,606583 pacientes/hora. E considerando o mesmo tempo médio de utilização dos leitos do Cenário II- dados reais da parte fora da amostra (30%), o  $\mu$  também será igual 0,006487 pacientes/hora.

Com os parâmetros  $\lambda$  e  $\mu$  definidos, determinou-se, para todo o cenário, o número mínimo de leitos, os tempos máximos de espera e as curvas de probabilidades. Os resultados foram apresentados na Figura 38.

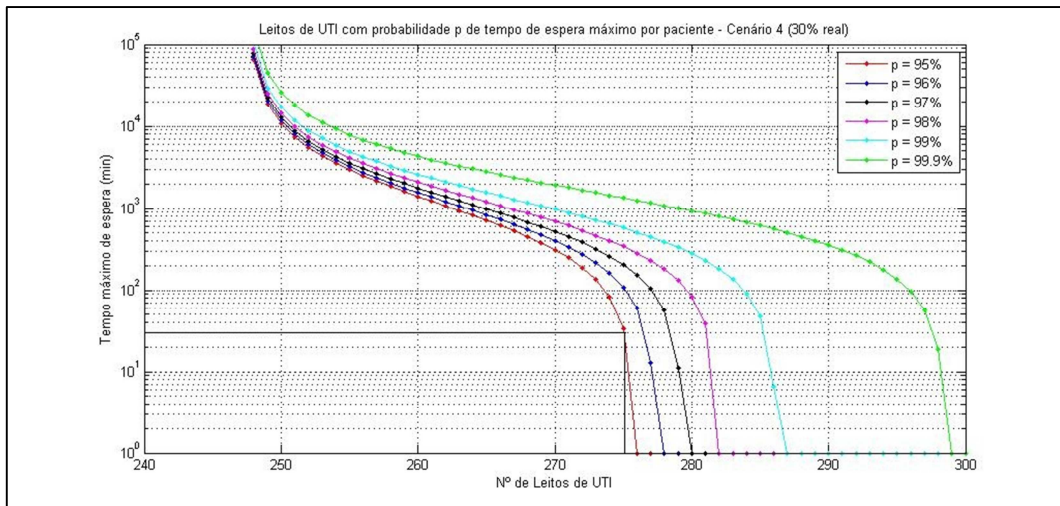


Figura 38 - Leitos de UTI com probabilidade p de tempo de espera máximo por paciente – Cenário IV (dados reais da parte fora da amostra - 30%)

Ao analisar a Figura 38, percebe-se que neste cenário são necessários:

- 248 leitos de UTI para o sistema ser estável.
- 275 leitos de UTI para a espera na fila ser no máximo de 30 minutos, com 95% de probabilidade.
- 299 leitos de UTI para a espera na fila ser no máximo de 1 minuto, com 99,9% de probabilidade.

### 6.5 Comparações dos resultados obtidos nos cenários propostos

Para facilitar a comparação, segue a tabela 8, com os quatro cenários dispostos resumidamente.

	<b>Cenário 1</b>	<b>Cenário 2</b>	<b>Cenário 3</b>	<b>Cenário 4</b>
<b>Taxa de Chegada</b>	Solicitações diária de pacientes	Solicitações diária de pacientes	Solicitações diária de pacientes - % Desistências	Solicitações diária de pacientes - % Desistências
<b>Taxa de Serviço</b>	Tempo médio de estadia dos hospitais da CER	Tempo médio de estadia dos hospitais do RJ	Tempo médio de estadia dos hospitais da CER	Tempo médio de estadia dos hospitais do RJ

Tabela 8 - Cenários propostos

Para previsões de 30 dias e 1 ano, identificou-se, para os quatro cenários apresentados anteriormente, o número mínimo de leitos que possibilite a estabilidade do sistema, uma espera na fila de no máximo 30 minutos, com 95% de probabilidade e uma espera na fila de no máximo 1 minuto, com 99,9% de probabilidade. Nas duas Figuras a seguir estão representados, esta quantidade de leitos previstos e as dos leitos atuais dos hospitais regidos pela CER, ressaltando que para os hospitais da CER também considerou-se como leitos de UTI-adulto, os leitos de UTI-adulto de cuidados intermediários.

