

ANÁLISE DE TEORIA DAS FILAS: SISTEMA DE FILAS DE UM SERVIÇO DE PRONTO ATENDIMENTO

Francieli de Fátima de Oliveira¹
Mara Lucia Grando; Marcel Belusso; Stefan Antônio Bueno²

RESUMO

A teoria das filas apresenta um problema comum nos dias atuais o congestionamento de sistemas, onde a principal característica é a presença de “clientes” solicitando “serviços” de alguma maneira, seja pessoa física ou matéria-prima em processo. Em sua expressão mais simples, um sistema de filas é composto de elementos que querem ser atendidos em um posto de serviço e que, eventualmente, devem esperar até que o posto esteja disponível. Quando se trata do setor da saúde, devem-se considerar ainda as consequências que o atraso gerado por espera em uma fila pode ocasionar no aumento da gravidade ou sequelas do problema de saúde, ou diminuição da chance de cura do indivíduo. A importância de gerar indicadores de desempenho, ainda que analíticos, é auxiliar a tomada de decisão de maneira que o tamanho das filas e a permanência de indivíduos nelas sejam minimizados. Neste trabalho, a metodologia da Teoria de Filas foi aplicada em um Pronto Atendimento Emergencial do (SUS) a fim de analisar o comportamento da fila existente. Algumas das medidas de desempenho analisadas fazem referência à quantidade média de indivíduos na fila e no sistema, o tempo médio de espera na fila e no sistema, assim como a sua taxa de ocupação, através da metodologia aplicada e cálculos quantitativos utilizados, observou-se que a quantidade de atendentes disponíveis supre a demanda de clientes que chegam a fila do Pronto Atendimento Emergencial analisado.

Palavras-Chaves: Teoria das Filas. Pronto Atendimento. Filas.

1 INTRODUÇÃO

A teoria das filas é uma probabilidade que estuda a formação de filas onde prove modelos para demonstrar o número de chegadas e de atendimento de clientes, através de cálculos matemáticas ela tenta encontrar um ponto de equilíbrio que satisfaça o cliente e seja viável ao servidor. As formações de filas ocorrem porque a procura pelo serviço é maior que a capacidade de atendimento.

No cenário atual, esta cada vez mais comum as empresas buscarem por soluções inovadoras para maximizar seu lucro e otimizar o seus processos, nesse caso a teoria das filas se tornou um método de avaliar a capacidade de atendimento levando em consideração o número de chegadas, tempo de espera e o tempo que leva cada atendimento. Através desses cálculos é possível identificar as necessidades do processo e a viabilidade de ter mais pessoas

¹ Acadêmico de Engenharia de Produção da UCEFF. E-mail: francieli.fa.oliveira@gmail.com

² Docentes de Engenharia de Produção da UCEFF. E-mail: maralucia35@gmail.com; marcel.belusso@uceff.edu.br; buenostefan@gmail.com.

para suprir a essa demanda. Segundo Dávalos (2012) a teoria parte do seguinte pressuposto “a formação de filas excede a capacidade de fornecer determinado serviço” e envolve o estudo matemático utilizando ferramentas de tratamento estatístico ou estocástico.

Neste contexto, torna-se relevante entender os diversos elementos que afetam a lista de espera por atendimento nas empresas, buscando ações que minimizem esse problema e, conseqüentemente, proporcione uma otimização dos serviços realizados. O estudo da teoria das filas contribui para o gerenciamento e melhorias no processo de atendimento, procurando evitar desperdícios e gargalos e satisfazendo ambas as partes.

O objetivo deste trabalho é analisar o sistema de atendimento no setor de Pronto Atendimento (SUS), ocupação do sistema e cenário de melhoria com disponibilidade de adição de novos atendentes.

2 REFERENCIAL TEÓRICO

2.1 FILAS

Para Lovelock e Wright (2002), uma fila é a representação de uma linha de pessoas, veículos, objetos físicos ou intangíveis que aguardam sua vez de serem atendidos. O processo de formação de filas ocorre quando o número de chegadas excede à capacidade do sistema de atendimento e geralmente está relacionado a problemas relativos à administração da capacidade. Portanto fila é uma sequência de pessoas, produtos ou serviços aguardando por atendimento, ou sequencia do processo, consiste do processo de chegada e distribuição do tempo de serviço, do número de servidores e da capacidade do sistema.

Gianesi e Corrêa (1994) citam que em estudos realizados observou-se que para os clientes, a sensação de espera é mais importante na formulação de sua percepção do que o tempo real gasto para esperar o serviço.

