



Recebido em: 25/06/2024.

Aceito em: 22/09/2024.

Revista SODEBRAS – Volume 19
Nº 222 – SETEMBRO/ DEZEMBRO - 2024

ANÁLISE E MODELAGEM DE FILAS NO CONTEXTO DA SAÚDE PÚBLICA BRASILEIRA: ESTUDOS E APLICAÇÕES DIVERSIFICADAS

QUEUE ANALYSIS AND MODELING IN THE CONTEXT OF BRAZILIAN PUBLIC HEALTH: DIVERSE STUDIES AND APPLICATIONS

Lucas Moreira¹
Yasmin Lírio Souza de Oliveira²

Resumo – O Sistema Único de Saúde (SUS) no Brasil enfrenta desafios complexos relacionados à gestão de filas em seus serviços, especialmente durante eventos como a pandemia de COVID-19 e crises decorrentes de desastres naturais. Essas situações destacam a importância de estudar e modelar as filas no SUS para garantir a eficiência e a qualidade dos serviços de saúde. Este estudo tem como objetivo principal analisar e modelar as filas de atendimento no SUS, considerando as demandas específicas decorrentes da pandemia de COVID-19 e de eventos climáticos extremos. Além disso, busca-se identificar estratégias que possam otimizar a gestão de filas e melhorar o acesso dos pacientes aos serviços de saúde. Os resultados demonstraram a aplicabilidade dos modelos de filas no contexto do SUS, destacando a importância de estratégias como a otimização do fluxo de pacientes, a implementação de sistemas de triagem eficientes e o uso de tecnologias para monitoramento e gestão das filas de espera. Logo, a análise e modelagem de filas no SUS são cruciais para aprimorar a eficiência e a qualidade dos serviços de saúde, principalmente diante de cenários de alta demanda, como em desastres naturais e surtos virais. Ressalta-se que a aplicabilidade dos resultados varia conforme as características e recursos de cada sistema de saúde, sendo essencial adaptar as estratégias de gestão de filas conforme as necessidades locais.

Palavras-chave: Sistema Único de Saúde (SUS) 1. Modelagem de Serviços de Saúde 2. Tecnologias em Saúde 3. Estratégias de Gestão de Filas 4.

Abstract – The Unified Health System (SUS) in Brazil faces complex challenges related to queue management in its services, especially during events such as the COVID-19 pandemic and crises arising from natural disasters. These situations highlight the importance of studying and modeling queues in the SUS to guarantee the efficiency and quality of health services. This study's main objective is to analyze and model service queues in the SUS, considering the specific demands arising from the COVID-19 pandemic and extreme weather events. Furthermore, we seek to identify strategies that can optimize queue management and

¹ Doutor em Estatística. Universidade de Brasília (UnB). lmoreira@unb.br

² Mestre em Estatística. Universidade de Brasília (UnB). yasmin-lirio@hotmail.com

improve patient access to health services. The results demonstrated the applicability of queuing models in the context of the SUS, highlighting the importance of strategies such as optimizing patient flow, implementing efficient triage systems and the use of technologies for monitoring and managing waiting lines. Therefore, the analysis and modeling of queues in the SUS are crucial to improving the efficiency and quality of health services, especially in high demand scenarios, such as natural disasters and viral outbreaks. It should be noted that the applicability of the results varies according to the characteristics and resources of each health system, and it is essential to adapt queue management strategies according to local needs.

Keywords: Unified Health System (SUS) 1. Modeling of Health Services 2. Health Technologies 3. Queue Management Strategies 4.

I. INTRODUÇÃO

A Teoria das Filas é utilizada para estudar sistemas em que usuários buscam um determinado serviço de atendimento. Nesse contexto, é desejável que se possa prever o número de clientes que chegam ou deixam o sistema, o tempo de atendimento e, nos casos em que há formação de filas, estimar o tempo de espera pelo serviço. Com isso, pode-se propor adequações na infraestrutura do atendimento e evitar situações incômodas que uma espera prolongada por um serviço possa causar.

Erlang (1909) realizou um dos primeiros estudos acerca da Teoria das Filas, cuja aplicação foi em circuitos telefônicos e ao problema de tráfego de chamadas nas linhas. Nesse trabalho, ele provou que as ligações chegavam segundo uma distribuição Poisson. Também, estudou-se o tempo médio no atendimento das ligações.

O arcabouço teórico da Teoria das Filas se dá por meio de conceitos de Processos Estocásticos. No entanto, é utilizada e documentada na literatura de áreas como Pesquisa Operacional (Perdoná *et al.*, 2017) e Engenharia Industrial (Camelo *et al.*, 2010). Os sistemas de filas podem ser utilizados em diversas situações. Fila de transplante de órgãos no Sistema Único de Saúde (Moreira & Bastos, 2017, 2018), controle de tráfego (aéreo, veículos, pessoas etc.), sistemas de comunicação e clientes que esperam por um serviço (em um supermercado, agência bancária, lotérica etc.) são alguns exemplos que podem ser estudados por meio dessa teoria.

A partir de um padrão das chegadas dos clientes ao serviço, dos atendimentos fornecidos pelo servidor e com base no número de canais de atendimento disponíveis, pode-se escolher um modelo estocástico de fila para analisar um sistema em particular. Uma vez que o tempo de espera para o acompanhamento reflete a qualidade dos serviços prestados, a Teoria das Filas também busca encontrar um ponto de equilíbrio que satisfaça o cliente e o orçamento disponibilizado pelo prestador de serviço.

Genericamente, um sistema de filas fica bem caracterizado quando se conhece o regime de chegada de clientes, a distribuição probabilística do tempo de serviço, o número de servidores, a capacidade de atendimento do complexo, a população de usuários e a disciplina de atendimento. Filas são formadas quando há maior procura em relação à capacidade da organização em suprir a demanda. Em muitas situações, não se pode evitar a formação de filas aumentando-se o número de servidores por questões de inviabilidade financeira ou limitações de espaço físico.

