Universidade de Brasília
Departamento de Estatística

Modelagem da fila de espera para leitos de UTI no DF

Yasmin Lírio Souza de Oliveira

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2022

Yasmin Lírio Souza de Oliveira

Modelagem da fila de espera para leitos de UTI no DF

Orientador: Prof. Dr. Lucas Moreira

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2022

Resumo

Os sistemas de filas são regidos por leis probabilísticas que englobam a Teoria de Processos Estocásticos. As linhas em que a taxa de chegadas de clientes e o número de clientes atendidos por unidade de tempo obedecem um processo de Poisson são exemplos especiais de Cadeias de Markov, onde as transições de estado são processos contínuos em relação ao tempo de permanência. A cadeia de Markov a tempo contínuo aplicada ao modelo de fila único é chamada Processo de Nascimento e Morte, onde uma chegada representada um nascimento e uma saída do sistema representa uma morte.

Os tempos de internação e espera por leitos em UTI prolongadados podem ser sinônimo de malefícios para o enfermo e para o hospital devido à geração de despesas excessivas e possíveis complicações por falta de atendimento imediato. Após observação das estimativas do tempo médio e do número médio de pessoas em espera por vagas em Unidades de Tratamento Intensivo no DF, foi proposto um modelo em que a adição de um segundo servidor gerou melhorias tanto em relação ao tamanho e tempo na fila quanto em relação ao atendimento dos pacientes.

Palavras-chaves: Teoria de Filas; Processos Estocásticos; Cadeia de Markov; Processo de Poisson;Processos de Nascimento e Morte; Modelos de Filas; Internação de pacientes; Unidade de Tratamento Intensivo; Tempo de espera; Tamanho da Fila;

Lista de Tabelas

1 Dedução de P_n para o modelo M/M/1. 20

2 Fragmento da lista de pacientes aguardando leitos de UTI. 28

3 Estimativas da simulação e valor teórico obtido através da fórmula. 31

4 Intervalos de Confiança para as estimativas médias. 32

5 Pacientes,horário de ingresso na Fila e Intervalo entre chegadas (em minutos). 33

6 Intervalos de Confiança para as estimativas médias do modelo M/M/1. . . 36

7 Intervalos de Confiança para as estimativas médias do modelo M/M/2. . . 38

8 Intervalos de Confiança para as estimativas médias do modelo M/M/3. . . 40

9 Estimativas das medidas de desempenho por modelo aplicado. 43

10 Intervalos de Confiança para as estimativas médias dos modelos. 44

Lista de Figuras

1 Grafo do Processo de Nascimento e Morte. 12
2 Diagrama de transição de estados do Sistema M/M/1. 20
3 Novos casos de COVID-19 no DF por dia. 25
4 Numéro de pacientes na fila de espera. 28
5 Taxa de ocupação total dos leitos COVID. 29
6 Número de leitos COVID por dia. 29
7 Histograma dos intervalos entre chegadas. 34
8 Gráfico Q-Q da distribuição dos intervalos entre chegadas. 34
9 Numéro de pacientes na fila total e com suspeita ou confirmação de covid-19. 41
10 Histograma dos intervalos entre chegadas. 42
11 Gráfico Q-Q da distribuição dos intervalos entre chegadas. 42

Sumário

1 Introdução 8

2 Fundamentação teórica 10

2.1 Processos estocásticos 10

2.2 Processo de Markov 10

2.3 Processos de Nascimento e Morte. 11

2.4 Processos de Poisson 15

2.4.1 Processos de Poisson Homogêneos 15

2.5 Distribuição Exponencial 16

2.6 Teste Kolmogorov-Smirnov de Aderência 16

2.7 Gráficos Q-Q 17

2.8 Sistema de Filas 17

2.9 Fila Única M/M/1 19

2.10 Modelo M/M/c 22

3 Leitos de UTI na rede pública de saúde do Distrito Federal 24

4 Metodologia 27

4.1 Conjunto de dados 27

5 Simulação para o Modelo M/M/1 31

5.1 Estimação 31

5.2 Intervalos de Confiança 32

6 Resultados 33

6.1 Seleção do modelo 33

6.2 Modelo M/M/1 35

6.3 Modelo M/M/2 37

6.4 Modelo M/M/3 39

6.5 Pacientes com suspeita/confirmação de covid-19 40

7 Conclusão 45

Referências 48

Anexo 50

8 Código da simulação 50

1 Introdução

O intuito da teoria de Filas Markovianas é prever a chegada de clientes e a geração de filas para adequar a infraestrutura do atendimento e amenizar o estresse e exaustão que uma espera prolongada possa causar através de uma melhoria na utilização dos serviços disponíveis.