### 6.5.1 Comparação dos cenários propostos com o atual - Previsão de 30 dias

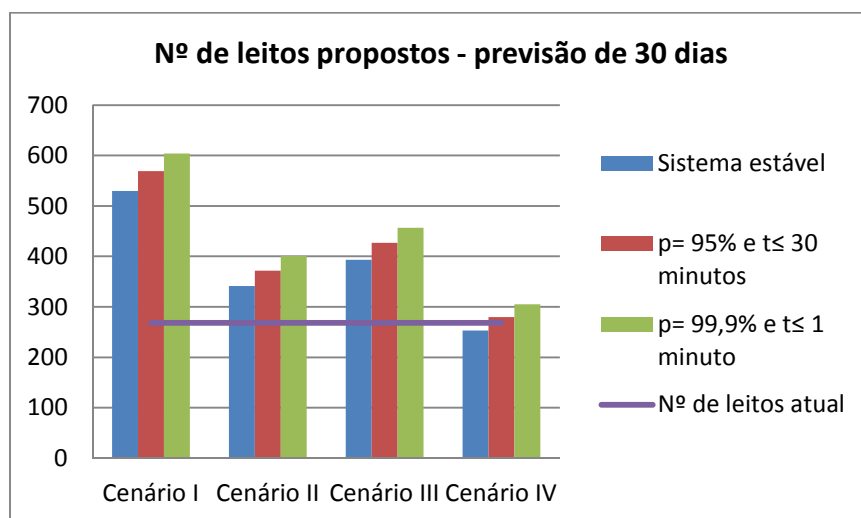


Figura 39 - Nº de leitos propostos x Nº de leitos atual (previsão de 30 dias)

Na Figura 39, percebe-se que, para os Cenários I,II e III, os 268 leitos atuais de UTI-adulto são insuficientes. Para o cenário IV, o número de leitos atual é suficiente apenas para estabilizar o sistema.

Na Figura 35 do cenário IV, percebe-se que se for considerado os 268 leitos atuais há 95% de probabilidade, da espera máxima na fila ser de aproximadamente 900 minutos e 99,9% de probabilidade, da espera máxima na fila ser de aproximadamente 3.200 minutos. Para os outros cenários não convém fazer esta análise, uma vez que, para nenhum deles o número de leitos atual é suficiente para estabilizar o sistema.

### 6.5.2 Comparação dos cenários propostos com o atual - Previsão de 1 ano

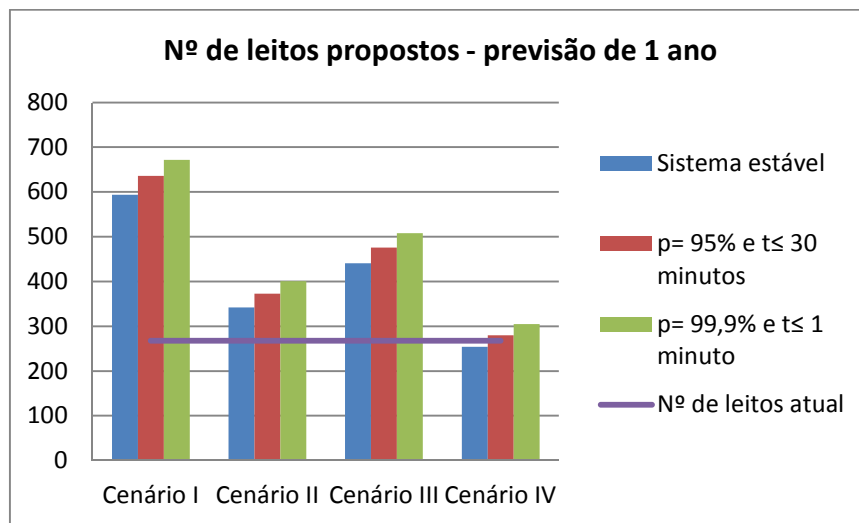


Figura 40 - Nº de leitos propostos x Nº de leitos atual (previsão de 1 ano)

Na Figura 40, percebe-se que, para os Cenários I,II e III, os 268 leitos atuais de UTI-adulto também são insuficientes. Para o cenário IV, o número de leitos atual é suficiente apenas para estabilizar o sistema.

Na Figura 36 do cenário IV, percebe-se que se for considerado os 268 leitos atuais há 95% de probabilidade, da espera máxima na fila ser de aproximadamente 1000 minutos e 99,9% de probabilidade, da espera máxima na fila ser de aproximadamente 3.300 minutos. Para os outros cenários não convém fazer esta análise, uma vez que, para nenhum deles o número de leitos atual é suficiente para estabilizar o sistema.

### 6.5.3 Comparação do número de leitos/1000 habitantes dos cenários propostos (previsão de 30 dias e 1 ano) e atual, conforme Portaria GM/MS nº 1.101

Sabe-se, como visto antes, que a Portaria 1.101 do Ministério da Saúde (2002) estabelece a seguinte quantidade de leitos:

- Leitos hospitalares totais - 2,5 a 3 leitos para cada 1.000 habitantes
- Leitos de UTI - 4% a 10% do total de leitos hospitalares

Baseando-se nesses parâmetros, obtém-se a relação de leitos de UTI por população. Assim, o número mínimo recomendado é de 4% de 2,5 leitos para cada 1000 habitantes, logo 0,1 leito de UTI para cada 1.000 habitantes. Já o número máximo recomendado é de 10% de 3 leitos para cada 1.000 habitantes, logo 0,3 leito de UTI para cada 1.000 habitantes.

Porém, este estudo refere-se a leitos de UTI-adulto o qual contempla Unidade de Cuidados Intermediários Adulto. Assim, de acordo com o CNES (2014), ambos representam 49,5302% dos leitos de UTI do Brasil, então o número mínimo de leitos de UTI-adulto é de

0,0495302 leitos para cada 1.000 habitantes e número máximo é de 0,148590 leitos para cada 1.000 habitantes.