2.2 TEORIA DAS FILAS

Todas as pessoas já passaram pelo aborrecimento de ter que esperar em uma fila, em supermercados aguardando para passar as compras, em hospitais esperando por atendimento médico, nos bancos e em muitas outras situações, mas não apenas pessoas esperam em filas:

nas indústrias produtos podem estar esperando para serem processados, aviões podem estar esperando pra aterrissar, navios podem estar aguardado para entrar no porto.

Fogliatti e Mattos (2007) definem que um sistema com fila consiste na chegada de usuários que necessitam de um determinado serviço, esperam numa fila – que se forma quando a taxa de atendimento é menor que a taxa de chegada de usuários –, são atendidos e saem do sistema após o serviço ser prestado.

A abordagem sobre a teoria das filas teve início em 1908, Copenhague, Dinamarca, com A. K. Erlang, considerado o pai da Teoria das Filas, quando trabalhava em uma companhia telefônica onde realizou um estudo sobre o congestionamento das linhas telefônicas da companhia, porém, somente a partir da Segunda Guerra Mundial ela foi aplicada a outros problemas de filas (PRADO, 2006).

Fogliatti e Mattos (2007) citam como objetivo da Teoria de Filas a determinação de medidas de desempenho que podem considerar a ótica do usuário – o tamanho médio da fila e os tempos médios na fila e de permanência no sistema – e a ótica da gerência – o tempo médio de serviço e a taxa de ociosidade dos atendentes.

2.2.1 Classificação de um sistema de Filas

Para Mendonça (2014) um sistema com fila é composto por usuários, por canais ou postos de serviço/atendimento e por um espaço designado para a espera. Os usuários chegam em um determinado tempo, caracterizando o processo de chegada, para serem atendidos em determinado canal, posto de atendimento ou sequência de processo de manufatura, enquanto os postos estão ocupados os usuários aguardam em uma fila com espaço determinado para mesma. Assim que um canal de serviço fica livre, um dos usuários da fila é chamado para atendimento, uma vez completado o serviço, o usuário é liberado do sistema.

Além das características gerais de um sistema de filas, é importante determinar a estrutura do sistema, que também é um elemento fundamental de estudo das filas as quais se classificam em: fila única e um servidor, fila única e múltiplos servidores em paralelo, múltiplas filas e múltiplos servidores em paralelo, fila única e múltiplos servidores em série. A disciplina da fila define qual a ordem em que os usuários serão atendidos (ARENALES, 2007).

Na teoria das filas existem indicadores que podem ser usados para mensurar a eficiência da prestação do serviço, entre os quais se destacam: tempo médio de um usuário na fila, tamanho médio da fila, tempo gasto pelo usuário na fila desde a chegada até o momento da

saída, tempo em porcentagem da ociosidade ou congestionamento no ponto de atendimento (ANDRADE, 2011).

Segundo Prado (1999) o estudo do comportamento de filas é uma forma de alterar os sistemas nos quais existem gargalos. O objetivo de estudar o comportamento das filas é prestar um melhor atendimento ou proporcionar uma redução de custos, já que filas são dispendiosas. Existem fatores que influenciam a operação de um sistema são eles: forma de atendimento, disciplina da fila, forma de chegada e estrutura do sistema de atendimento, um fator importante é determinar a estrutura da fila, pois cada um exige um sistema de estudo analítico diferente.

As equações são classificadas baseadas nas seguintes características dos processos de chegada e de atendimentos aos clientes: As chegadas se processam segundo uma distribuição de Poisson com média λ chegadas/tempo. Os tempos de atendimento seguem a distribuição exponencial de Poisson com média μ . O atendimento à fila é feito por ordem de chegada. O número de clientes potenciais é suficientemente grande para que a população possa ser considerada infinita, (ARENALES, 2007).

O levantamento de dados estatísticos tem como finalidade determinar a distribuição de probabilidades do número de atendimentos ou da duração de cada atendimento.

2.2.2 Processo de chegada e tempo de atendimento

O processo de chegada indica qual o padrão de chegada dos clientes no sistema. Apresenta comportamento estocástico, ou seja, as chegadas ocorrem no tempo e no espaço de acordo com as leis da probabilidade; assim, é preciso conhecer qual a distribuição de probabilidade que descreve os tempos entre as chegadas dos clientes. (AQUILANO, CHASE, JACOBS, 2006).