Erlang (1909) realizou um dos primeiros estudos acerca da teoria das Filas, cuja aplicação foi em relação a circuitos telefônicos e ao problema de tráfegos nas linhas, onde provou que ligações aleatoriamente distribuídas seguiam uma distribuição de Poisson e definiu a expressão do tempo médio da demora no atendimento de ligações.

A Teoria das Filas é um ramo da probabilidade, mas também é documentada na literatura de pesquisa operacional (PERDONÁ et al., 2017) e engenharia industrial (CAMELO et al., 2010), utiliza conceitos de processos estocásticos e da matemática aplicada na análise da formação e comportamento das filas. Controle de tráfego (aéreo, veículos, pessoas), sistema de comunicação e clientes que esperam atendimento em uma lotérica são alguns exemplos que adotam essa teoria.

A partir do padrão probabilístico das chegadas dos clientes à fila, dos atendimentos fornecidos pelo servidor e a partir do número de canais de atendimento disponíveis podemos indicar um modelo quantitativo de fila para avaliar esse sistema ou situação em particular. Uma vez que o tempo de espera para o atendimento reflete a qualidade dos serviços prestados por uma empresa, a Teoria das Filas busca encontrar um ponto de equilibrio que satisfaça o cliente e o orçamento do prestador de serviço.

Um sistema de filas consiste no processo de chegada de clientes, da distribuição do tempo do serviço, do número de servidores, da capacidade de atendimento do complexo, da população de usuários e da disciplina de atendimento. Filas são formadas quando há um excesso de procura em relação à capacidade da organização de responder a essa demanda. Esse problema muitas vezes não pode ser resolvido aumentando a capacidade do sistema de atendimento por questões de inviabilidade financeira e limitações de espaço.

O presente trabalho aplica modelos da teoria de Filas Markovianas para analisar a lista de espera por leitos gerais e leitos COVID da Unidade de Terapia Intensiva na rede pública de saúde do Distrito Federal. Uma Unidade de Terapia Intensiva (UTI) é uma área altamente especializada do hospital, contendo equipe técnica qualificada e multidisciplinar destinada ao tratamento e monitoramento de pacientes oferecendo cuidados personalizados e instrumentos tecnológicos primorosos a fim de reestabelecer a saúde do

paciente (MORAES et al., 2018).

Apesar de uma UTI ser uma unidade de cuidado intensivo, a internação de pacientes em seus leitos por muito tempo pode ser sinônimo de infortúnios para o paciente e para o hospital. A geração de custos exorbitantes, a possibilidade de enfermos em estado crítico não serem admitidos devido a carência de leitos (ocasionando em aumento da morbidade intra-hospitalar), o maior risco de infecções e maior chance de desenvolverem problemas psicológicos, como ansiedade e depressão, são alguns dos empecilhos que uma internação por tempo prolongado na UTI pode provocar. 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 alguns modelos de filas, de que forma a classificação desse modelo é feita e uma simulação utilizando o software R para a estimação dos parâmetros tempo de espera médio na fila, taxa de ocupação média e tamanho médio da fila. Posteriormente a Teoria das Filas é aplicada ao alinhamento de espera por leitos de UTI do Sistema Único de Saúde (SUS) no Distrito Federal.

2 Fundamentação teórica

2.1 Processos estocásticos

Um processo estocástico $\{X_t, t \in T\}$ é uma coleção de variáveis aleatórias, ou seja, para cada t , X_t é uma variável aleatória correspondente, onde T é dito como o conjunto de índices. O índice t é usualmente lido como um instante no tempo e X_t por consequência é interpretado como o estado do processo nesse instante t. Por exemplo, X_t pode representar o número pessoas que entraram em um banco no instante t. Caso o conjunto T seja contável $\{X_t, t = 0, 1, ...\}$ o processo estocástico será dito como um processo de tempo discreto, caso T seja um intervalo dos números reais $\{X_t, t \in \mathbb{R}\}$ ou uma combinação de intervalos reais, o processo será dito como um processo de tempo contínuo (ROSS, 2014).