Uma consideração importante deve ser ressaltada quanto à população de referência. Pois, de acordo com a ANS (2013), 63,4% da população do estado do Rio de Janeiro dependem exclusivamente do SUS e 36,6% possuem planos privados de saúde, os quais nem sempre cobrem leitos de UTI. Entretanto, o escopo deste estudo considera como dependentes do SUS apenas a parcela não possuidora de plano privado de saúde. Assim, como o estado do Rio de Janeiro, de acordo com o IBGE (2014) possui 16,447129 milhões de habitantes, inicialmente a população referência será de 10,427479 milhões, contudo esta será mais especificada a seguir.

De acordo com a CNES (2014), no estado do Rio de Janeiro, o número total de leitos de UTI-adulto existentes no SUS é de 1187. Contudo, para este estudo há 268 leitos de UTI-adulto no SUS do Rio de Janeiro, pois apenas os leitos dos hospitais regidos pela CER são considerados para tanto.

Assim, a definição da população referência a ser analisada foi extraída da razão entre os leitos UTI-adulto da CER e os leitos totais do estado do Rio de Janeiro multiplicada pela população do Rio de Janeiro dependente exclusivamente do SUS, resultando na população referência a ser considerada de 2,3543087 milhões de habitantes. Sobre os quais calculou as taxas de UTI-adulto para cada 1000 habitantes nos cenários propostos para previsões de 30 dias e 1 ano, além do cenário atual.

E, por fim, nas Figuras 41 e 42, comparou as taxas de UTI-adulto/1000 habitantes com os parâmetros estabelecidos de acordo com quantidade de leitos definida pela portaria nº 1.101 do Ministério da Saúde (2002). Também, vale ressaltar que as Figuras 41 e 42 são compostas respectivamente por cenários propostos para 30 dias e 1 ano.

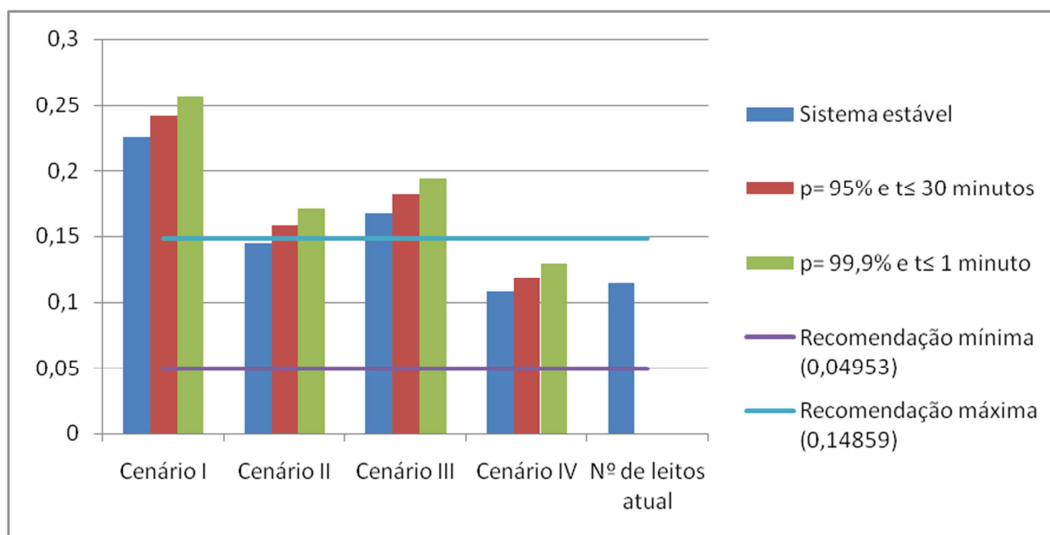


Figura 41 - Leitos de UTI adulto (SUS)/1000 habitantes para cenários com previsão de 30 dias

Da Figura 41, percebe-se que o cenário IV e o atual estão inteiramente dentro do limite estabelecido pela portaria GM/MS nº 1.101, no Cenário II apenas o número mínimo de leitos necessário para tornar o sistema estável esta dentro dos limites estabelecidos, já o restante do Cenário II e os Cenários I e III encontram-se acima do limite de recomendação máxima.