As equações são classificadas baseadas nas seguintes características dos processos de chegada e de atendimentos aos clientes: As chegadas se processam segundo uma distribuição de Poisson com média λ chegadas/tempo. Os tempos de atendimento seguem a distribuição exponencial de Poisson com média μ . O atendimento à fila é feito por ordem de chegada. O número de clientes potenciais é suficientemente grande para que a população possa ser considerada infinita (ARENALES, 2007).

O levantamento de dados estatísticos tem como finalidade determinar a distribuição de probabilidades do número de atendimentos ou da duração de cada atendimento. O processo de chegadas de um usuário no sistema, segundo Fogliatti e Mattos (2007), pode ser

determinístico, quando a quantidade de chegadas e os momentos em que elas ocorrem são conhecidos, ou estocástico, quando se tem um comportamento aleatório.

2.2.3 Números de servidores

O número de servidores também conhecido como número de canais de serviço indica a quantidade de "pontos de atendimento" do sistema, de forma a servir aos clientes paralelamente sendo possível mais de um servidor (atendente). Em um sistema de fila única, existe uma única fila para todos os servidores, como em um caixa de banco. Em um sistema de múltiplas filas, existe uma fila para cada servidor, como em um caixa de supermercado (ARENALES, 2007).

2.2.4 Capacidade do Sistema

Representa o número máximo de clientes que o sistema suporta, incluindo os que estão em espera e os que estão sendo atendidos. A capacidade pode ser infinita ou finita. Se a capacidade for finita, quando o sistema estiver lotado nenhum cliente pode entrar até que um cliente saia do sistema, liberando espaço (AQUILANO, CHASE, JACOBS, 2006).

2.2.5 Disciplina da Fila

A disciplina de filas refere-se à maneira como os clientes são escolhidos para entrar em serviço após uma fila ser formada. A maioria das disciplinas comuns que podem ser observadas na vida diária é FIFO (First In First Out), ou seja, o primeiro a chegar é o primeiro a ser atendido. Entretanto, existem outras disciplinas, tais como, LIFO (Last In First Out), ou seja, último a chegar primeiro a ser atendido (ANDRADE, 2011).

2.2.6 Medidas de Desempenho

Segundo Pinto (2011) tendo em conta que o tempo perdido em filas de espera pode constituir um custo para as organizações, entende-se a preocupação dos gestores em melhorar as características dos sistemas de fila de espera considerando diversas alternativas. Para isso é útil quantificar o desempenho de cada sistema através de medidas de desempenho que constituirão uma valiosa informação para quem decide.

Segundo Pinto (2011), estas medidas são as seguintes:

- Número Médio de Clientes na Fila de Espera (L_q)
- Número Médio de Clientes no Sistema (L)
- Tempo Médio que um cliente Espera na Fila (W_q)
- Tempo Médio que um cliente Espera no sistema (W)
- Probabilidade de que o sistema esteja Ocupada (ρ)
- Probabilidade de que o sistema esteja Desocupada (P_0).

2.3 M/M/S

É um modelo baseado num processo com número de servidores disponíveis: S , o sistema tem uma distribuição das chegadas de Poisson e dos tempos de atendimento exponencial, a capacidade do sistema e da população são infinitas e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair (PINTO,2011). O Quadro 1 mostra as formulas do sistema.

Quadro 1 - Fórmulas do Sistema M/M/S

Descrição	Fórmulas
Probabilidade de que o Sistema Esteja Ocupada	$\rho = \frac{\lambda}{\mu}$
Probabilidade de que n Clientes Encontrem-se no Sistema	$P_n = (1 - \rho)\rho^n$
Probabilidade de que o sistema esteja Desocupada	$P_0 = (1 - \rho)$
Numero Médio de Clientes no Sistema de Atendimento	$L = \frac{\lambda}{\lambda - \mu}$
Numero Médio de Clientes na Fila de Espera	$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Tempo Médio Gasto no Sistema pelo Cliente	$W = \frac{1}{\mu - \lambda}$
Tempo Médio de Espera na Fila por Cliente	$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$

Fonte: Adaptado de Pinto (2011).