Um processo estocástico descreve o comportamento de um sistema durante um período de tempo através de uma família de variáveis aleatórias. O conjunto de todos os possíveis valores que as variáveis X_t podem assumir é denominado como espaço de estados S do processo estocástico (JÚNIOR, 2017).

A probabilidade de um sistema ir do estado i para o estado j depende dos estados passados do sistema

$$P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0), \tag{2.1.1}$$

onde $n \in T$, $i, j \in S$ e i_0 o estado inicial do processo.

Neste trabalho, t representará o instante no qual o paciente ingressou no sistema de filas por leitos de UTI, portanto abordaremos processos de tempos contínuos. As análises foram realizadas adotando T em minutos.

2.2 Processo de Markov

Um processo de Markov $\{X_t, t \in \mathbb{R}\}$ é um processo estocástico onde a probabilidade de qualquer comportamento futuro do processo, quando o estado atual é conhecido, não é alterado pelo conhecimento adicional do comportamento passado. Pode-se dizer que, dado a condição "presente"do processo, o "futuro"é independente do "passado", ou seja, o processo está sem memória (ARAÚJO, 2015). Isto é, para todo instante arbritário

n

No caso discreto,

$$\mathbb{P}(X_{n+1} = j|X_n = i, ..., X_0 = i_0) = \mathbb{P}(X_{n+1} = j|X_n = i); \tag{2.2.1}$$

No caso contínuo,

$$\mathbb{P}(X_{n+1} \leq j|X_n \leq i, ..., X_0 \leq i_0) = \mathbb{P}(X_{n+1} \leq j|X_n \leq i). \tag{2.2.2}$$

A probabilidade condicional $\mathbb{P}(X_{n+1} = j|X_n = i)$ é chamada de probabilidade de transição do estado i para j a um passo, $i, j \in S$. Assumindo a propriedade de Markov que se refere à propriedade de perda de memória de um processo estocástico, as propriedades de transição a um passo não variam ao longo do tempo, ou seja, para um k arbitrário:

$$p_{ij} = \mathbb{P}(X_{n+k} = j|X_n = i) = \mathbb{P}(X_k = j|X_0 = i), n = 0, 1, ... \tag{2.2.3}$$

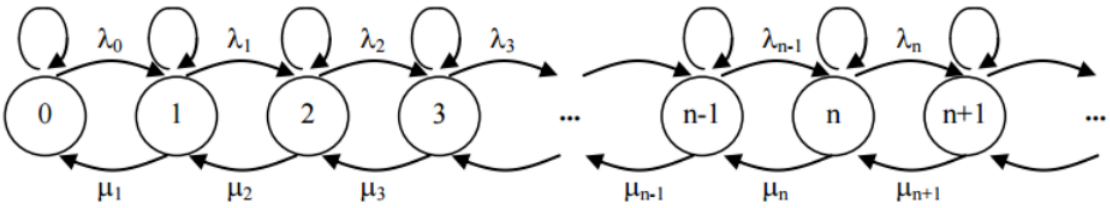
Onde $i, j, p_{ij} \geq 0$ e $\sum_{j=0}^{\infty} p_{ij} = 1$.

Se ambos espaços de estados S e de parâmetros T forem contínuos então o processo de Markov será chamado de processo de Markov de parâmetro contínuo. Se o espaço de estados for contínuo e o de parâmetros discreto ele será dito como um processo de Markov de parâmetro discreto. Quando o espaço de estado for discreto, será dito como uma cadeia de Markov. Os sistemas de fila em que os intervalos entre chegadas e os tempos de serviço seguem a distribuição exponencial podem ser modelados como uma cadeia de markov considerando-se como estado o número de usuários na fila.

2.3 Processos de Nascimento e Morte

Os Processos de Nascimento e Morte são processos de Markov à tempo contínuo com espaço de estados S discreto e com suas transições de estado só ocorrendo entre estados vizinhos, ou seja, estando no estado n, no próximo passo poderá estar apenas em n, $n + 1$ ou $n - 1$.