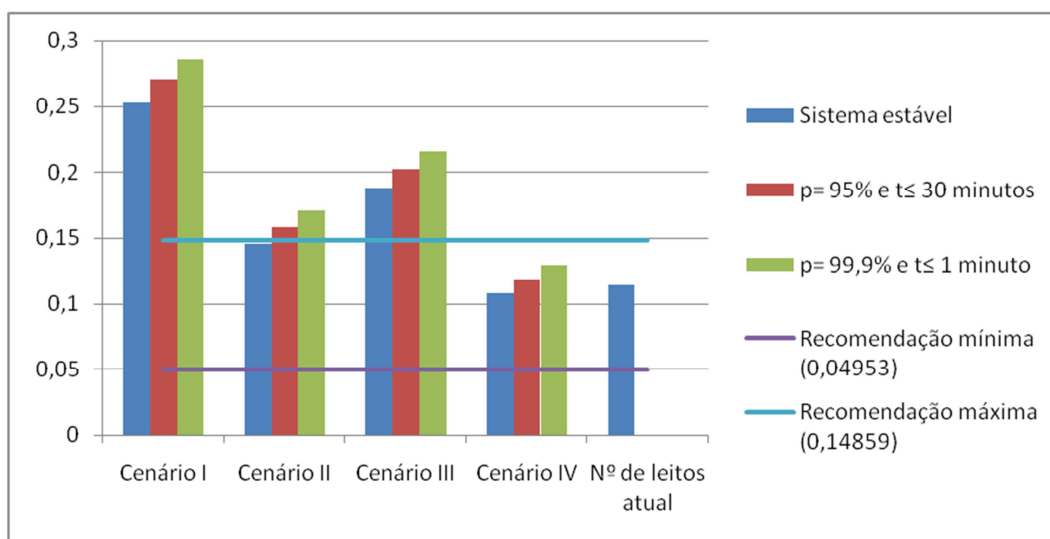


Figura 42 - Leitos de UTI adulto (SUS)/1000 habitantes para cenários com previsão de 1 ano

Da Figura 42, também percebe-se que o cenário IV e o atual estão inteiramente dentro do limite estabelecido pela portaria GM/MS nº 1.101, no Cenário II apenas o número mínimo de leitos necessário para tornar o sistema estável esta dentro dos limites estabelecidos, já o restante do Cenário II e os Cenários I e III encontram-se acima do limite de recomendação máxima.

### 6.5.4 Comparação do número de leitos da previsão fora da amostra x dados reais da parte fora da amostra

Para a previsão fora da amostra (30%) e os dados reais da parte fora da amostra (30%) identificou-se, para os quatros cenários apresentados anteriormente, o número mínimo de leitos que possibilite a estabilidade do sistema, uma espera na fila de no máximo 30 minutos, com 95% de probabilidade e uma espera na fila de no máximo 1 minuto, com 99,9% de probabilidade. Estes números foram representados nas Figuras 43, 44 e 45.

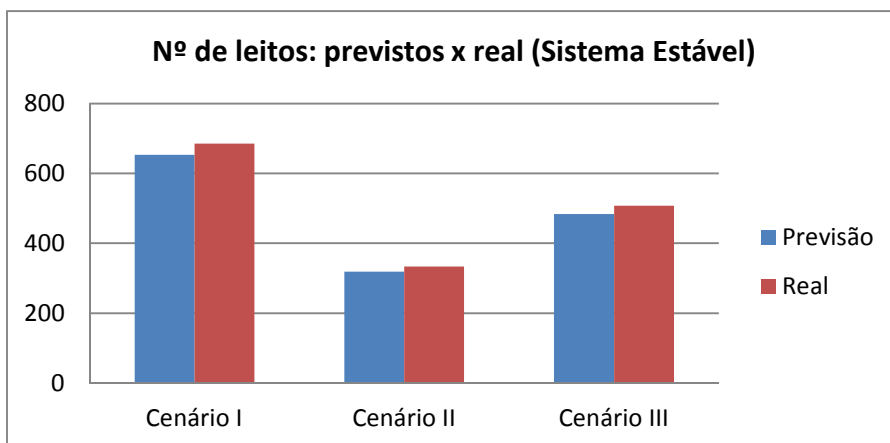


Figura 43 -Nº de leitos: Previsto x Real - Sistema Estável

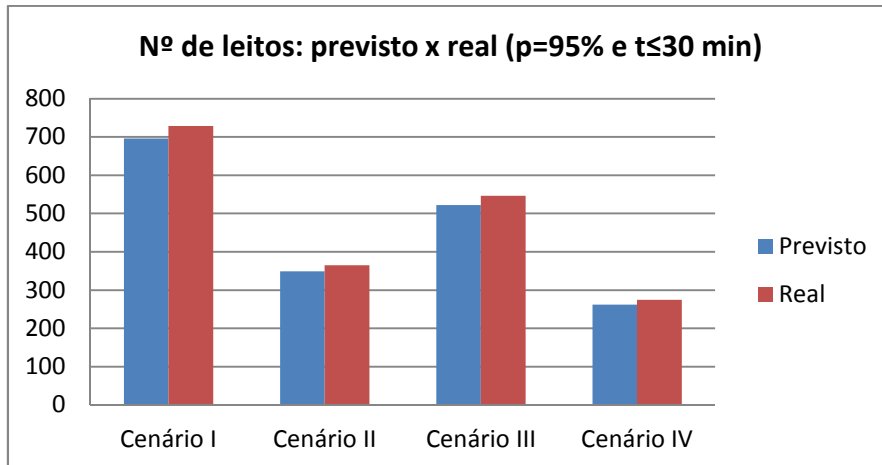


Figura 44 - Nº de leitos: Previsto x Real - (p=95% e t≤30 min)