O presente trabalho aplica modelos de Filas Markovianas para analisar a lista de espera por leitos gerais e leitos COVID da Unidade de Terapia Intensiva (UTI) na rede pública de saúde do Distrito Federal (DF). Uma UTI representa uma área específica do hospital altamente especializada, contendo equipe técnica qualificada e multidisciplinar

destinada ao tratamento e monitoramento de pacientes que se apresentam em estado de saúde grave. Sendo assim, consiste em maior nível de complexidade por oferecer cuidados profissionais especializados e recursos tecnológicos sofisticados a fim de reestabelecer a saúde do paciente (Moraes *et al.*, 2018).

Apesar de uma UTI representar uma área de cuidado intensivo, a internação de pacientes em seus leitos por período prolongado pode ser sinônimo de malefícios tanto para o enfermo como para o próprio hospital. A geração de despesas excessivas, a possibilidade de rejeição da admissão de pacientes críticos devido à falta de leitos (resultando em aumento da morbidade intra-hospitalar), o maior risco de exposição a infecções e a alta probabilidade de desenvolvimento de problemas psicológicos, como ansiedade e depressão, são algumas das adversidades causadas pela internação por tempo prolongado na UTI. Sendo assim, a Teoria de Filas será aplicada para avaliar o desempenho do Sistema Público de Saúde no DF e propor a implementação de melhorias significativas.

Neste trabalho é apresentada uma revisão bibliográfica sobre alguns modelos de filas e a forma de classificação desses modelos. Posteriormente, a Teoria das Filas é aplicada ao alinhamento de espera por leitos de UTI do Sistema Único de Saúde (SUS) no Distrito Federal. A estimação dos parâmetros (tempo de espera médio na fila, taxa de ocupação média e tamanho médio da fila) é feita por meio do software livre R.

II. REFERENCIAL TEÓRICO

O referencial teórico deste estudo é fundamentado na Teoria de Filas, uma área da matemática aplicada amplamente utilizada para modelar sistemas de atendimento e filas de espera em diferentes contextos. A análise crítica e organizada da literatura pertinente ao tema proporciona uma compreensão aprofundada dos conceitos-chave, teorias e modelos que embasam a aplicação da Teoria de Filas em um conjunto de dados reais relacionados à espera por leitos de UTI no sistema público de saúde do Distrito Federal (DF) durante a terceira onda da pandemia.

A estrutura básica de um sistema de filas é composta pela Fonte, Chegada, Fila e Serviço. A Fonte representa a população de clientes potenciais, a Chegada indica como os clientes entram no sistema de atendimento, a Fila refere-se ao número de clientes aguardando atendimento, e o Serviço é o modo como o atendimento é prestado aos clientes. Essa estrutura pode variar em complexidade, incluindo sistemas com múltiplos estágios de atendimento e diferentes distribuições de tempo de serviço.

A taxa de serviço (μ) indica quantos clientes (ou pacientes) são atendidos em média por unidade de tempo e pode ser constante ou aleatória.

O regime de serviço abrange a disponibilidade, a capacidade do sistema e a duração do atendimento de cada usuário, que pode ser determinístico ou aleatório. Segundo Mendonça (2014), a disciplina de atendimento consiste na maneira pela qual os usuários que estão na fila são selecionados para serem atendidos. Os tipos de disciplinas de atendimento mais utilizados são:

- FIFO (first in - first out): os usuários são atendidos na ordem das chegadas. Essa disciplina de atendimento é a mais comumente adotada.
- LIFO (last in - first out): o primeiro usuário a ser atendido é o que chegou por último.
- PRI (priority service): o atendimento aos usuários segue uma ou mais prioridades preestabelecidas pela gerência do sistema.
- SIRO (service in random order): o atendimento aos usuários segue uma ordem aleatória.

Vale salientar que há outros tipos de disciplinas de atendimento, inclusive considerando aspectos como atendimento prioritário e desistências.

Em muitos sistemas, a disciplina FIFO (First In, First Out) é comum, onde o primeiro cliente a chegar é o primeiro a ser atendido.

A notação de Kendall (1956) é utilizada para descrever sistemas de filas (A/B/c/K/N/Z), onde cada símbolo representa um aspecto específico:

- A: Distribuição do tempo entre chegadas.
- B: Distribuição do tempo de serviço.
- c: Número de servidores.
- K: Capacidade total do sistema.
- N: Número de usuários potenciais.
- Z: Disciplina de atendimento.

Para simplificar, considera-se frequentemente um sistema com capacidade ilimitada ($K=\infty$), população infinita ($N=\infty$) e disciplina FIFO ($Z=\text{FIFO}$).

As filas Markovianas (M/M/m) assumem que tanto a chegada dos clientes quanto o tempo de serviço seguem uma distribuição exponencial, indicadas por "M" (Markoviana). O sistema M/M/1 é o modelo básico com um único servidor:

- Taxa média de chegada (λ): Número de clientes chegando por unidade de tempo.
- Taxa média de atendimento (μ): Número de clientes atendidos por unidade de tempo.
- Disciplina de atendimento: FIFO.

Para um sistema M/M/1, a função densidade para o tempo entre chegadas é $f(t) = \lambda e^{-\lambda t}$ e para os tempos de atendimento é $f(t) = \mu e^{-\mu t}$. O tempo médio entre chegadas é $\frac{1}{\lambda}$ e o tempo médio de serviço é $\frac{1}{\mu}$.

Um sistema de filas pode estar em condição transiente inicialmente, mas tende a atingir um estado de equilíbrio ao longo do tempo. Para um sistema em equilíbrio, as principais medidas de desempenho são:

- Fator de Utilização do Servidor (U): $U = \frac{\lambda}{\mu}$.
- Número Médio de Clientes no Sistema (L): $L = \frac{\lambda}{\mu - \lambda}$.
- Variância do Número de Clientes no Sistema ($\text{Var}(n)$): $\text{Var}(n) = \frac{\lambda}{(\mu - \lambda)^2}$.
- Tempo Médio de Cada Unidade no Sistema (W): $W = \frac{1}{\mu - \lambda}$.
- Número Médio de Clientes na Fila (L_Q): $L_Q = \frac{\lambda^2}{\mu(\mu - \lambda)}$.
- Tempo Médio de Cada Unidade na Fila (W_Q): $W_Q = \frac{\lambda}{\mu(\mu - \lambda)}$.

