

ATENDIMENTO EM SAÚDE À LUZ DA TEORIA DAS FILAS

JUNIOR, Homero da Silva Nahum^{1,2}; BARRETO, Ana Cristina Lopes y Glória²; BRASIL, Roxana Macedo³; BRITO, Diogo de Freitas^{1,4}

113

Resumo

A investigação objetivou caracterizar serviços de saúde com configurações de atendimento sem (M/M/1) e com (M/M/1/K) restrição ao atendimento, em clínica médica da Cidade do Rio de Janeiro, considerando somente uma especialidade. Os resultados indicaram ajuste exponencial para os tempos de Chegada (λ) e Atendimento (μ). O M/M/1 tinha fator de utilização de 75,00%, para valores médios de 3,00 pacientes no sistema, 2,25 pacientes na fila, e 1,00h demanda pelo sistema, incluindo 45 minutos na fila. A limitação de cinco pessoas nas dependências (M/M/1/5) proporcionou similar ociosidade, porém com 1,70 pacientes no sistema, aguardando 48,89 minutos no total, sendo 28,88 minutos na fila com, em média, 1,00 paciente. A conclusão foi que os modelos corresponderam ao objetivo.

Palavras-chave: Pesquisa operacional. Processo estocástico. Planejamento. Administração. Negócio.

Abstract

The research aimed to characterize health services with service configurations without (M/M/1) and with (M/M/1/K) service restrictions, in a medical clinic in the city of Rio de Janeiro, considering only one specialty. The results indicated exponential adjustment for the Arrival (λ) and Service (μ) times. M/M/1 had a utilization factor of 75.00%, for average values of 3.00 patients in the system, 2.25 patients in the queue, and 1.00h demand for the system, including 45 minutes in the queue. The limitation of five people in the facilities (M/M/1/5) provided similar idleness, but with 1.70 patients in the system, waiting 48.89 minutes in total, with 28.88 minutes in the queue with, on average, 1.00 patients. The conclusion was that the models corresponded to the objective.

Keywords: Operational research. Stochastic process. Planning. Administration. Business.

Introdução

A Teoria das Filas (TF) seria empregada em áreas diversas como comércio (Leite, 2024; Almeida *et al.*, 2024), serviço (Simoyama, Diniz e Battisti, 2022; Pincaço *et al.*, 2024; Mathias *et al.*, 2024; Nassy e Simões, 2024; Assis, Quedas e Silva, 2024), automação (Girão *et al.*, 2024; Abensur *et al.*, 2003; Pereira, 2023), indústria (Melo, 2024), transporte

¹ Docente da Escola de Saúde da Universidade Candido Mendes;

² Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

³ Docente Ph.D. em Educação Física;

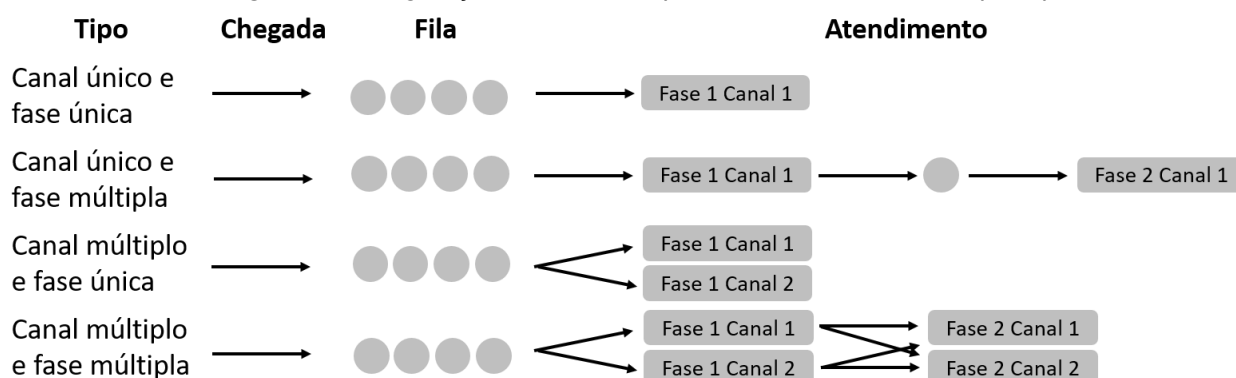
⁴ Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa.

(Silva, Costa e Jesus, 2024; Andrade *et al.*, 2024) e saúde (Lima e Villela, 2024; Freitas *et al.*, 2024).

Historicamente, a primeira investigação sobre TF foi de Agner Krarup Erlang para redimensionamento da *Copenhagen Telephone Company* (Dinamarca), que apresentava volumoso congestionamentos nas centrais, o que, também, culminou na determinação do número de operadores necessários ao serviço (Erlang, 1925; Erlang, 1909) e os tempos de espera e total (Erlang, 1917).