Figura 1: Grafo do Processo de Nascimento e Morte.



Considere um sistema onde seu estado em um instante é representado pelo número de pessoas que estão no composto naquele instante. Suponha que quando existem n pessoas no sistema:

- Novas chegadas entram no sistema em uma taxa exponencial λ_n, $\lambda_n \in \mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$;
- As pessoas saem do sistema em uma taxa exponencial μ_n, $\mu_n \in \mathbb{R}_+$.

Assim, quando acham-se n pessoas no sistema, o tempo até a próxima chegada tem distribuição exponencial com média $1/\lambda_n$ e independe do tempo até a próxima saída do sistema que segue uma distribuição exponencial com média $1/\mu_n$. Esse sistema é chamado de Processo de Nascimento e Morte. Os parâmetros $\{\lambda_n\}_{n=0}^{\infty}$ é denominado taxa de chegada ou nascimento e $\{\mu_n\}_{n=1}^{\infty}$ é chamado de taxa de saída ou morte (ROSS, 2014).

O tempo entre chegadas e o tempo de serviço são supostos exponencialmente distribuídos com taxas λ e μ, respectivamente. Então, a quantidade de clientes que entram ou saem do sistema têm distribuição Poisson com taxa $\lambda + \mu$. A ligação fluxo de entrada igual ao fluxo de saída resulta nas equações de balanceamento. De acordo com Ross (2014), a matriz geradora ou matriz de intensidade Q descreve o movimento entre os estados e é constituída pela taxa na qual o processo move do estado i para o estado j (q_{ij}), ou seja,

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

O estado i representa a quantidade de pessoas no instante t, e o estado j a quantidade de pessoas no próximo instante, onde q_{ij} é a taxa no qual o processo move do estado i para o estado j. Quando $j = i$ temos que $q_{ij} = -\sum_{j \neq i} q_{ij}$, ou seja,

$$q_{ij} = \begin{cases} \lambda_i, & \text{se } j = i+1; \\ \mu_i, & \text{se } j = i-1. \ (\mu_i \neq 0); \\ -(\lambda_i + \mu_i), & \text{se } j = i. \ (q_0 = \lambda_0); \\ 0, & \text{se } j = i+2 \text{ ou } j = i-2. \end{cases}$$

Pode-se destacar algumas propriedades importantes dos Processos de Nascimento e Morte para o modelo de filas:

1. Se a taxa de ocupação do serviço, $\rho = \frac{\lambda}{\mu} > 1$, em média, o fluxo de clientes que entram no sistema é maior que o fluxo de saídas do sistema. Resultando em um sistema instável, uma vez que se o número de clientes tender para o infinito($N \to \infty$) quando o intervalo de análise é suficientemente grande;

2. Se $\rho < 1$, o fluxo de clientes que saem do sistema é maior que o fluxo de chegada. Indicando que existe uma solução estacionária para o sistema em que todos os clientes que entram no sistema são atendidos em algum momento t, o que torna o sistema estável.

Ao estudar a Teoria das Filas os principais interesses são a variação na quantidade de clientes esperando atendimento, quanto tempo o sistema fica em um dado espaço, e qual caminho ele seguirá no espaço de estados dado o estado atual. A notação p_{ij} conforme equação (2.2.3), é a probabilidade de transição de um estado i para um estado j e pode ser descrita como um processo de Markov de tempo contínuo, com probabilidade de movimento saindo de n somente para o estado $n+1$ ou $n-1$, pois a alteração do estado do sistema ocorre apenas com a chegada ou saída de um cliente. A matriz de transição P reproduz a variação na quantidade de clientes a espera do serviço e é definida por