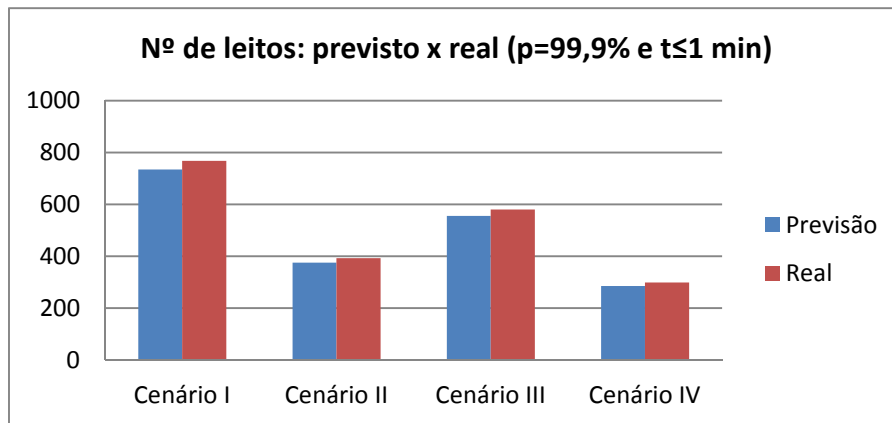


Figura 45- Nº de leitos: Previsto x Real - (p=95% e t≤30 min)

A partir das Figuras 43,44 e 45, percebe-se que em todas as comparações o número de leitos previstos foi menor que o real. De acordo com os dados da Figura 43, a diferença média do nº de leitos: previsto x real foi de 4,8%, para os dados da Figura 44 essa diferença foi de 4,72% e para os dados da Figura 45 foi de 4,55%.



## 7. Conclusão e Comentários Finais

Este trabalho se propôs a facilitar o planejamento da disponibilidade de leitos de UTI por meio da previsão de demanda de chegada de pacientes, a partir da qual foi dimensionado o número ótimo dos respectivos leitos sob critérios pré-estabelecidos.

A previsão de demanda foi fundamental para este estudo, pois aumentou a possibilidade do modelo de filas ser alimentado com dados mais realistas, já que a previsão capta características da série como tendência e sazonalidade. Além disso, a previsão alimenta um modelo estocástico de filas, o qual considera variações em torno da média e as quantifica.

A aplicação da teoria de filas neste estudo se mostrou simples e eficiente, uma vez que os dados apresentaram distribuição probabilística exponencial, logo foi possível aplicar o modelo tradicional M/M/s. Entretanto, para distribuições probabilísticas não-exponenciais a análise matemática dos modelos de fila se torna bem mais difícil e nem sempre matematicamente tratável.

Agora, quanto à quantidade ótima de leitos para os quatro cenários obtida com a teoria das filas, constata-se que os cenários II e IV apresentam os melhores resultados por requererem um número menor de leitos e, conseqüentemente, apresentarem uma quantidade mais próxima do número atual de leitos. Porém, o melhor desempenho desses cenários se deve a suposição de uma possível redução de seu tempo de serviço, ou seja, supõe-se que o tempo médio de estadia dos hospitais regidos pela CER do estado do Rio de Janeiro seja reduzido para o mesmo tempo médio de estadia em todos os hospitais do estado do Rio de Janeiro. No entanto, a viabilidade desta redução não é conhecida, assim, para trabalhos futuros sugere-se que investiguem-se a grande diferença do tempo médio de serviço de todos os hospitais do estado do Rio de Janeiro com aquele dos hospitais regidos somente pela CER, bem como para determinar a viabilidade ou não da redução do tempo de serviço.

Na comparação dos cenários II e IV, este último se destaca uma vez que neste cenário desconta-se a porcentagem de desistência por motivos de alta ou melhora clínica. E, assim demonstra a importância de uma solicitação correta.

Da comparação do número de leitos/1000 habitantes dos cenários propostos e atual, conforme a portaria GM/MS nº 1.101, verifica-se que apenas o cenário IV encontra-se dentro do limite de recomendação, desta forma pode-se dizer que para os dados deste estudo a quantidade de leitos determinada pela portaria não é suficiente para o restante dos cenários. No entanto, ainda não se pode afirmar que a metodologia utilizada para estabelecer a portaria pode não ter sido a ideal ou ter se baseado em metodologias inadequadas à realidade atual, uma vez que um cenário se encontrou dentro dos limites estabelecidos e também porque os dados deste estudo representam uma parcela pequena e são bem específicos, isto é, considera-se apenas UTI-adulto

de alguns hospitais do estado do Rio de Janeiro. Enquanto que para a portaria foi considerado todos os tipos de UTI para uma dada população, a qual pode abranger até mesmo todo o país. Assim, para futuros trabalhos também sugere-se a expansão deste estudo para todos os tipos de leitos e para todos os hospitais do estado do Rio de Janeiro.

Da comparação do número de leitos previsto e real, verifica-se que apesar do número real ter sido menor que o previsto em todas as comparações, a diferença foi menor que 5%, assim pode-se dizer que o método utilizado no estudo foi válido.

## Referências Bibliográfica

ABENSUR, E. O. et al. A strategic approach based on the queuing theory. **Revista de Administração Mackenzie**, n. 2, p. 39–59, 2003.