A TF seria o método estatístico para estimar os tempos de demora, total e de atendimento, inerentes à prestação de serviço, quando clientes externos chegariam ao acaso (Torres, 1966), e sendo a demanda maior que a capacidade de atendimento. Portanto, a modelagem objetivaria encontrar a viabilidade do serviço, evitando gargalos ou ineficiências, e satisfazer a exigência do cliente externo (Abensur, 2018), independentemente da configuração da fila (Figura 1).

Figura 1: Configurações de Fila. Adaptado de Heizer e Render (2001).



Fonte: Os Autores (2025).

Essa consideraria, comumente, população infinita, capacidade ilimitada do sistema e disciplina de atendimento como Primeiro a Entrar – Primeiro a Sair (PEPS ou FIFO: *First in – First On*), portanto as estimativas tomariam as distribuições de probabilidade (exponencial ou Poisson) dos tempos entre as chegadas e de serviço, e o número de canais (atendentes), os quais seriam denominados por Processo de Chegada, Tempo de Atendimento e Canais (Harada, 2017), respectivamente, levando o modelo a ser representado e estimado, conforme a existência de um único atendente (estágio), M/M/1, ou vários, M/M/m, pelas fórmulas constantes no Quadro I (Abensur, 2018). Independentemente da configuração, os tempos de chegada e atendimento seriam aleatórios e exponenciais, com médias dadas por $E(X) = 1/\lambda$ e $E(S) = 1/\mu$, nessa ordem.

Quadro I: Fórmulas às Estimativas do Modelo M/M/1 e M/M/m.

Medida	M/M/m	M/M/1
Número médio de usuários no sistema (L)	$L = \frac{\lambda \mu \left(\frac{\lambda}{\mu}\right)^m}{(m-1)!(m\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$	$L = \frac{\lambda}{\mu - \lambda}$
Número médio de usuário na fila (L _q)	$L_q = L - \frac{\lambda}{\mu}$	$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Tempo médio do usuário no sistema (W)	$W = \frac{L}{\lambda}$	$W = \frac{1}{\mu - \lambda}$
Tempo médio do usuário na fila (W _q)	$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda}$	$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$
Probabilidade de haver nenhum usuário no sistema (P ₀)	$P_0 = \frac{1}{\left(\sum_{n=0}^{m-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n\right) + \frac{1}{m!} \left(\frac{\lambda}{\mu}\right)^m \frac{m\mu}{m\mu - \lambda}}$ $\forall m\mu > \lambda$	$P_0 = 1 - \frac{\lambda}{\mu}$
Probabilidade de haver n usuários no sistema (P _n)	$P_n = \begin{cases} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} P_0; 0 \leq n \leq m \\ \frac{\left(\frac{\lambda}{\mu}\right)^n}{m! m^{n-m}}; n > m \end{cases}$	$P_n = \rho^n (1 - \rho)$
Fator de utilização do sistema ou Ocupação do sistema (ρ)	$\rho = \frac{\lambda}{m\mu}$	$\rho = \frac{\lambda}{\mu}$

Fonte: Adaptado de Abensur (2018).

Com base no exposto, possível seria deduzir que número de clientes externos existentes, tamanho da fila e tempos de espera no sistema e na fila em determinado instante t seriam variáveis aleatórias dependentes do estado inicial t_0 . Dito de outra maneira, a probabilidade no momento t dependeria da probabilidade em t_0 . Logo, a fila se caracterizaria como processo estocástico, sequência de variáveis aleatórias indexadas ao tempo (Jandre *et al.*, 2024; Akkari, 2018; Clarke e Disney, 1979).

Correntemente, a influência de t_0 desapareceria quando t tendesse ao infinito ($t \rightarrow \infty$), significando que o sistema alcançou o equilíbrio (regime permanente). Nesse, as probabilidades também não seriam alteradas pelo tempo decorrido (Taylor e Karlin 1998; Torres, 1966). Todo sistema que conquista tal propriedade (situação de equilíbrio) seria denominado ergódico (Jones e Smith, 2009; Kovács, 1996). A vantagem dessa característica residiria na possibilidade de utilização do equilíbrio para estimar a probabilidade de qualquer outro estado (Torres, 1966).