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ \frac{\mu_1}{(\lambda_1+\mu_1)} & 0 & \frac{\lambda_1}{(\lambda_1+\mu_1)} & 0 & 0 & \cdots \\ 0 & \frac{\mu_2}{(\lambda_2+\mu_2)} & 0 & \frac{\lambda_2}{(\lambda_2+\mu_2)} & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Logo,

$$p_{ij} = \begin{cases} \frac{\lambda_i}{(\lambda_i+\mu_i)}, & \text{se } j = i+1 \ \forall i \geq 1, \\ \frac{\mu_i}{(\lambda_i+\mu_i)}, & \text{se } j = i-1 \ \forall i \geq 1, \\ 1, & \text{se } i = 0, \ j = 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Um Processo de Nascimento e Morte com um padrão constante de chegadas e

saídas possui a distribuição de probabilidade independente do tempo. Os modelos de filas apresentam distribuição constante ao longo do tempo, exceto quando

- O número médio de entradas por unidade de tempo(λ) e o número médio de saídas (μ) variam ao longo de T.
- O número esperado de clientes no sistema tende ao infinito.
- $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P_{ij} \neq 1$.

A probabilidade do sistema estar ocioso é representada por P_0 e a probabilidade do sistema possuir n clientes é denotada por P_n. As estimações dos valores de P_n e P_0 são obtidas quando o sistema é estável (todos os pacientes são atendidos em algum momento t) através das relações entre λ_i e μ_i. Por conseguinte, segundo (ROSS, 2014)

$$\lambda_{j-1}P_{j-1} + \mu_{j+1}P_{j+1} - (\lambda_j + \mu_j)P_j = 0, \text{ para } j \geq 1. \tag{2.3.1}$$

Igualando $j = 0$ na equação (2.3.1)

$$P_1 = \frac{\lambda_0}{\mu_1}P_0. \tag{2.3.2}$$

Isolando P_{j+1} em (2.3.1)

$$P_{j+1} = \frac{(\lambda_j + \mu_j)P_j}{\mu_{j+1}} - \frac{\lambda_{j-1}P_{j-1}}{\mu_{j+1}} \text{ para } j = 0, 1, 2, \cdots, n-1. \tag{2.3.3}$$

Isolando a equação (2.3.3) em função de P_0 para $j = \{0, 1, 2, \cdots, n-1\}$ temos

- $j = 1: \ P_1 = \frac{\lambda_0}{\mu_1}P_0$
- $j = 2: \ P_2 = \frac{(\lambda_1+\mu_1)P_1}{\mu_2} - \frac{\lambda_0}{\mu_2}P_0 = \frac{\lambda_1\lambda_0P_0}{\mu_1\mu_2}$

  $\vdots$
- $j = (n-1): \ P_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}P_0$

Com isso, podemos reescrever P_n em função de P_0 como

$$P_n = P_0\prod_{i=1}^{n}\left(\frac{\lambda_{i-1}}{\mu_i}\right), \ n \geq 1. \tag{2.3.4}$$

Como a soma das probabilidades é restrita e igual a 1 ($\sum_{n=0}^{\infty} P_n = 1$) temos

$$1 = P_0 + \sum_{n=1}^{\infty} \left[P_0 \prod_{i=1}^{n} \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right];$$

$$1 = P_0 + P_0 \sum_{n=1}^{\infty} \left[\prod_{i=1}^{n} \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right].$$

Consequentemente

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left[\prod_{i=1}^{n} \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right] \right)^{-1} \tag{2.3.5}$$

2.4 Processos de Poisson

2.4.1 Processos de Poisson Homogêneos

Um processo estocástico $\{X_t,\ t \in T\}$ onde X_t representa o número total de chegadas ocorridos até o instante t é um processo de Poisson Homogêneo com taxa $\lambda \in \mathbb{R}$ (número de clientes que chegam ou são atendidos por unidade de tempo), que assume valores inteiros não negativos tal que:

1. $X_0 = 0$ e $X_t > 0$;

2.Se $s < t$, então $X_s < X_t$;

3. $X_t - X_s$ é o número de chegadas que o correm no intervalo $(s,\ t]$.

Para qualquer coleção de instantes $t_0 < t_1 < ... < t_n$, os incrementos do processo $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ são estacionários, ou seja, a distribuição de $X_{s+t} - X_s$ para um tempo arbitrário $t \geq 0$, é independente do número de chegadas até o tempo t.