ALAGOZ, O. et al. The optimal timing of living-donor liver transplantation. *Management Science*. **Management Science**, v. 50, n. 10, p. 1420–1430, 2004.

ALBUQUERQUE, N. **Avaliação do acesso a leitos de regulação hospitalar do SUS em pé no ano de 2009**. Recife, PE, Brasil: Monografia Espec., Centro de Pesquisas Aggeu Magalhães/Fundação Oswaldo Cruz, 2009.

**ANS - Agência Nacional de Saúde Suplementar**. Disponível em: <[www.ans.gov.br](http://www.ans.gov.br)>. Acesso em: 14 ago. 2013. 2013.

BARBOSA, A. P. Terapia intensiva neonatal e pediátrica no Brasil : o ideal , o real e o possível. **Jornal de pediatria**, v. 80, n. 6, p. 437–438, 2004.

BARD, J. F. e PURNOMO, H. W. Preference scheduling for nurses using column generation. **European Journal of Operational Research**, v. 164, n. 2, p. 510–534, 2005.

BOSE, S. K. **An introduction to queueing systems**. 1. ed. New York: Springer US, 2001.

BOWERMAN, B. L.; O'CONNELL, R. T. e KOEHLER, A. B. **Forecasting, time series and regression: an applied approach**. 4. ed. Belmont CA: Cengage Learning, 2005.

BOX, G. E. P. e JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. 4. ed. New Jersey: Wiley, 1970.

BRASIL. Lei 8.080 de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. **Diário Oficial da União**, Brasília, DF, 20 set. 1990. Seção 1. p.1. Disponível em <<http://www.in.gov.br>>. Acesso em: 21 jul. 2013

BRASIL. Lei 8.142 de 28 de dezembro de 1990. Dispõe sobre a participação da comunidade na gestão do Sistema Único de Saúde (SUS) e sobre as transferências intergovernamentais de recursos financeiros na área da saúde e dá outras providências. **Diário**

**Oficial da União**, Brasília, DF, 31 dez. 1990. Seção 1. p.4. Disponível em <<http://www.in.gov.br>>. Acesso em: 19 jul. 2013

BROCKWELL, P. J. e DAVIS, R. A. **Time Series: Theory and Methods**. 2. ed. New York: Springer, 1991.

BROWN, R. G. **Statistical forecasting for inventory control**. New York: McGraw-Hill, 1959.

CARDOEN, B.; DEMEULEMEESTER, E. e BELIËN, J. Operating room planning and scheduling: A literature review. **European Journal of Operational Research**, v. 201, n. 3, p. 921–932, 2010.

CHATFIELD, C. **TIME-SERIES FORECASTING**. 1. ed. New York: Chapman and Hall, 2000.

CLEMENTE, A. e CLEMENTE, L. T. Aplicação da metodologia Box-Jenkins para previsão do ICMS do estado do Paraná de agosto de 2011 a julho de 2012. **Economia e Tecnologia**, v. 27, p. 47–58, 2011.

**CNES - Cadastro Nacional dos Estabelecimentos de Saúde**. Disponível em: <[cnes.datasus.gov.br](http://cnes.datasus.gov.br)>. Acesso em: 23 mai. 2014. 2014.

**DATASUS - Departamento de Informática do SUS**. Disponível em: <[www.datasus.gov.br](http://www.datasus.gov.br)>. Acesso em: 18 ago. 2013 2013.

DOBSON, G.; HASIJA, S. e PINKER, E. J. Reserving Capacity for Urgent Patients in Primary Care. **Production and Operations Management**, v. 20, n. 3, p. 456–473, 2011.

GARDNER, E. S. J. Exponential smoothing: the state of the art. **Journal of Forecasting**, v. 4, p. 1–28, 1985.

GARDNER, E. S. J. Exponential smoothing : The state of the art — Part II. **Journal of Forecasting**, v. 22, p. 637–666, 2006.

GARDNER, E. S. e MCKENZIE, E. Why the damped trend works. **Journal of the Operational Research Society**, v. 62, n. 6, p. 1177–1180, 2010.

GONÇALVES, R. M. e PONTES, E. P. **Estudo de taxa de ocupação de leitos de UTI do estado de Minas Gerais**. Brasília/DF: Trabalho apresentado no V Congresso Consad, 2012.

GONÇALVES, T. C. R. **Análise e simulação computacional de processos aplicada ao fluxo de pacientes no hospital**. São Paulo,SP,Brasil: Monografia, Escola Politécnica/USP, 2010.

GREEN, L. V. e SAVIN, S. Reducing Delays for Medical Appointments: A Queueing Approach. **Operations Research**, v. 56, n. 6, p. 1526–1538, 2008.

GRIFFITHS, J. D. and L. N. P. e SMITHIES, M. AND WILLIAMS, J. A queueing model of activities in an intensive care unit. **IMA Journal of Management Mathematics**, v. 17, p. 277–288, 2006.

HILLIER, F. S. e LIEBERMAN, G. J. **Introdução à Pesquisa Operacional**. 8<sup>a</sup>. ed. São Paulo: Mc Graw Hill, 2010.