Quadro II: Fórmulas às Estimativas do Modelo M/M/1/K e M/M/m/K

Medida	M/M/m/K	M/M/1/K
Número médio de usuários no sistema (L)	$L = \frac{\lambda(1 - P_n)}{\mu} + L_q$	$L = \begin{cases} \frac{\rho}{(1 - \rho)} - \frac{(K + 1)\rho^{K+1}}{(1 - \rho^{K+1})}; \rho \neq 1 \\ \frac{K}{2}; \rho = 1 \end{cases}$
Número médio de usuário na fila (L _q)	$L_q = \begin{cases} \frac{(\rho m)^m \rho}{m! (1 - \rho)^2} [1 - \rho^{K-m+1} - (K - m + 1)\rho^{K-m}(1 - \rho)] \\ \frac{(\rho m)^m (K - m + 1)(K - m)}{2m!} P_0; \rho = 1 \end{cases}$	$L_q = \begin{cases} \frac{\rho}{(1 - \rho)} - \frac{K\rho^{K+1} + \rho}{(1 - \rho^{K+1})}; \rho \neq 1 \\ \frac{K(K - 1)}{2(K + 1)}; \rho = 1 \end{cases}$
Tempo médio do usuário no sistema (W)	$W = \frac{L}{\lambda(1 - P_0)}$	$W = \frac{L}{\lambda(1 - P_0)}$
Tempo médio do usuário na fila (W _q)	$W_q = \frac{L_q}{\lambda(1 - P_0)}$	$W_q = \frac{L_q}{\lambda(1 - P_0)}$
Probabilidade de haver nenhum usuário no sistema (P ₀)	$P_0 = \begin{cases} \frac{1}{\sum_{n=0}^{m-1} \frac{(\rho m)^n}{n!} + \frac{(\rho m)^m}{m!} \cdot \frac{1 - \rho^{K-m+1}}{1 - \rho}}; \rho \neq 1 \\ \frac{1}{\sum_{n=0}^{m-1} \frac{(\rho m)^n}{n!} + \frac{(\rho m)^m}{m!} (K - m + 1)}; \rho = 1 \end{cases}$	$P_0 = \begin{cases} \frac{1 - \rho}{1 - \rho^{K+1}}; \rho \neq 1 \\ \frac{1}{K + 1}; \rho = 1 \end{cases}$
Probabilidade de haver n usuários no sistema (P _n)	$P_n = \begin{cases} \frac{(\rho m)^n}{n!} P_0; n = 1, 2, \dots, m - 1 \\ \frac{\rho^n m^m}{m!} P_0; n = m, m + 1, \dots, K \end{cases}$	$P_n = \begin{cases} \frac{\rho^n (1 - \rho)}{1 - \rho^{K+1}}; \rho \neq 1, n = 0, 1, \dots, K \\ \frac{1}{K + 1}; \rho = 1, n = 0, 1, \dots, K \\ 0; n > K \end{cases}$
Fator de utilização do sistema ou Ocupação do sistema (ρ)	$\rho = \frac{\lambda}{m\mu}$	$\rho = \frac{\lambda}{\mu}$

Fonte: Adaptado de Abensur (2018).

As especificações matemáticas anteriores consideraram que as populações atendidas seriam infinitas. Porém, tal consideração poderia não ser razoável, correntemente, filas em saúde teriam limitação de *k* usuários no sistema, por exemplo, nos casos em que houvesse distribuição de determinado número de senhas ou o espaço físico

fosse limitado. Em resultância, a fila teria $k-1$ por tamanho máximo (Akkari, 2018), como ocorreria em consultórios médicos, odontológicos, terapêuticos ou de nutricionistas particulares. Como consequência, o modelo seria representado e estimado pelas fórmulas do Quadro II, para um único atendente, $M/M/1/K$, ou vários, $M/M/m/K$ (Abensur, 2018). Com base no exposto, a presente investigação objetivou modelar duas filas em serviços de saúde com atendimentos infinitos e limitados.

Metodologia

Uma clínica médica situada na zona sul da Cidade do Rio de Janeiro, forneceu os tempos de chegada e atendimento de pacientes nos últimos 18 meses, de uma única especialidade com apenas um canal, profissional atendendo. A investigação sobre o ajuste a distribuições de probabilidade revelou a Exponencial como adequada em ambos os casos, o que possibilitou a estimativas dos tempos médios de Chegada (λ) e Atendimento (μ). A partir dessas o modelo real foi obtido o modelo real ($M/M/1$) e feitas simulações considerando dois e três profissionais, bem como limitação de pacientes nas dependências da instituição.

Resultados e Discussão

A Situação Real apresentava taxa de Ocupação do sistema (ρ) de 75,00%, tendo somente 25,00% de chance de inexistência de paciente no sistema (P_0), isso tendo em média três indivíduos no sistema (L) e 2,25 pessoas na fila (L_q), o que determinaria a permanência média no sistema de uma hora (W) e 0,75 h ($W_q = 45,00$ minutos) na fila (Tabela 1).

Essas características poderiam determinar espera excessiva pelo paciente. Todavia, as Simulações $M/M/2$ e $M/M/3$ indicaram subutilização dos sistemas (Tabela 1), $\rho = 37,50\%$ e $\rho = 25,00\%$ nessa ordem, portanto a ociosidade seria demasiadamente elevada, indicando desequilíbrio entre custo e efetividade. Esse indicado seria corroborado pela possível existência de menos de um paciente no sistema, $L_{M/M/2} = 0,8727$ e $L_{M/M/3} = 0,7647$, e nenhum na fila, $L_q < 1,00$. O tempo médio dispendido no sistema seria 0,2909h (17,45 minutos) e 0,2549h (15,29 minuto).