Para um estudo mais aprofundado dos diferentes tipos de disciplinas de atendimento mais utilizados, como SIRO, ou para a obtenção dos principais parâmetros dos modelos M/M/2 e M/M/3, é altamente recomendada a leitura de Ross (2014). Esta referência oferece uma visão detalhada e abrangente desses aspectos, fornecendo uma base sólida para a compreensão e análise mais aprofundada desses conceitos na teoria das filas e sistemas de atendimento.

III. METODOLOGIA

A estratégia metodológica empregada neste estudo foi adequada aos objetivos de analisar o processo de geração de filas de pacientes à espera por leitos de UTI na Rede Pública de Saúde do Distrito Federal (INFOSAÚDE, 2022) e propor melhorias no sistema para reduzir o tempo de espera e minimizar complicações para os pacientes.

Os participantes do estudo foram pacientes internados ou à espera de internação na UTI da rede pública de saúde do Distrito Federal. O período de análise compreendeu do dia 4 de fevereiro até 3 de março, durante a terceira onda de casos da pandemia de COVID-19 no DF. A população alvo incluiu 1092 pacientes observados durante o período de estudo.

Para a obtenção dos dados, devido à inexistência de uma opção de download direto, foram realizadas capturas de tela das informações. Essas imagens foram posteriormente convertidas em arquivos de texto para análise. A lista de pacientes em espera por leito de UTI foi monitorada, em média, a cada 6 horas. A atualização das informações era realizada periodicamente durante o dia pela Central de Regulação da Internação Hospitalar, com base na atualização do estado clínico do paciente.

Os dados coletados incluíram:

- Horário de chegada e saída estimado dos pacientes.
- Subtipo de leito (Coronariano, Neurocirúrgico, Clínico, Cirúrgico, Materno e Não Informado).
- Status de Suspeito/Confirmado com COVID-19.
- Necessidade de Suporte Dialítico.

A taxa de ocupação total dos leitos COVID também foi coletada diariamente durante o mesmo período.

Todos os cálculos e gráficos presentes no trabalho foram produzidos utilizando o software R versão 4.0.5. A análise dos dados seguiu os seguintes passos:

1. Verificação de Distribuição Exponencial: Para verificar se o tempo entre as chegadas possuía distribuição Exponencial, foi realizada uma análise gráfica por meio de um gráfico Q-Q. As observações distribuídas linearmente permitiram assumir que os intervalos entre as chegadas seguiam uma distribuição Exponencial.

2. Estimativas de Parâmetros:

- Foram calculados o número de pacientes que chegam por minuto e o intervalo médio entre chegadas.
- Para pacientes na fila, o último momento registrado foi considerado como o horário de atendimento, excluindo os 62 pacientes que estavam na fila na última verificação.
- Como o tempo de atendimento não estava disponível para cada paciente individualmente, um teste Kolmogorov-Smirnov de Aderência foi realizado para verificar se o número de clientes atendidos por dia seguia uma distribuição Poisson.

3. Modelagem com M/M/1:

- O modelo M/M/1 foi utilizado com taxas de chegada e atendimento de 0,02589 e 0,02601 por minuto, respectivamente.
- Os resultados incluíram a probabilidade do sistema estar vazio e a proporção do tempo em que há n pacientes na lista de espera.

4. Modelagem com M/M/2 e M/M/3:

- A adição de servidores foi modelada para avaliar a otimização do atendimento.

A coleta e análise de dados foram limitadas pela falta de opções de download direto e pela natureza irregular das atualizações dos dados. No entanto, os métodos

empregados garantiram a precisão necessária para a análise pretendida. A utilização do software R para análise e modelagem estocástica assegurou a robustez dos resultados.

Os intervalos de confiança (95%) para as estimativas foram calculados e apresentados, garantindo a confiabilidade das conclusões. As modelagens M/M/1, M/M/2 e M/M/3 forneceram informações valiosas sobre a operação do sistema e as melhorias potenciais, demonstrando a validade dos instrumentos analíticos utilizados no estudo.

IV. RESULTADOS

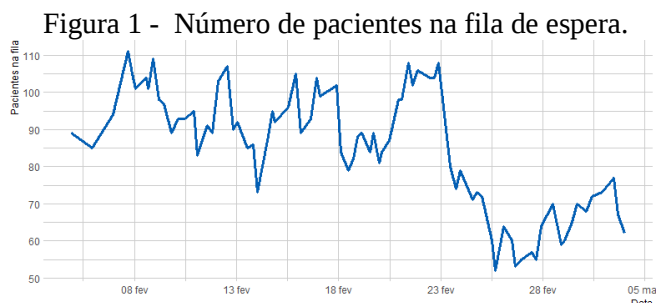
Nesse estudo de caso, foi analisado o processo de geração de filas de pacientes à espera por leitos de UTI na Rede Pública de Saúde do Distrito Federal durante um período específico. O objetivo principal foi observar o número de pessoas em espera por um leito e propor melhorias no desempenho do sistema para que o atendimento fosse efetuado mais rapidamente, diminuindo a probabilidade de perdas e complicações para os pacientes que aguardam uma vaga. De acordo com a Secretaria de Saúde, o tempo de espera começa a contar a partir do momento em que o paciente é inserido no sistema de regulação, após avaliação do quadro de saúde e diagnóstico da doença, até a disponibilidade de uma vaga para transferência.

4.1 – Coleta e Análise de Dados

Para a obtenção dos dados, foram realizadas capturas de tela das informações e, posteriormente, a conversão dessas imagens para arquivos de texto, uma vez que a opção de download dos dados não estava disponível. Todos os cálculos e gráficos presentes neste trabalho foram produzidos pelos autores utilizando o software R, versão 4.0.5.