3 MÉTODO DE PESQUISA

Para Marconi e Lakatos (2006), pesquisa é uma atividade voltada à busca de respostas e à solução de problemas para questões propostas, para isso utilizam-se métodos científicos. Este artigo resulta dos dados recolhidos e analisados em campo associados aos conhecimentos técnicos acadêmicos com a aplicação de técnicas da Teoria das Filas. Empregou-se a pesquisa

de campo através de observação e pesquisa bibliográfica, pois a obtenção de dados foi na área operacional, sendo todas as observações e coletas de dados realizadas sem interferência no processo natural de atendimento. A técnica utilizada na pesquisa em questão foi à observação, que “[...] é uma técnica de coleta de dados para conseguir informações e utiliza os sentidos na obtenção de determinados aspectos da realidade” (MARCONI; LAKATOS, 2006, p. 192).

A população deste estudo foi pesquisada por meio da técnica de amostragem que, segundo Marconi e Lakatos (2006), consiste em escolher uma parte representativa do todo para o estudo, de maneira que os resultados referentes à amostra pesquisada sejam suficientes para gerar conclusões a respeito da população total. A amostra pesquisada consiste nas pessoas que chegam ao pronto atendimento em determinado dia e horário, mas apenas as pessoas usuárias do sistema, ou seja, aquelas que chegam e recebem atendimento na recepção. Os dados foram coletados em dias da semana diferentes, bem como em horários diferentes visando ter uma diversidade e maior confiabilidade dos dados obtidos tendo em vista o resultado final do artigo que visa entender a realidade dos serviços de Pronto Atendimento (SUS) nos períodos diurno e noturno.

4 RESULTADOS

O sistema de filas analisado no presente trabalho é baseado em um serviço de pronto atendimento localizado na Grande Efapi. Determinado através do método de observação pelo número de chegadas à recepção do mesmo, e pela fila de atendimento da sala de recepção dos usuários, na recepção encontram-se 2 guichês de atendimento permanentes e um guichê para auxílio em eventuais situações de maior demanda.

Os usuários chegam ao sistema pela porta de entrada do PA, e seguem para a frente do guichê, onde se inicia a formação de fila única, conseqüentemente pela chegada de mais usuários. No guichê, o atendimento consiste no preenchimento de uma ficha com os dados do paciente dentre outras burocracias de atendimento. A disciplina da fila segue o padrão FIFO, de maneira que o atendimento ocorre de acordo com a ordem de chegada à fila, não tendo preferência por grau de idade, deficiência ou gestante.

O período analisado foi de 10 dias, sendo 1 hora por dia, variando nos dois turnos de atendimento. O pesquisador foi à unidade de Pronto Atendimento a fim de observar características do sistema que tem importância para o estudo, as quais estão descritas a seguir:

- Número de usuários na fila no período de 1 hora;

- Número de atendimentos realizados nos guichês no período de 1 hora.

Em seguida, os dados coletados foram tabelados para serem utilizados posteriormente. Foram coletados e analisados 311 amostras de chegadas de clientes, e 302 amostras de atendimentos. Permanentemente o sistema conta com dois guichês em período integral em atendimento sendo o terceiro guichê somente acionado caso ocorra uma demanda maior do serviço, o que não ocorreu nos dias das respectivas observações. No primeiro dia o guichê 1 houve a ausência da atendente por 30 minutos.

A Tabela 1, mostra um comparativo e uma análise dados, a fim de identificar a chegada de clientes de taxa de atendimento durante o período diurno e no período noturno, assim como os dias das amostras.

Tabela 1: Dados coletados

DIAS DE AMOSTRAS	HORÁRIO DE AMOSTRA	PERÍODO DE OBSERVAÇÃO	CHEGADAS (PESSOAS)
QUI - 11/05/2017	07:00/08:00	DIURNO	60
SEX - 12/05/2017	19:00/20:00	NOTURNO	23
SAB - 13/05/2017	23:00/00:00	NOTURNO	29
SEG - 15/05/2017	11:00/12:00	DIURNO	8
TER - 16/05/2017	17:00/18:00	DIURNO	16
QUA - 17/05/2017	07:00/08:00	DIURNO	55
QUI - 18/05/2017	07:00/08:00	DIURNO	48
SEX - 19/05/2017	21:00/22:00	NOTURNO	36
SAB - 20/05/2017	20:00/21:00	NOTURNO	20
SEG - 22/05/2017	22:30/23:30	NOTURNO	16

Fonte: Dados da pesquisa (2017).

Tabela 2: Resultados Parciais das amostras (chegadas)

RESULTADOS PARCIAIS CHEGADAS	
TOTAL CHEGADAS	311
MÉDIA CHEGADAS	31,1

Fonte: Dados da pesquisa (2017).