Em qualquer das situações, a probabilidade de haver, um número de pacientes superior a três seria inferior aos 10,00%. A existência de exatamente três pessoas teria

chance de 10,55%, exceto quando houvesse trio de profissionais ($P_3 \cong 3,31\%$). Em suma, conforme expectativa, o aumento da quantidade de atendentes promoveria decaimento acentuado da probabilidade de permanência de pacientes em número crescente no sistema (Tabela 1), ainda assim, a Situação Real seria a adequada ao serviço, dadas as condições reinantes.

Tabela 1: Resultados da Modelagem sem Limite de Atendimentos. Tempo Médio de Chegada (λ) < Tempo Médio de Atendimento (μ) para Situação Real (M/M/1) e Simulações (M/M/2 e M/M/3).

Características	M/M/1	M/M/2	M/M/3
m, nº	1	2	3
λ , paciente/h	3	3	3
μ , paciente/h	4	4	4
L, pacientes	3,0000	0,8727	0,7647
Lq, pacientes	2,2500	0,1227	0,0147
W, h	1,0000	0,2909	0,2549
Wq, h	0,7500	0,0409	0,0049
ρ , %	0,7500	0,3750	0,2500
P_0 , %	0,2500	0,4545	0,4706
P_n , %			
1	0,1875	0,3409	0,3529
2	0,1406	0,1278	0,1324
3	0,1055	0,1055	0,0331
4	0,0791	0,0396	0,0176
5	0,0593	0,0148	0,0044
6	0,0445	0,0056	0,0011
7	0,0334	0,0021	0,0003
8	0,0250	0,0008	0,0001
9	0,0188	0,0003	0,0000
10	0,0141	0,0001	0,0000

Fonte: Os Autores (2025).

O segundo grupo de situações simuladas considerou a igualdade entre os tempos médios de Chegada e Atendimento, $\lambda = \mu = 4,00$ pacientes/hora, dito de outra maneira, investigou-se a possibilidade de a capacidade dos profissionais atenderem adequadamente aos pacientes ser análoga ao regime com que eles chegariam à clínica (Tabela 2). Nessas condições, a Situação Real (M/M/1) atuaria na capacidade máxima ($\rho = 100,00\%$), com nula possibilidade de ociosidade. Todas as demais características não seriam possíveis de serem estimadas e representadas por 0,0000% ou divisão por zero (-).

Continuaria subutilizado o sistema M/M/3, $\rho = 33,33\%$, novamente, tendo menos de duas pessoas no sistema ($L = 1,0455$ pacientes) e tempo nesse de 0,2614h ($W = 15,68$

minutos), por consequência a fila tenderia a demandar pouco tempo ($L_q = 0,0455h = 2,73$ minutos). As estimativas de P_n decairiam mais lentamente do que aquela constante na Tabela 1. Na modelagem para M/M/2, o comportamento de P_n convergiria ao descrito, porém a ociosidade seria de 50,00% (ρ) com média de 1,3333 pacientes no sistema e 0,3333 pacientes na fila, aguardando respectivamente, $W = 20,00$ minutos (0,3333h) e $W_q = 5,00$ minutos (0,0833h). Em suma, para $\lambda = \mu = 4,00$ pacientes/hora, a contratação do segundo especialista seria necessária e suficiente à relação custo – efetividade.

Tabela 2: Resultados da Modelagem sem Limite de Atendimentos. Testando Tempo Médio de Chegada (λ) = Tempo Médio de Atendimento (μ) para Situação Real (M/M/1) e Simulações (M/M/2 e M/M/3).

Características	M/M/1	M/M/2	M/M/3
m, nº	1	2	3
λ , paciente/h	4	4	4
μ , paciente/h	4	4	4
L, pacientes	-	2,5263	0,3365
L_q , pacientes	-	1,5263	-0,6635
W, h	-	0,6316	0,0841
W_q , h	-	0,3816	-0,1659
ρ , %	1,0000	0,5000	0,3333
P_0 , %	0,0000	0,3333	0,3636
P_n , %			
1	0,0000	0,3333	0,3636
2	0,0000	0,1667	0,1818
3	0,0000	0,2500	0,0606
4	0,0000	0,1250	0,0556
5	0,0000	0,0625	0,0185
6	0,0000	0,0313	0,0062
7	0,0000	0,0156	0,0021
8	0,0000	0,0078	0,0007
9	0,0000	0,0039	0,0002
10	0,0000	0,0020	0,0001

Fonte: Os Autores (2025).