Para $s \geq 0$ e $t > 0$, a variável aleatória $X_{s+t} - X_s$ segue distribuição Poisson com taxa λ_t e sua distribuição é

$$\mathbb{P}(X_{s+t} - X_s = n) = \frac{e^{-\lambda_t}(\lambda_t)^n}{n!},\ \text{para } n = 0, 1, ...$$

onde n é o número de chegadas que ocorrem no intervalo $(s,\ s + t]$.

Assumindo que X_t é um Processo de Poisson com taxa $\lambda_t > 0$, a média e a variância de X_t são iguais a

$$E[X_t] = VAR(X_t) = \lambda_t.$$

2.5 Distribuição Exponencial

A distribuição exponencial possui a propriedade de não apresentar memória, isto é, o próximo estado só depende do estado atual e não dos estados anteriores.

Se os intervalos entre chegadas são distribuídos exponencialmente com média $1/\lambda$, o tempo esperado para a próxima chegada sempre será $1/\lambda$, independentemente do tempo transcorrido antes da última chegada (BRESSAN, 2002).

Uma variável aleatória X é dita ter distribuição exponencial com parâmetro λ, $\lambda > 0$ se sua função densidade de probabilidade é dada por:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{se } t \geq 0, \\ 0, & \text{se } t < 0. \end{cases} \tag{2.5.1}$$

Ou, equivalentemente, se sua função de distribuição acumulada $F(x)$ é dada por:

$$F(t) = \mathbb{P}(T \leq t) = \begin{cases} 1 - e^{-\lambda t}, & \text{se } t \geq 0, \\ 0, & \text{se } t < 0. \end{cases} \tag{2.5.2}$$

A média e variância da distribuição exponencial são dadas por

$$E(X) = \frac{1}{\lambda} \text{ e } \mathrm{VAR}(X) = \frac{1}{\lambda^2}. \tag{2.5.3}$$

A distribuição Exponencial possui a propriedade de não apresentar memória, ou seja, o próximo estado depende apenas da diferença entre os instantes desejado e atual, pois

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s), \text{ para todo } t,\ s \geq 0. \tag{2.5.4}$$

2.6 Teste Kolmogorov-Smirnov de Aderência

Para verificar se os instantes seguem distribuição Exponencial para os tempos entre as chegadas é realizado o teste Kolmogorov-Smirnov de aderência.

As hipóteses do teste são

$$\begin{cases} H_0 : \text{A amostra segue uma distribuição Exponencial,} \\ H_1 : \text{A amostra não segue uma dist. Exponencial.} \end{cases}$$

Dada uma amostra aleatória $x_1, x_2, ..., x_n$, $n \in \mathbb{N}$ referente a uma variável aleatória

X, a estatística do teste é a diferença máxima absoluta entre as funções de distribuição acumulada teórica e empírica (SAMEJIMA, 2021)

$$D = max_x|S_n(x) - F_0(x)|. \tag{2.6.1}$$

Se $X \sim Exp(\lambda)$, então:

1. $F_0(x) = \int_0^z e^{-\lambda u}\, du = 1 - e^{-\lambda z}, (z \leq n)$,
2. $S_n(x) = \frac{N(x)}{n}$, onde $N(x)$ é o número de observações $\leq x_n$.

Quando $F_0(X)$ é contínua e a hipótese nula é verdadeira, a distribuição exata de D é apresentada na tabela para os quantis do teste de Kolmogorov Smirnov (CONOVER, 1998) quando $n \leq 40$ e uma aproximação assintótica é usada para $n > 40$.

Rejeita-se a hipótese nula ao nível de significância α se D excede o quantil de ordem $(1 - \alpha)$, ou com base no p-valor que depende das hipóteses consideradas.

2.7 Gráficos Q-Q

Os gráficos Q-Q, ou gráficos quantil-quantil são utilizados como suporte para julgar se um conjunto de dados veio eventualmente de alguma distribuição teórica, como Normal ou Exponencial. É apenas uma confirmação visual, não uma prova complexa, por isso é um pouco abstrata, mas nos permite ver rapidamente se nossa suposição é plausível e, se não, como a suspeita é violada e quais pontos dos dados contribuem para o descumprimento.

Um gráfico Q-Q é um gráfico de dispersão criado pela plotagem de dois conjuntos de quantis um contra o outro. Se ambos os grupamentos vierem da mesma distribuição, os pontos devem formar uma linha aproximadamente reta.