HOLT, C. C. Forecasting trends and seasonals by exponentially weighted averages, O.N.R. Memorandum 52/1957. **Carnegie Institute of Technology**, 1957.

HOUT, W. B. VAN DEN et al. The heart-allocation simulation model: A tool for comparison of transplantation allocation policies. **Transplantation**, v. 76, n. 10, p. 1492–1497, 2003.

HYNDMAN, R. J. et al. It's time to move from “what” to “why”—comments on the M3-competition. **International Journal of Forecasting**, v. 17, n. 4, p. 567–570, 2001.

HYNDMAN, R. J. et al. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439–454, 2002.

HYNDMAN, R. J. et al. **Forecasting with Exponential Smoothing**. New York: Springer, 2008.

HYNDMAN, R. J. e KHANDEKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. **Journal of Statistical Software**, v. 27, 2008.

**IBGE – Instituto Brasileiro de Geografia e Estatística**. Disponível em: <[www.ibge.gov.br](http://www.ibge.gov.br)>. Acesso em: 12 mar. 2014. 2014.

JONES, S. S. et al. Forecasting daily patient volumes in the emergency department. **Academic emergency medicine : official journal of the Society for Academic Emergency Medicine**, v. 15, n. 2, p. 159–70, 2008.

KENDALL, M. G. e ORD, J. K. **Time Series**. 3. ed. Lond: Oxford University Press, 1990.

KLEINROCK, L. **QUEUEING SYSTEMS volume I: theory**. New York: John Wiley & Sons, 1975.

MABERT, V. A. **An Introduction to short term forecasting using Box-Jenkins Methodology**. Norcross, Georgia: Monograph, American Institute of Industrial Engineers, 1975.

MAKRIDAKIS, S. e HIBON M. Accuracy of forecasting: an emperial investigation (with discussion). **Journal of the Royal Statistical Society**, v. 142, n. 2, p. 97–145, 1979.

MAKRIDAKIS, S.. ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E., e WINKLER, R., **The Forecasting Accuracy of Major Time Series Methods**. London: Wiley, 1983.

MAKRIDAKIS, S. , CHATFIELD C., HIBON M., LAWRENCE M., MILLS T., ORD. K. E SIMMONS, L.F. The M -2 Competition: a real-time judgmentally based forecasting study. **International Journal of forecasting**, v. 9, n. 2, p. 5–23, 1993.

MAKRIDAKIS, S., WHEELWRIGHT, W. C. e HYNDMAN, R. J. **Forecasting, Methods and Applications**. 3. ed. New York: John Wiley & Sons, 1998.

MAKRIDAKIS, S. e HIBON M. The M -3 Competition: results, conclusions and implications. **International Journal of forecasting**, v. 16, p. 451–476, 2000.

MARINHO, A. Um estudo sobre as filas para transplantes no Sistema Único de Saúde brasileiro. **Cad. de saúde Pública**, v. 22, n. 10, p. 2229–2239, 2006.

MAYHEW, L. e SMITH, D. Using queuing theory to analyse the Government's 4-h completion time target in Accident and Emergency departments. **Health Care Management Science**, v. 11, n. 1, p. 11–21, 2007.

MCMANUS, M. L. et al. Queuing Theory Accurately Models the Need for Critical Care Resources. **American Society of Anesthesiologists**, v. 100, n. 5, p. 1271–1276, 2004.

MILNE, E. e WHITTY, P. Calculation of the need for paediatric intensive care beds. **Archives of Disease in Childhood**, v. 73, p. 505–507, 1995.

Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Resolução nº 7 de 24 de fevereiro de 2010. Estabelece padrões mínimos para o funcionamento das Unidades de Terapia Intensiva, visando à redução de riscos aos pacientes, visitantes, profissionais e meio ambiente. **Diário Oficial da União**, Brasília, DF, 24 fev. 2010.

Ministério da Saúde. Gabinete do Ministro. Portaria 3.432/98. Estabelece critérios de classificação para as Unidades de Tratamento ou Terapia Intensivo – UTI. **Diário Oficial da União**, Brasília, DF, 13 ago. 1998.

Ministério da saúde. Gabinete do Ministro. Portaria Nº 466, DE 04 DE JUNHO DE 1998. Estabelece o Regulamento Técnico para o Funcionamento dos Serviços de Tratamento Intensivo e sua respectiva classificação de acordo com o grau de complexidade, capacidade de atendimento e grau de risco inerente ao tipo de atendimento prestado. **Diário Oficial da União**, Brasília, DF, 5 jun. 1998. Seção 1, p. 9. Disponível em <<http://www.in.gov.br>>. Acesso em: 21 mar. 2014. 1998.

Ministério da Saúde. Secretaria de Assistência à Saúde. Portaria 1.101/2002. Estabelece os parâmetros de cobertura assistencial do Sistema Único de Saúde - SUS. **Diário Oficial da União**, Brasília, DF, 13 jun. 2002. Seção 1. p.36. Disponível em <<http://www.in.gov.br>>. Acesso em: 14 jul. 2013.