A lista de pacientes em espera por leito de UTI no sistema público de saúde do DF foi monitorada, em média, a cada 6 horas no período de 4 de fevereiro até 3 de março, quando a terceira onda de casos da pandemia estava ocorrendo no DF. A atualização das informações era realizada periodicamente durante o dia pela Central de Regulação da Internação Hospitalar, baseada na atualização do estado clínico do paciente.

Os dados coletados incluem os tempos entre as chegadas e os atendimentos dos pacientes. Considerando que a lista de pacientes em espera foi monitorada em intervalos, os dados coletados não são completos, pois a lista é atualizada com uma periodicidade pequena e irregular, e sua versão anterior fica indisponível a cada nova atualização.



Fonte: Autores, 2022.

Para cada paciente, foram registradas as seguintes informações:

- Horário de chegada e horário de saída estimado.

- Subtipo de leito (Coronariano, Neurocirúrgico, Clínico, Cirúrgico, Materno e Não Informado).
- Status de Suspeito/Confirmado com COVID-19.
- Necessidade de Suporte Dialítico.

O posicionamento na lista de espera na UTI obedece aos critérios de prioridade (PRI) de acordo com o quadro clínico do paciente e o tipo de leito solicitado. Portanto, os que estão no topo da lista não serão necessariamente os primeiros a serem direcionados aos leitos, pois dependem do tipo de leito disponível.

Figura 2 – Fragmento da lista de pacientes aguardando leitos de UTI.

Prioridade	ID	Data de inserção	Hora de inserção	Subtipo de leito	Suspeito/Confirmado COVID-19	Suporte Dialítico
1	12957504	04/02/2022	19:42	Neurocirúrgico	1	0
1	12959116	08/02/2022	09:35	Coronariano	1	1
1	12951832	10/02/2022	15:00	Cirúrgico	1	0
1	12960228	10/02/2022	11:12	Cirúrgico	0	1
1	12961020	11/02/2022	21:13	Clínico	1	1

Fonte: Autores, 2022.

A informação da Taxa de Ocupação Total dos leitos COVID também foi coletada no mesmo período. O método de cálculo da taxa de ocupação total de leitos é o somatório de pacientes internados no dia (total de leitos ocupados) dividido pelo somatório de Leitos Operacionais no dia (total de leitos COVID menos bloqueados/aguardando liberação).

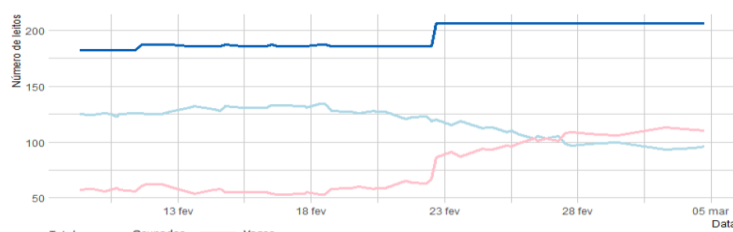
Figura 3 – Taxa de ocupação total dos leitos COVID.



Fonte: Autores, 2022.

A partir do dia 23 de fevereiro, observou-se que o valor da taxa de ocupação começou a diminuir de valores em torno de 95% para valores menores que 60% após o dia 28. Uma explicação provável para essa redução é a diminuição dos casos de COVID-19 no DF e o aumento no número total de leitos.

Figura 4 – Número de leitos COVID por dia.



Fonte: Autores, 2022.

No dia 22 de fevereiro, houve uma adição de 20 leitos na rede pública do DF e, a partir do dia 26 de fevereiro, o número de leitos vagos superou o número de leitos

ocupados. Isso pode ser explicado pela menor quantidade de novos casos diários de COVID-19 no DF e o aumento no número de leitos. Entretanto, o grande número de leitos vagos não é necessariamente tranquilizante, uma vez que os leitos vagos são compostos pelos leitos disponíveis e bloqueados.

4.2 – Aplicações do Modelo M/M/c

Na Figura 5 são apresentados os horários de chegada e o intervalo entre a chegada do paciente atual e do anterior para os últimos 10 pacientes dos 1029 que ingressaram na fila de leitos para a UTI durante o período de coleta dos dados.

Figura 5 – Horário de ingresso na Fila e Intervalo entre chegadas (em minutos).

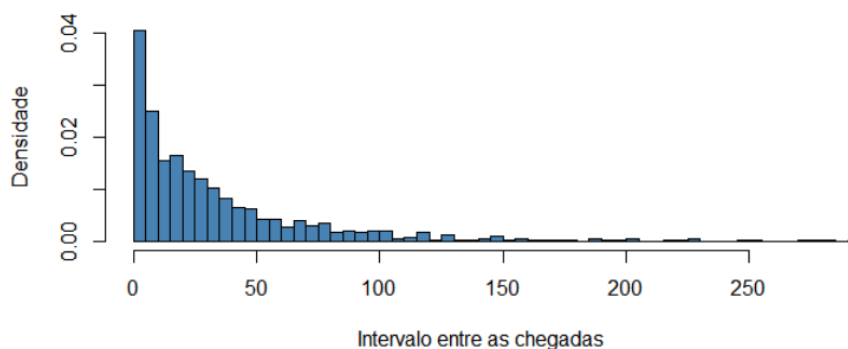
Paciente	Ingresso	Intervalo
1020	2022-03-03 20:25:00	...
1021	2022-03-03 20:32:00	7
1022	2022-03-03 20:35:00	3
1023	2022-03-03 21:15:00	40
1024	2022-03-03 21:22:00	7
1025	2022-03-03 21:49:00	27
1026	2022-03-03 22:51:00	62
1027	2022-03-03 23:24:00	33
1028	2022-03-03 23:40:00	16
1029	2022-03-04 00:17:00	37

Fonte: Autores, 2022.