As simulações foram estendidas ao quadro no qual o Tempo Médio de Chegada (λ) superaria o Tempo Médio de Atendimento (μ), quando o Sistema M/M/1 estaria colapsado, pois o Fator de Ocupação do Sistema = 125,00% > 100,00%, acarretando na negatividade das demais estimativas (Tabela 3). Talvez, o modelo com três profissionais (M/M/3) ainda seria excessivo, pois a ociosidade (58,33%) superaria a ocupação (41,67%), em média, aproximadamente, um paciente estaria no sistema (L) e menos de um na fila (L_q), o que culminaria em 16,33 minutos ($W = 0,2722h$) e 1,33 minutos ($W_q = 0,0222h$),

respectivamente. A maior probabilidade de pacientes no sistema seria de 34,83% correspondente à P_1 . Essa foi de 28,85% no modelo M/M/2, o qual apresentaria 62,50% de ocupação, portanto a probabilidade de tempo ocioso seria de 37,50%, o que indicaria que dois especialistas seria a quantidade adequada. Hipótese ratificada pelas médias de pessoas na fila ($L_q = 0,8013$ pacientes) e no sistema ($L = 2,053$ pacientes), e tempos médios de espera total, $W = 0,4103h$ (24,62 minutos), e na fila, $W_q = 0,1603h$ (9,62 minutos), culminando em 23,08% de chance de não haver qualquer pessoa no sistema (P_0).

Tabela 3: Resultados da Modelagem sem Limite de Atendimentos. Testando Tempo Médio de Chegada (λ) > Tempo Médio de Atendimento (μ) para Situação Real (M/M/1) e Simulações (M/M/2 e M/M/3).

Características	M/M/1	M/M/2	M/M/3
m, n^o	1	2	3
λ , paciente/h	5	5	5
μ , paciente/h	4	4	4
L , pacientes	-5,0000	2,0513	1,3611
L_q , pacientes	-6,2500	0,8013	0,1111
W , h	-1,0000	0,4103	0,2722
W_q , h	-1,2500	0,1603	0,0222
ρ , %	1,2500	0,6250	0,4167
P_0 , %	-0,2500	0,2308	0,2786
P_n , %			
1	-0,3125	0,2885	0,3483
2	-0,3906	0,1803	0,2177
3	-0,4883	0,4883	0,0907
4	-0,6104	0,3052	0,1356
5	-0,7629	0,1907	0,0565
6	-0,9537	0,1192	0,0235
7	-1,1921	0,0745	0,0098
8	-1,4901	0,0466	0,0041
9	-1,8626	0,0291	0,0017
10	-2,3283	0,0182	0,0007

Fonte: Os Autores (2025).

A investigação sob a condição de limitação nas dependências da instituição, tomando $K = 5$ pacientes, revelou, tomando $\lambda < \mu$ (Tabela 4), que a existência de um e somente um profissional consumiria 75,00% da ocupação do sistema, com 30,41% de chance de não haver qualquer paciente nele. Isso acompanhando pela média de 1,70 pacientes no sistema (L), exigindo 0,8148h ($W = 48,89$ minutos) de tempo médio para vencer todo o sistema, e 1,00 paciente na fila (L_q), demandando em média 0,4814h de

espera ($Wq = 28,89$ minutos). Razoável seria supor, então que a configuração M/M/1/5 seria razoável, especialmente quando considerada $P_1 = 22,81\%$.

Tabela 4: Resultados da Modelagem com Limite de Pacientes nas Dependências ($K = 5$). Tempo Médio de Chegada (λ) < Tempo Médio de Atendimento (μ) para Simulações (M/M/1/K, M/M/2/K e M/M/3/K).

Características	M/M/1/K	M/M/2/K	M/M/3/K	M/M/3a/K
m, nº	1	2	3	3
K, pacientes	5	5	5	5
λ , paciente/h	3	3	3	3
μ , paciente/h	4	4	4	4
L, pacientes	1,7009	0,5122	-0,5831	0,0234
Lq, pacientes	1,0050	0,1045	0,0486	0,0234
W, h	0,8148	0,3141	0,2308	0,0078
Wq, h	0,4814	0,0641	-0,0192	0,0078
ρ , %	0,7500	0,3750	0,2500	0,2500
P_0 , %	0,3041	0,4564	1,8423	1,0000
P_n , %				
1	0,2281	0,3423	1,3817	0,7500
2	0,1711	0,1284	0,5181	0,2813
3	0,1283	0,0481	0,1295	0,0703
4	0,0962	0,0181	0,0324	0,0176
5	0,0722	0,0068	0,0081	0,0044
6	0,0541	0,0025	0,0020	0,0011
7	0,0406	0,0010	0,0005	0,0003
8	0,0304	0,0004	0,0001	0,0001
9	0,0228	0,0001	0,0000	0,0000
10	0,0171	0,0001	0,0000	0,0000

Fonte: Os Autores (2025).

Desenhos com $m \geq 2,00$ profissionais, tornariam os respectivos sistemas subutilizados, pois as taxas de ociosidade superariam aquelas de utilização, $\rho(M/M/2/5) = 37,50\%$ e $\rho(M/M/3/5) = 25,00\%$, e as relações custo – efetividade seriam demasiadamente desfavoráveis ao negócio, que pese os números de pessoas aguardando atendimento e os tempos em tais situações decrescessem rapidamente (Tabela 4). Isso seria tão abrupto em M/M/3/5 que $P_0 = 184,23\%$ e $P_1 = 138,17\%$. Essa possibilidade levou ao desenvolvimento da simulação M/M/3a/5, na qual $P_0 = 100,00\%$ foi imposta, apenas para exemplificação comparativa, culminando em $P_1 = 75,00\%$. Para além disso, respectivamente todas as estimativas foram menores, em termos absolutos, que aquelas conquistadas por M/M/3/5, exceto a taxa de ocupação, dado que considerou características constantes, número de estágios, tempo médio de chegada e tempo médio de atendimento. Pragmaticamente, as

constatações ratificaram que se a configuração do sistema conquistou $P_0 = 100,00\%$, então haveria superdimensionamento.