Os gráficos Q-Q classificam os dados da amostra em ordem crescente e, em seguida, os distribui no gráfico em relação aos quantis calculados a partir de uma distribuição teórica. O número de quantis é selecionado para corresponder ao tamanho dos dados da amostra.

2.8 Sistema de Filas

Os sistemas de filas são descritos, de forma geral, pelo processo de chegada de clientes a um conjunto de atendimento, realizados por uma certa quantidade de servido-

res. As formações de filas ocorrem quando a demanda por assistência é maior do que a capacidade de suporte do sistema. A estrutura básica de um sistema de filas consiste de uma fonte que é a população de todos os usuários (unidade que requer atendimento), da fila que são os clientes em espera pelo atendimento, e o canal de atendimento que é o processo ou sistema que realiza a assistência ao cliente.

Os tipos de modelos de filas são definidos a partir da Notação de Kendall, que representa cada cadeia de filas da seguinte forma (KENDALL, 1953)

$$A/S/m/k/N/Q,$$

onde,

A: Distribuição dos tempos entre as chegadas;

S: Distribuição dos tempos de serviços;

m: Número de servidores no sistema ($m \in N$);

k: Capacidade do sistema;

N: Tamanho da população;

Q: Disciplina da fila.

Ao suprimir os três últimos símbolos descritores da fila, acorda-se que a capacidade do sistema é ilimitada, a população é infinita e a disciplina da fila segue o critério FIFO (First in First Out) ou FCFS (First Come First Served), em que o primeiro a chegar é o primeiro a ser atendido.

Quando um processo está no estado estacionário, são aplicáveis as fórmulas de Little, que afirmam que

$$L = \lambda W, \tag{2.8.1}$$

onde $L = \sum_{n=0}^{\infty} nP_n$ (número médio de clientes no sistema), λ representa o número médio de clientes que chegam por unidade de tempo e W o tempo médio de espera no sistema.;

$$L_Q = \lambda W_Q, \tag{2.8.2}$$

onde L_Q é o número médio de clientes em espera por atendimento e W_Q o tempo médio de espera na fila.

2.9 Fila Única M/M/1

As Filas Markovianas são as filas únicas em que número de chegadas e atendimentos por unidade de tempo são processos de Poisson, são indicadas de forma genérica como M/M/m. Em que a distribuição do tempo entre novas chegadas ao sistema segue uma distribuição exponencial (M) e o tempo necessário para a realização do serviço também segue uma distribuição exponencial (M). Neste caso, consideramos que o sistema se encontra em uma posição de equilíbrio em que a taxa de chegada (λ) e as taxas de atendimento dos servidores (μ) não se alteram (BRESSAN, 2002).

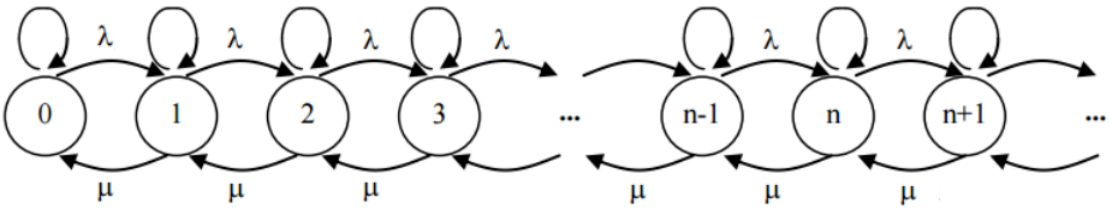
Os sistemas M/M/1 da notação de Kendall correspondem ao modelo básico e mais utilizado como modelo de fila Markoviana. A capacidade máxima do sistema e o tamanho da população são considerados infinitos. Nas aplicações desse modelo, temos:

- Taxa média de chegada = λ;
- Taxa média de atendimento = μ;
- Número de atendentes = 1;
- Disciplina de atendimento = FIFO.
- λ_n: a taxa média de chegada de novos clientes, ou número esperado de chegadas por tempo unitário, quando n clientes estão no sistema;
- μ_n: o tempo esperado de atendimento dos clientes sendo atendidos no sistema;

A função densidade para o tempo entre as chegadas é dada por $f(t) = \lambda e^{-\lambda t}$ e a função densidade para os tempos de atendimento é dada por $f(t) = \mu e^{-\mu t}$. Onde o tempo médio entre chegadas e o tempo médio de serviço são dados por $1/\lambda$ e $1/\mu$, respectivamente, sendo os tempos independentes.