O número de pacientes que chegam por minuto e o intervalo médio entre chegadas, para os 27 dias e meio analisados são 0,02589 e 38,8, respectivamente. Esses resultados indicam que chegam em média 0,02589 pacientes por minuto, ou seja, 38 pacientes por dia e o intervalo médio entre cada chegada é de 38 minutos e 48 segundos.

Antes do início da coleta de dados já havia 63 pacientes na fila com horários de chegadas disponíveis, entretanto, estes foram desconsiderados nos cálculos pois seus intervalos entre chegadas estavam incompletos, uma vez que a maior parte dos pacientes que entraram na fila em períodos próximos já haviam deixado a fila e seus dados estavam indisponíveis.

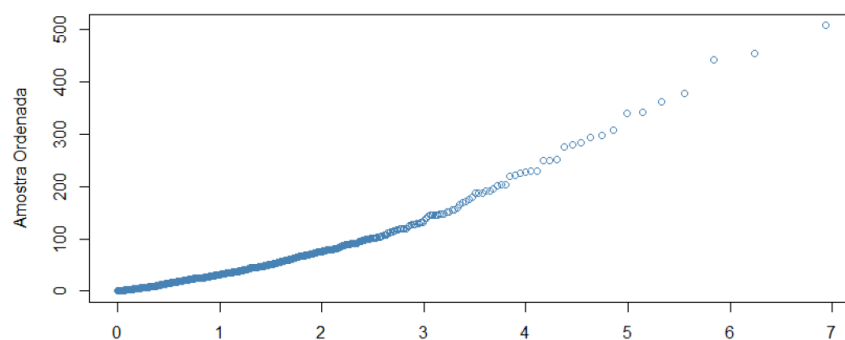
Figura 6 – Intervalos entre chegadas.



Fonte: Autores, 2022.

Para verificar se o tempo entre as chegadas possuía distribuição exponencial, foi realizada uma análise gráfica por meio de um gráfico Q-Q (MORETTIN&BUSSAB, 2017).

Figura 7 – Gráfico Q-Q da distribuição dos tempos intervalos entre chegadas.



Fonte: Autores, 2022.

Como as observações estão distribuídas linearmente, podemos assumir que os intervalos entre as chegadas seguem uma distribuição exponencial.

O número de pacientes que chegam por minuto e o intervalo médio entre as chegadas é de 0,02601 e 38,45, respectivamente. Para esse cálculo, consideramos o último momento registrado do paciente na fila como seu horário de atendimento; os 62 pacientes que estavam na fila no momento da última verificação não foram considerados atendidos.

Como o tempo de atendimento não estava disponível para cada paciente individualmente, apenas em porcentagem por categorias (até 15 dias, entre 16 e 30 dias, mais de 30 dias), foi realizado o teste Kolmogorov-Smirnov de Aderência para verificar se o número de clientes que são atendidos por dia segue uma distribuição Poisson. Com base no p-valor = 0,1076, aceita-se que os dados possam ser representados pelo modelo Poisson com média 37,45 e assumimos que o tempo de atendimento possui uma distribuição exponencial com taxa 0,02601 (MORETTIN & BUSSAB, 2017).

O modelo M/M/1 será usado nesse estudo considerando-se 0,02589 chegadas por minuto e 0,02601 atendimentos por minuto. Além disso, para a aplicação desse modelo, tem-se que a capacidade máxima do sistema e o tamanho da população são considerados infinitos; há apenas um único servidor e a ordem de atendimento segue a disciplina FIFO. A probabilidade de o sistema estar vazio e a proporção do tempo em que há n pacientes na lista de espera são 0,0046 e $(0,0046)(0,9954)^n$, respectivamente.

Os valores encontrados para cada medida de desempenho foram os seguintes:

- Taxa de ocupação média: 0,9954;
- Número médio de clientes no sistema: 216,84;
- Probabilidade de se ter 00 ou mais pacientes na fila: 0,631;
- Tempo médio de cada unidade no sistema (minutos): 8375,172, ou seja, 139,5862 horas ou 5 dias, 19 horas e 35 minutos;
- Tempo médio de cada unidade na Fila: 8336,726, o que equivale a 138,9454 horas ou 5 dias, 18 horas e 57 minutos.
- Número médio de clientes na fila: 216.

Considerando um nível de confiança de 95%, os intervalos de confiança para as estimativas do tempo de espera médio na fila, taxa de ocupação média e tamanho médio da fila foram calculados e são apresentados na Figura 8.

Figura 8 – Intervalos de Confiança para as estimativas do modelo M/M/1.

Estimativa	Limite Inferior	Limite Superior
Tempo de espera médio na fila (horas)	138,8493	139,0415
Tamanho médio da fila	184,2395	247,4478
Número médio de pacientes no sistema	176,6874	256,9908
Tempo médio de cada paciente no sistema (horas)	127,4593	151,7131

Fonte: Autores, 2022.

O tempo de espera médio é muito alto para os pacientes que esperam por vagas na unidade de tratamento intensiva, o que pode gerar complicações para o paciente e um aumento no seu tempo de internação. Para reduzir o tempo na fila e diminuir a chance de depreciação da saúde do paciente, maximizando o atendimento, uma parametrização em que existam dois ou três servidores é proposta na próxima seção.

O aumento no número de servidores neste caso pode ser considerado como uma contratação de leitos que não são parte integrante da rede SES/DF, estes sendo de uma instituição privada que participa de forma complementar ao Sistema Único de Saúde, segundo diretrizes deste, mediante contrato público de direito ou convênio, tendo preferência as entidades filantrópicas e as sem fins lucrativos (Manual de Orientações para Contratação de Serviços no Sistema Único de Saúde, 2007).

Sob a proposta de um segundo guichê para o atendimento dos pacientes, temos que a probabilidade de atendimento imediato e a proporção do tempo em que existam n pacientes no sistema, são dados: por $P_0=0,3354$, $P_n= 0,3354(0,02589)^n/n!(0,02601)^n$, se $n < 2$ e $P_n= (0,02589)^n/2^{n-2}n!(0,02601)^n$, se $n \geq 2$.