Simulando a possibilidade de $\lambda = \mu$ (Tabela 5), tal como ocorrido na Tabela 2, M/M/1/5 atingiria a ocupação máxima, porém possível seria estimar as demais características, portanto existiriam 2,50 pacientes no sistema e 1,67 pacientes na fila. Os tempos médios de espera seriam 45,00 minutos (0,7500h) e 30,00 minutos (0,5000h), respectivamente para sistema (W) e fila (Wq). Mereceu destaque imperativo, $P_n = 0,0000\% \forall n > K$, e $P_n = 16,6700\% \forall n \leq K$.

Tabela 5: Resultados da Modelagem com Limite de Pacientes nas Dependências ($K = 5$). Tempo Médio de Chegada (λ) = Tempo Médio de Atendimento (μ) para Simulações (M/M/1/K, M/M/2/K e M/M/3/K).

Característica	M/M/1/K	M/M/2/K	M/M/3/K
m, nº	1	2	3
K, pacientes	5	5	5
λ , paciente/h	4	4	4
μ , paciente/h	4	4	4
L, pacientes	2,5000	1,5000	0,8350
Lq, pacientes	1,6667	0,7500	0,1650
W, h	0,7500	0,5000	0,3116
Wq, h	0,5000	0,2500	0,0616
ρ , %	1,0000	0,5000	0,3333
P_0 , %	0,1667	0,2500	0,3300
P_n , %			
1	0,1667	0,2500	0,3300
2	0,1667	0,1250	0,1650
3	0,1667	0,0625	0,0550
4	0,1667	0,0313	0,0183
5	0,1667	0,0156	0,0061
6	0,0000	0,0078	0,0020
7	0,0000	0,0039	0,0007
8	0,0000	0,0020	0,0002
9	0,0000	0,0010	0,0001
10	0,0000	0,0005	0,0000

Fonte: Os Autores (2025).

A custo – efetividade seria comercialmente desfavorável com M/M/3/5, porque $\rho = 33,33\%$, novamente, mesmo tendo os demais indicativos favoráveis aos pacientes (Tabela 5). Em M/M/2/5, a utilização seria de 50,00%, com médias de 1,50 paciente no sistema (L) e menos de um na fila (Lq), aguardando 30 minutos (W) e 15 minutos (Wq), nessa ordem. As respectivas probabilidades de existência de indivíduos no sistema seriam inferiores aquelas de M/M/1/5, exceto para P_0 e P_1 . No contexto do negócio, a contratação de outro

profissional médico poderia ser adequada se convergisse ao planejamento estratégico e à visão da Organização. Portanto, a tomada de decisão quanto ao número de atendentes, um ou dois, seria circunstancial.

Quando $\lambda > \mu$ (Tabela 6), a circunstancialização desapareceria, pois M/M/1/5 colapsaria e M/M/3/5 apresentaria elevada probabilidade de inexistência de pacientes aguardando (P_0), assim como de haver somente um (P_1), diante do exposto o dimensionamento do sistema seria excessivo, representado pelo alteroso custo, o qual não teria correspondência na efetividade. Em razão disso, o modelo M/M/3a/5 foi implementado somente com fins ilustrativos.

Tabela 6: Resultados da Modelagem com Limite de Pacientes nas Dependências ($K = 5$). Tempo Médio de Chegada (λ) > Tempo Médio de Atendimento (μ) para Simulações (M/M/1/K, M/M/2/K e M/M/3/K).