Tabela 3: Cálculos das amostras (atendimentos)

DIAS DE AMOSTRA	NUMERO DE ATENDIMENTOS POR ATENDENTE		TOTAL HORA
	ATENDENTE 1	ATENDENTE 2	
QUI - 11/05/2017	16	35	51
SEX - 12/05/2017	12	11	23
SAB - 13/05/2017	13	16	29
SEG - 15/05/2017	5	3	8
TER - 16/05/2017	6	10	16
QUA - 17/05/2017	24	31	55
QUI - 18/05/2017	21	27	48
SEX - 19/05/2017	13	23	36
SAB - 20/05/2017	12	8	20
SEG - 22/05/2017	9	7	16

Fonte: Dados da pesquisa (2017).

Tabela 4: Resultados parciais das amostras (atendimentos)

RESULTADOS PARCIAIS ATENDIMENTOS	
TOTAL ATENDIMENTOS	302
MÉDIA ATENDIMENTOS	30,2

Fonte: Dados da pesquisa (2017).

Utilizando as informações obtidas através do método de observação, e aplicando o modelo M/M/S de teoria das filas, se têm os seguintes resultados:

Tabela 5: Resultados através da aplicação das amostras

RESULTADOS OBTIDOS	
S (Atendentes)=	2
λ (Chegadas)=	31,10 / hr
μ (Atendimentos)=	30,2 / hr
P (Taxa Utilização)=	0,5149 (51,49%)
P0 (Taxa de Desocupação)=	0,1894 (18,94%)
Lq (Nº Clientes na Fila)=	0,2072
Wq (Tempo na Fila)=	0,0066 horas ou 23,76 segundos
L (Nº Clientes no Sistema)=	1,237
W (Tempo no Sistema)=	0,0331 horas ou 142,92 segundos

Fonte: Dados da pesquisa (2017).

5 CONSIDERAÇÕES FINAIS

Com a realização deste trabalho foi possível medir a capacidade operacional do sistema de filas que caracteriza a unidade objeto de estudo. Os resultados apontam para o cumprimento dos objetivos estipulados. Não se tem congestionamentos no sistema de atendimento da recepção. Através da técnica de múltiplos servidores observamos que o acionamento do terceiro guichê para atendimento faz-se desnecessária haja visto termos um tempo de espera considerado relativamente pequeno.

REFERÊNCIAS

- ANDRADE, Eduardo Leopoldino. **Introdução a Pesquisa Operacional**. 4ª Ed, LTC – Livros Técnicos e Científicos Editora Ltda, Rio de Janeiro, RJ, 2011.
- AQUILANO, N.J.; CHASE, R.B.; JACOBS, F.R. **Administração da Produção e Operações para Vantagens Competitivas**. 11ª.ed. São Paulo, 2006.
- ARENALES, Et-al, Armentano, Morabito, Yanasse. **Pesquisa Operacional**. 6ª Tiragem, ED; Campus- Abrepro, Rio de Janeiro, RJ. 2007.
- DÁVALOS, Pablo Bezzera. **Hidroquímica do Estuário do Rio Caravelas**, Caravelas – BA / Pablo Bezzera Dávalos – Natal, RN, 2012
- FOGLIATTI, M. C.; MATTOS, N. M. C. **Teoria de Filas**. Editora Interciência. Rio de Janeiro, 2007.
- GIANESI, I. G. N; CORRÊA, H. L. **Administração estratégica de serviços: operações para satisfação do cliente**. São Paulo: Atlas, 1994.
- LOVELOCK, C.; WRIGHT, L. **Serviços: marketing e gestão**. São Paulo: Saraiva, 2002.
- MARCONI, M. de A; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2006.
- MENDONÇA, E. **Teoria de filas markovianas e aplicações**, 2014. Disponível em: <http://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/5265/1/PDF%20-%20Edn%C3%A1rio%20Barbosa%20de%20Mendon%C3%A7a.pdf>. Acesso em maio 2017.
- PRADO, Darci. **Teoria das Filas e da Simulação**. v.2. 4. Ed. Belo Horizonte: editora de Desenvolvimento Gerencial. Serie Pesquisa Operacional, 2009.
- PINTO, Ângelo Santos. **Aplicação da Teoria de Filas na Análise da Capacidade Operacional de um Sistema** - Estudo Caso BCA Porto Novo. Disponível em: <http://www.portaldoconhecimento.gov.cv/bitstream/10961/497/2/Trabalho%20Fim%20de%20Curso.pdf>. Acesso em maio 2017.