O resultado de P_0 indica que em 33,54% do tempo o sistema fica vazio, e os atendentes dos guichês ficam ociosos. As medidas de desempenho para esse modelo são:

- Número médio de clientes no sistema: 0, 328;
- Tempo médio de cada unidade no sistema (minutos): 51, 106;
- Tempo médio de cada unidade na Fila: 12, 66;
- Número médio de clientes na fila: 1,323.

Considerando um nível de confiança de 95% os intervalos de confiança para as estimativas são apresentados na Figura 9.

Figura 9 – Intervalos de Confiança para as estimativas do modelo M/M/2.

Estimativa	Limite Inferior	Limite Superior
Tempo de espera médio na fila (minutos)	6,1775	19,1415
Tamanho médio da fila	0	31,932
Número médio de pacientes no sistema	0	41,4749
Tempo médio de cada paciente no sistema (horas)	0	12,97864

Fonte: Autores, 2022.

Esses resultados atestam que a adição de um novo servidor otimiza o atendimento e reduz o tempo de espera na fila pelos pacientes na unidade de tratamento intensivo. O número médio de usuários na fila passaria de 216 para menos de 32 e o número médio de usuários no sistema é pouco maior que um na estimativa pontual. Além disso, o tempo médio de espera na fila não seria superior a 20 minutos (com 95% de confiabilidade). No modelo M/M/1 esse tempo era superior a 5 dias.

Se considerarmos um sistema com três servidores, podemos mostrar que em 36,54% do tempo o sistema fica vazio, e que os intervalos de confiança (95%) para as estimativas são aqueles apresentados na Figura 10.

Figura 10 – Intervalos de Confiança para as estimativas do modelo M/M/3.

Estimativa	Limite Inferior	Limite Superior
Tempo de espera médio na fila (minutos)	0	9,2125
Tamanho médio da fila	0	31,6465
Número médio de pacientes no sistema	0	38,5652
Tempo médio de cada paciente no sistema (horas)	0	12,8251

Fonte: Autores, 2022.

Com isso, é percebido que um sistema com três servidores não acarretaria melhoras significativas no sistema, quando comparados com um que possua dois servidores. Uma vez que as estimativas pontuais e intervalares para as medidas de desempenho são muito próximas.

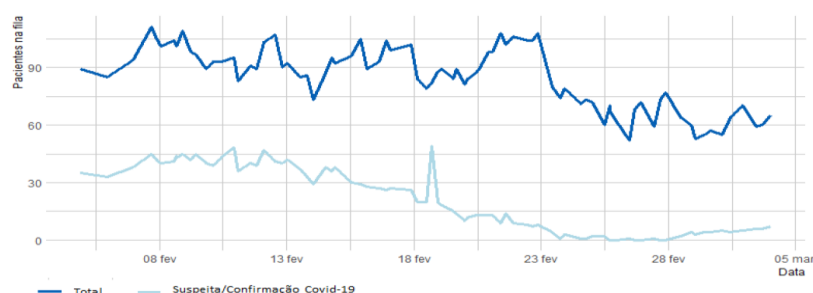
4.3 – Pacientes com Suspeita ou Confirmação de Covid-19

Os leitos de UTI COVID-19 são aqueles com Suporte Ventilatório Pulmonar, seguindo requisitos de Resoluções de Diretoria Colegiada da ANVISA (RDC Nº 07, de 24 de fevereiro de 2010) e de Portarias do Ministério da Saúde. A quantidade de leitos é variável ao longo do tempo, de acordo com flutuações no número de casos e da necessidade de leitos.

A internação por Síndrome Respiratória Aguda Grave em leitos de UTI é feita quando existem indícios de que o pulmão do paciente não está conseguindo oxigenar o sangue e os órgãos de forma conciliável com a vida em curto prazo.

As Unidades de Terapia Intensiva têm sido um instrumento de grande destaque na recuperação de pacientes internados por Covid-19. Sabendo que existe uma população de risco, com maior predisposição à demanda de suporte ventilatório, faz-se necessária a internação em UTI, visando o cuidado periódico dessas vidas (F. CARVALHO; ELIAS; T. CARVALHO, 2021). No período observado, cerca de 23,6% dos pacientes aguardando em fila possuíam suspeita ou confirmação do vírus SARS-CoV2.

Figura 11 – Número de pacientes na fila com suspeita ou confirmação de Covid-19.



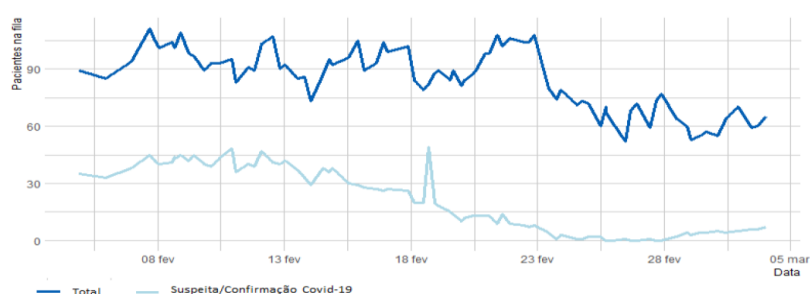
Fonte: Autores, 2022.

Pela Figura 11, é perceptível que o número de pacientes na fila diminuiu de forma similar ao número de novos casos de Covid-19 no DF. Em relação ao total de pacientes em espera por leitos de UTI, no período de 4 a 18 de fevereiro, a proporção de pacientes com confirmação ou suspeita da doença era maior, e à medida que os casos diminuíram, o tamanho da fila também reduziu.

O número de pacientes com suspeita ou confirmação de Covid-19 que chegam por minuto e o intervalo médio entre suas chegadas, para os 27 dias e meio analisados, foram 0,0065 e 0,118, respectivamente. Esses resultados indicam que em média chegam 0,0065 pacientes por minuto, ou 9 pacientes por dia, ou em média 9 dos 38 pacientes totais que chegam. Sendo assim, a cada 10 pacientes que ingressam, 2,37 possuem no mínimo a suspeita da doença. O intervalo médio entre cada chegada é de 0,108 dias, ou 2 horas e 35 minutos.