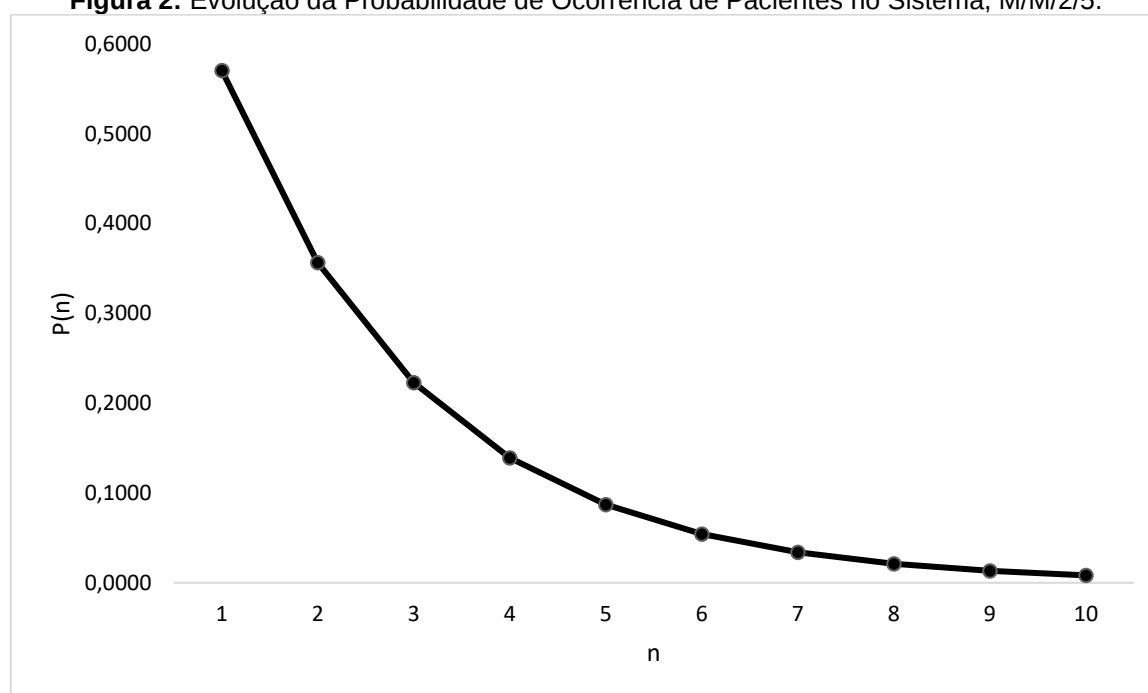
Característica	M/M/1/K	M/M/2/K	M/M/3/K	M/M/3a/K
m, nº	1	2	3	3
K, pacientes	5	5	5	5
λ , paciente/h	5	5	5	5
μ , paciente/h	4	4	4	4
L, pacientes	-	1,4421	-0,5948	0,2487
Lq, pacientes	-	0,7626	0,4581	0,2487
W, h	-	0,5306	0,1412	0,0497
Wq, h	-	0,2806	-0,1088	0,0497
ρ , %	-	0,6250	0,4167	0,4167
P_0 , %	-	0,4564	1,8423	1,0000
P_n , %				
1	-	0,5705	2,3029	1,2500
2	-	0,3566	1,4393	0,7813
3	-	0,2228	0,5997	0,3255
4	-	0,1393	0,2499	0,1356
5	-	0,0870	0,1041	0,0565
6	-	0,0544	0,0434	0,0235
7	-	0,0340	0,0181	0,0098
8	-	0,0213	0,0075	0,0041
9	-	0,0133	0,0031	0,0017
10	-	0,0083	0,0013	0,0007

Fonte: Os Autores (2025).

Em vista do exposto, M/M/2/5 seria a única decisão adequada, necessária e suficiente à demanda e ao atendimento (Tabela 6), mesmo tendo $\rho = 62,50\%$ (Ociosidade = 37,50%), e permitindo, em média, 31,83 minutos no sistema ($W = 0,5306h$) e 16,83 minutos na fila ($Wq = 0,2806h$) para 1,44 pacientes (L) aguardando, desses 0,76 pessoas na fila (Lq), médias aproximadas. Assim, a chance de não haver qualquer pessoa para os

dois especialistas seria, aproximadamente, 45,64% (P_0), mas a probabilidade superior seria $P_1 = 57,05\%$ com rápido decaimento (Figura 2), tornando a curva assintótica à abscissa a partir de oito pacientes no sistema ($n = 8$). Em última análise, quando o número de pacientes no sistema tenderia ao infinito ($n \rightarrow \infty$), a probabilidade de ocorrência desse quantitativo tenderia a zero ($P_n \rightarrow 0$).

Figura 2: Evolução da Probabilidade de Ocorrência de Pacientes no Sistema, M/M/2/5.



Fonte: Os Autores (2025).

Considerações Finais

Objetivando caracterizar serviços de saúde de uma especialidade ofertada em determinada clínica, estimados foram os modelos M/M/1 e M/M/1/K. Os resultados indicaram que o modelo com um profissional seria adequado na existência ou não de restrição desde a média de chegada dos pacientes fosse inferior àquela do atendimento. Então, possível foi concluir que os modelos atenderam adequadamente ao proposto.

Aos estudos futuros recomenda-se a consideração dos custos de manutenção e implementação dos modelos como tática à adequação dos modelos. A consideração da relação custo – efetividade poderia pormenorizar os resultados. A análise inferencial das curvas de probabilidade poderia indicar o melhor ajuste em razão da probabilidade de existência de pacientes no sistema. Refazer as estimativas considerando a chegada ou

atendimento como obedientes à distribuição de Poisson poderia fornecer resultados distintos.