MONTGOMERY, D. C.;; JOHNSON, L. A. e GARDINER, J. **Forecasting and Time Series Analysis**. 2. ed. New York: McGraw-Hill, 1990.

MORETTIN, P. A. e TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: McGraw-Hil, 2006.

NDIAYE, M. e ALFARES, H. Modeling health care facility location for moving population groups. **Computers & Operations Research**, v. 35, n. 7, p. 2154–2161, 2008.

NISHIDE, V. M.;; CINTRA, E. de A. e NUNES, W. A. **Assistência de Enfermagem ao Paciente Gravemente Enfermo**. 2. ed. São Paulo: ATHENEU, 2005.

NOGUEIRA, L. D. S. et al. Características clínicas e gravidade de pacientes internados em UTIs públicas e privadas. **Texto & Contexto - Enfermagem**, v. 21, n. 1, p. 59–67, 2012.

OLIVEIRA, M. D. e BEVAN, G. Modelling hospital costs to produce evidence for policies that promote equity and efficiency. **European Journal of Operational Research**, v. 185, n. 3, p. 933–947, 2008.

PAPAGEORGIU, J. C. Some operations research applications to problems of health care systems (a survey). **International Journal Biomedical Computation**, v. 9, n. 2, p. 101–104, 1978.

PEARSON, G. A.; REYNOLDS, F. e STICKLEY, J. Calculating the need for intensive care beds. **Archives of disease in childhood**, v. 97, p. 943–946, 2012.

PEGEL, C. C. Exponential forecasting: some new variations. **Management Science**, v. 15, n. 5, p. 311–315, 1969.

PEREIRA, B. B. **Tópicos em series temporais: Métodos automáticos de previsão**. COPPE/UFRJ, Rio de Janeiro, Brasil, 1980.

PERSSON, M. e PERSSON, J. A. Health economic modeling to support surgery management at a Swedish hospital. **Omega**, v. 37, n. 4, 2009.

**PORTAL DA SAÚDE - Ministério da Saúde**. Disponível em: <portalsaude.saude.gov.br>. Acesso em: 01 Jun. 2013. 2013.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>. 2012. Acesso em 15/01/2013. 2012.

RAIS, A. e VIANA, A. Operations Research in Healthcare: a survey. **International Transactions in Operational Research**, v. 18, n. 1, p. 1–31, 2011.

RHODES, a et al. The variability of critical care bed numbers in Europe. **Intensive care medicine**, v. 38, n. 10, p. 1647–53, 2012.

ROYSTON, G. One hundred years of operational research in Health – UK 1948–2048. **Quarterly Journal of Economics**, v. 60, n. 1, p. 169–179, 2009.

SALLES, A. **Notas de Aula de Séries Temporais: Introdução aos Modelos de Previsão**. Escola Politécnica - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, 2007.



SANTOS, Ia. C. Dos;; VARGAS, M. A. de O. e SCHNEIDER, N. Encaminhamento do paciente crítico para UTI por decisão judicial: situações vivenciadas pelos enfermeiros. **Enfermagem em Foco**, v. 1, n. 3, p. 94–97, 2010.

**SES - Secretaria de Estado de Saúde.** Disponível em: <[www.rj.gov.br/web/imprensa/exibeconteudo?article-id=1876643](http://www.rj.gov.br/web/imprensa/exibeconteudo?article-id=1876643)>. Acesso em: 02 Ago. 2013. 2013.

**SESDEC - Secretaria de Estado de Saúde e Defesa Civil/Superintendência de regulação.** Disponível em: <<http://www.saude.rj.gov.br>>. Acesso em: 02 jun. 2013. 2010.

SHMUELI, A.;; SPRUNG, C. L. e KAPLAN, E. H. Optimizing admissions to an intensive care unit. **Health care management science**, v. 6, n. 3, p. 131–6, 2003.

SIDDHARTHAN, K.;; JONES, W. J. e JOHNSON, J. a. A priority queuing model to reduce waiting times in emergency care. **International journal of health care quality assurance**, v. 9, n. 5, p. 10–6, 1996.

SMITH, H. K. et al. Planning sustainable community health schemes in rural areas of developing countries. **European Journal of Operational Research**, v. 193, n. 3, p. 768–777, 2009.

TAKAGI, H. **Queueing analysis, vol 1.** Amsterdam: North Holland Elsevier, 1991.

TAYLOR, J. W. Exponential Smoothing with a Damped Multiplicative Trend Exponential Smoothing with a Damped Multiplicative Trend. **International Journal of Forecasting**, v. 19, n. 0, p. 715–725, 2003.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages, Management Science. **Management Science**, v. 6, n. 324–342, 1960.

XUE, J. L. et al. Forecast of the number of patients with end-stage renal disease in the United States to the year 2010. **Journal of the American Society of Nephrology : JASN**, v. 12, n. 12, p. 2753–8, 2001.

ZHU, Z.;; HEN, B. H. e TEOW, K. L. Estimating ICU bed capacity using discrete event simulation. **International Journal of Health Care Quality Assurance**, v. 25, p. 134–144, 2010.