Antes do início da coleta de dados, já havia 34 pacientes na fila com horários de chegadas disponíveis. Entretanto, foram desconsiderados nos cálculos acima, pois seus intervalos entre chegadas estavam incompletos, uma vez que a maior parte dos pacientes que entraram na fila em períodos próximos já haviam deixado a fila e seus dados estavam indisponíveis.

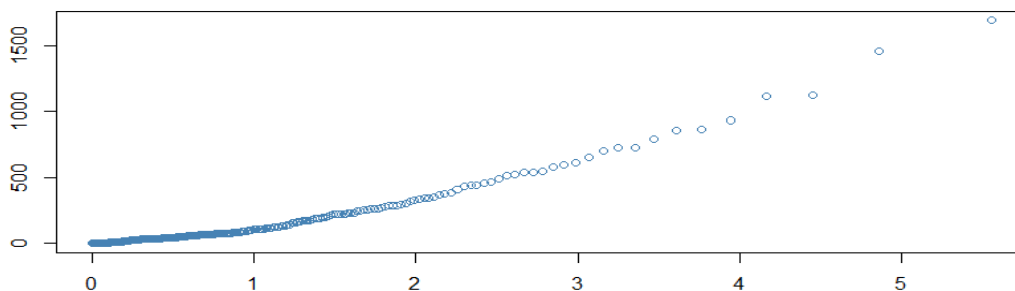
Figura 12 – Comportamento dos intervalos entre chegadas.



Fonte: Autores, 2022.

O teste Kolmogorov-Smirnov de Aderência foi realizado para verificar se o intervalo entre as chegadas possui distribuição exponencial. Para um nível de significância de 10%, o valor da estatística de teste foi de 0,183, portanto, maior que o valor crítico. Logo, aceita-se a hipótese de que o intervalo entre as chegadas têm distribuição exponencial.

Figura 13 – Gráfico Q-Q da distribuição dos intervalos entre chegadas.



Fonte: Autores, 2022.

Além disso, ao analisar o gráfico Q-Q, como as observações se apresentam de forma linear, podemos assumir que a distribuição dos intervalos é a mesma da distribuição teórica utilizada, ou seja, pode-se consentir que os intervalos entre as chegadas seguem uma distribuição Exponencial.

O número médio de pacientes atendidos por minuto foi de 0,0072. Para se obter esse valor, considerou-se o último momento registrado do paciente na fila como seu horário de atendimento, os 7 pacientes que estavam na fila no momento da última verificação não foram considerados atendidos.

Aplicando os modelos M/M/1, M/M/2 e M/M/3, onde o aumento no número de servidores neste caso pode ser considerado como um ou dois hospitais de campanha com novos leitos para pacientes com Covid-19, obtém-se as seguintes estimativas de desempenho apresentadas na Figura 16.

Figura 14 – Estimativas das medidas de desempenho por modelo aplicado.

Estimativa	Símbolo	M/M/1	M/M/2	M/M/3
Número médio de pacientes na fila	$\hat{L}_Q$	8,4	0,23	0,06
Tempo médio de cada paciente na fila (minutos)	$\hat{W}_Q$	1289,7	35,54	9,35
Número médio de pacientes no sistema	$\hat{L}$	9,3	1,13	0,96
Tempo médio no sistema (minutos)	$\hat{W}$	1428,6	174,43	148,24

Fonte: Autores, 2022.

Comparando as estimativas pontuais dos parâmetros de desempenho dos modelos M/M/1 e M/M/2, percebe-se que há uma melhora significativa no tempo médio de espera, diminuído de 21 horas e 38 minutos para cerca de 36 minutos. Assim como uma diminuição na quantidade média de pacientes na fila. Em relação ao modelo M/M/3, o tempo médio de cada paciente na fila diminuiu em apenas 26 minutos. Comparado com o M/M/2, do ponto de vista da economicidade, a redução não foi significativa.

Figura 15 – Intervalos de Confiança de 95% para as estimativas médias dos modelos.

Estimativa	IC M/M/1	IC M/M/2	IC M/M/3
$\hat{L}_Q$	[0; 41,4]	[0; 33,2]	[0; 33,02]
$\hat{W}_Q$	[1282; 1297,4]	[27,9; 43,2]	[1,7; 17,02]
$\hat{L}$	[0; 46,9]	[0; 32,1]	[0; 31,9]
$\hat{W}$	[1285,5; 1571,6]	[31,4; 317,5]	[5,2; 291,3]

Fonte: Autores, 2022.

Ao se analisar os intervalos de confiança, percebe-se que o tempo de espera médio em fila que era superior a 21,3 horas no modelo M/M/1, não supera 43,2 minutos no modelo com dois servidores. O tempo de espera médio no sistema com um servidor era maior que 21,4 horas, e no M/M/2 é menor que 5,3 horas.

O tempo médio de espera na fila com dois servidores, pelo intervalo de confiança, era inferior a 44 minutos. Acrescentando-se mais um servidor, esse tempo foi reduzido para menos de 17 minutos, aproximadamente. Essa melhora não é tão significativa considerando-se a diferença de mais de 20 horas entre os modelos M/M/1 e M/M/2.

Além disso, também é perceptível que um modelo em que existam três servidores disponíveis não é significativamente vantajoso do que o modelo com apenas dois servidores, quando se observa o número médio de pacientes. Na estimativa pontual, esse valor diminuiu de 1,13 para 0,96. Enquanto isso, os intervalos de 95% de confiança são praticamente idênticos.

V. CONCLUSÃO

Este estudo examinou a geração e gestão de filas de espera para leitos de UTI na rede pública de saúde do Distrito Federal durante a terceira onda da pandemia de COVID-19, utilizando a Teoria de Filas como principal referencial teórico. Os dados analisados, coletados entre 4 de fevereiro e 3 de março, indicaram que o sistema atual

operava sob um regime de alta ocupação e significativa espera, gerando complicações para os pacientes.