Referências

- ABENSUR, EO *et al.* Tendências para o auto-atendimento bancário brasileiro: um enfoque estratégico baseado na teoria das filas. **Revista de Administração Mackenzie**, ano 4, n. 2, p. 39-59, 2003.
- ABENSUR, EO. **Pesquisa operacional para cursos de engenharia de produção**. São Paulo: Blucher, 2018.
- AKKARI, ACS. **Processos estocásticos**. Londrina (PR): Editora e Distribuidora Educacional, 2018.
- ALMEIDA, FP. *et al.* Análise do processo de picking em uma loja de rede varejista. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 07, e296, 2024.
- ANDRADE, KC *et al.* Simulando o fluxo de bagagens: método para reduzir extravios nos aeroportos com o uso do sistema Arena. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 09, e441, 2024.
- ASSIS, AR; QUEDAS, AC; SILVA, VD. Análise do acesso de estudantes ao campus da Fatec zona leste: um estudo de caso logístico. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 06, p. 90–100, 2024.
- CLARKE, AB; DISNEY, RL. **Probabilidade e processos estocásticos**. Rio de Janeiro: LTC, 1979.
- ERLANG, AK. Calcul des probabilités et conversations téléphoniques. **Revue Générale de l'Electricité**, v. 18, n. 8, p. 305–309, 1925.
- ERLANG, AK. Løsning af nogle problemer fra sandsynlighedsregningen af betydning for de automatiske telefoncentraler. **Elektroteknikeren**, v. 13, p. 5–13, 1917.
- ERLANG, AK. Sandsynlighedsregning og telefonsamtaler. **Nyt Tidsskrift for Matematik**, v. 20, n. B, p. 33–39, 1909.
- FREITAS, ADS *et al.* Aplicação da simulação computacional no fluxo de atendimento em uma unidade hospitalar no município de Natal/RN. **Revista Foco**, v. 17, n. 1, e4198, 2024.
- GIRÃO, GS *et al.* Análise da fila de atendimento na bilheteria do Zoo Safári - SP: um estudo sobre implementação de automação na redução das filas. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 08, e348, 2024.
- HARADA, DYB. **Entendendo as filas de espera**: uma abordagem para o ensino médio. Dissertação (Mestrado Profissional em matemática em Rede Nacional) – Faculdade de Ciências e Tecnologia. Universidade Estadual Paulista Júlio Mesquita Filho. Presidente Prudente (SP), 2017.
- JANDRE, GRS *et al.* Modelagem estocástica da variabilidade da frequência cardíaca. **Revista Presença**, v. 10, n. 22, p. 237-252, 2024.
- JONES, PW; SMITH, P. **Stochastic processes an introduction**. Boca Raton (USA): CRC Press, 2009.
- KOVÁCS, ZL. **Teoria da probabilidade e processos estocásticos**. São Paulo: Edição Acadêmica, 1996.
- LEITE, IP. **Melhoria do nível de serviço no atendimento de um supermercado usando teoria das filas e simulação**. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Produção) - Universidade Federal do Amazonas. Itacoatiara (AM), 2024.

LIMA, MAN; VILLELA, DAM. Demanda e ocupação de leitos do Sistema Único de Saúde para os principais tipos de câncer no Brasil, de 2018 a 2021: estudo transversal. **Epidemiologia e Serviços de Saúde**, v. 33, e20231172, 2024.

MATHIAS, AS *et al.* Aplicação da Teoria das Filas em uma Cantina. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 09, e409, 2024.

MELO, LM. **Análise, modelagem e simulação de um processo produtivo de massa de linguiça frescal**. Dissertação (Mestrado em Engenharia de Alimentos) - Universidade Federal de Uberlândia. Patos de Minas (MG), 2024.

NESSY, VM; SIMÕES, WL. Otimização do fluxo de atendimento multidisciplinar a crianças com dificuldades de aprendizagem. **Brazilian Journal of Production Engineering**, v. 10, n. 5, p. 184–195, 2024.

PEREIRA, LMS. **Teoria das filas aplicada ao sistema de carregamento de etanol de uma empresa do setor sucroenergético distribuidora de etanol no estado de São Paulo**. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Produção) – Centro de Tecnologia. Universidade Federal da Paraíba. João Pessoa (PB), 2023.

PICANÇO, ARS *et al.* Modelagem e otimização do sistema de atendimento de um restaurante baseado em Teoria de Filas. **Revista de Gestão e Secretariado**, p. 15, n. 5, e3290, 2024.

SILVA, GC; COSTA, JS; JESUS, RG. Simulação aplicada no transporte público urbano: análise da linha 407L/10 Barro Branco – Guilhermina Esperança em horário de fluxo pela manhã. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 09, e413, 2024.

SIMOYAMA, F; DINIZ, DM; BATTISTI, MCG. Aplicação de teoria das filas no atendimento telefônico: configuração de turnos em uma autarquia federal. **@REGIT**, v. 18, n. 2, p. 09-24, 2022.

TAYLOR, HM; KARLIN, S. **An introduction to stochastic modeling**. New York (USA): Academic Press, 1998.

TORRES, OF. Elementos da teoria das filas. **Revista de Administração de Empresas**, v. 6, n. 20, p. 111-127, 1966.