A aplicação dos modelos M/M/1, M/M/2 e M/M/3 demonstrou que o sistema atual (M/M/1) apresentava uma taxa de ocupação média extremamente alta (99,54%), com um tempo médio de espera na fila de aproximadamente 5 dias. A análise sugeriu que a introdução de um segundo servidor (modelo M/M/2) poderia reduzir drasticamente o tempo médio de espera na fila para menos de 20 minutos. A adição de um terceiro servidor (modelo M/M/3) mostrou apenas uma melhora marginal, não justificando o custo adicional.

Observou-se que 23,6% dos pacientes na fila tinham suspeita ou confirmação de COVID-19, e a diminuição dos casos de COVID-19 levou a uma redução proporcional no número de pacientes na fila de espera.

Utilizando ferramentas de análise gráfica e testes de aderência, confirmou-se que os intervalos entre chegadas dos pacientes seguem uma distribuição exponencial, validando a aplicação dos modelos Markovianos.

A coleta de dados foi realizada manualmente através de capturas de tela, o que pode introduzir erros e limitações na completude e precisão dos dados. A atualização irregular e não sistemática das listas de espera limitou a análise temporal contínua e detalhada dos fluxos de pacientes.

As conclusões deste estudo são específicas ao contexto da pandemia de COVID-19. No entanto, surtos de viroses, como os de dengue, e as altas demandas por serviços hospitalares após recentes desastres climáticos, como os ocorridos no Rio Grande do Sul e no Rio de Janeiro, evidenciam a importância de estudar e modelar as filas no SUS. Vale destacar que a aplicabilidade dos resultados pode variar em outros contextos ou sistemas de saúde que possuam características e recursos diferentes.

A falta de dados específicos sobre o tempo de atendimento para cada paciente obrigou a utilização de uma distribuição Poisson para a modelagem, o que pode não representar perfeitamente a realidade.

A implementação de sistemas de coleta de dados automatizados para garantir a precisão e completude dos dados, facilitando análises mais detalhadas e contínuas. Também, a realização de estudos comparativos em diferentes contextos e sistemas de saúde para avaliar a aplicabilidade dos modelos e propostas de melhorias sugeridas.

Seria interessante integrar análises econômicas para avaliar a viabilidade das propostas de adição de servidores, considerando os custos associados e os benefícios em termos de redução de tempo de espera e melhoria na saúde dos pacientes.

Em trabalhos futuros, almeja-se também explorar outros modelos de filas e distribuições de tempo de serviço para verificar se resultados mais precisos e adequados podem ser obtidos para diferentes realidades.

Em síntese, este estudo contribui significativamente para a compreensão e melhoria do sistema de gestão de filas de espera para leitos de UTI durante situações de alta demanda, como pandemias, destacando a importância da aplicação da Teoria de Filas para otimização do atendimento em sistemas de saúde públicos.

VI. REFERÊNCIAS

BRASIL, **Centro de Operações de Emergência em Saúde Pública**, Governo do Distrito Federal. “Plano de Mobilização de Leitos COVID-19 do Distrito Federal”, 2021.

BRASIL, **Ministério da Saúde**, Departamento de Regulação, Avaliação e Controle de Sistemas. “Manual de Orientações para Contratação de Serviços no Sistema Único de Saúde”, 2007.

BRESSAN, G. “Modelagem e Simulação de Sistemas Computacionais - Sistemas de Filas Simples”, **LARC-PCS/EPUSP**, 2002.

CAMELO, G. R., COELHO, A. S., BORGES, R. M., SOUZA, R. M. “Teoria das Filas e da Simulação Aplicada ao Embarque de Minério de Ferro e Manganês no Terminal Marítimo de Ponta da Madeira”, **Cadernos do IME**, 2010.

ERLANG, A. K. “The Theory of Probabilities and Telephone Conversations”, **Nyt Tidsskrift for Matematik**, 1909.

INFOSAÚDE, **Secretaria de Saúde do Distrito Federal**, 2022. Disponível em <https://info.saude.df.gov.br/leitospublicosutigeraissalasit/>

KENDALL, D. G. “Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Embedded Markov Chain”, **The Annals of Mathematical Statistics**, 1953.

MORAES, V., VENTURINI, R., TAVARES, M., de SÁ, M., SALMER, W. "Boletim Epidemiológico: Análise da Média de Permanência UTI Adulto Município de Jataí", Observatório de Epidemiologia e Serviços de Saúde - **EpiServ**, 2018.

MENDONÇA, E. B. D. (2014). Teoria de filas Markovianas e Aplicações. (Tese de doutorado). Orientadora: Divanilda Maia Esteves, v. 63.

MOREIRA, L.; BASTOS, E. F. "Study of the System of Files of Transplant of Organs in SUS", **Revista Sodebras [on line]**, v. 12, p. 151, 2017.

MORETTIN, P. A.; BUSSAB, W. O. (2017). Estatística Básica. [S.l.]: **Saraiva Educação SA**.

MOREIRA, L.; BASTOS, E. F. “How to Decrease Waiting Time in the Queue of Transplantation of Organs in SUS”, **Revista Sodebras [on line]**, v. 13, p. 105-109, 2018.

PERDONÁ, I. I., NUNES, R. V., NEVES, R. M., NAIMER, S. C., GODOY, L. P. "Sistema de Manufatura: Otimização de Processos em uma Unidade Fabril de Cimento Através da Teoria das Filas", **Exacta**, 2017.

ROSS, S. M. “Introduction to Probability Models”, **Elsevier**, 2014.

VII. AGRADECIMENTOS

Agradeço a Deus, à minha orientanda, à minha família e à Universidade de Brasília (UnB) pelo apoio e incentivo durante a realização deste trabalho científico. Suas contribuições foram fundamentais para o desenvolvimento deste estudo.

VIII. COPYRIGHT

Direitos autorais: O(s) autor(es) é(são) o(s) único(s) responsável(is) pelo material incluído no artigo.