O modelo M/M/1 é um processo de nascimento e morte onde a estimação para um determinado estado independe dos estados passados, portanto, as chegadas são consideradas como os nascimentos e as conclusões dos serviços são consideradas com mortes. As taxas de chegada e serviço independem do estado do sistema, isto é, $\lambda_n = \lambda$ e $\mu_n = \mu$, para todo $n = \{1, 2, 3, ...\}$.

Figura 2: Diagrama de transição de estados do Sistema M/M/1.



Utilizando-se as equações de balanceamento (Seção 2.3), temos que Fluxo de Entrada = Fluxo de Saída, ou seja, todos os pacientes que entram na fila serão atendidos, logo, para que haja um estado de equilíbrio é necessário que o número de entradas e de saídas por unidade de tempo sejam iguais. Com isso, podemos calcular P_{i+1}, equação (2.3.4), de cada estado i (número de clientes no sistema), $i \geq 0$ em função de P_0. Os passos são descritos na seguinte tabela:

Tabela 1: Dedução de P_n para o modelo M/M/1.

Estado	Taxa de Entrada	=	Taxa de Saída	P_{i+1}
0	μP_1	=	λP_0	$P_1 = \frac{\lambda}{\mu} P_0$
1	$\lambda P_0 + \mu P_2$	=	$\lambda P_1 + \mu P_1$	$P_2 = \frac{1}{\mu}(\lambda + \mu)P_1 - \lambda P_0 = \frac{\lambda^2}{\mu^2} P_0$
2	$\lambda P_1 + \mu P_3$	=	$\lambda P_2 + \mu P_2$	$P_3 = \frac{\lambda^3}{\mu^3} P_0$
⋮				
n-1	$\lambda P_{n-2} + \mu P_n$	=	$\lambda P_1 + \mu P_1$	$P_n = \frac{\lambda^n}{\mu^n} P_0$
n	$\lambda P_{n-1} + \mu P_{n+1}$	=	$\lambda P_n + \mu P_n$	$P_{n+1} = \frac{1}{\mu}(\lambda + \mu)P_n - \frac{\lambda}{\mu} P_{n-1} = \frac{\lambda}{\mu} P_n$

Por indução, pode-se mostrar que

$$P_{n+1} = \left(\frac{\lambda}{\mu}\right)^{n+1} P_0, \, \forall n \geq 0. \tag{2.9.1}$$

Por definição temos que $\sum_{n=0}^{\infty} P_n = 1$. Então

$$\begin{aligned} 1 - \sum_{n=0}^{\infty} P_n = \\ = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n P_0 \\ = P_0 \frac{1}{1 - \lambda/\mu}. \end{aligned}$$

Com isso temos que $P_0 = 1 - \frac{\lambda}{\mu}$.

Baseado nesta solução e fazendo $p = \frac{\lambda}{\mu}$, podemos derivar os principais parâmetros do sistema M/M/1:

- Fator de utilização do servidor(taxa de ocupação média):

$$U = \frac{\lambda}{\mu};$$

- Número médio de clientes no sistema:

$$L = E(n) = \sum_{n=1}^{\infty} nP_n = \sum_{n=0}^{\infty} n(1-p)p^n = \frac{p}{(1-p)};$$

- Variância do número de clientes no sistema:

$$Var(n) = E(n^2) - E(n)^2 = \left[\sum_{n=1}^{\infty} n^2(1-p)p^n\right] - E(n)^2 = \frac{p}{(1-p)^2};$$

- Probabilidade de se ter n ou mais clientes no sistema:

$$\mathbb{P}_{\geq n} = \sum_{j=n}^{\infty} P_j = \sum_{j=n}^{\infty} p^j(1-p) = p^n;$$

- Tempo médio de cada unidade no sistema:

  Pela fórmula de Little (2.8.1),

$$W = \frac{L}{\lambda} = \frac{p}{(1-p)}\frac{1}{\lambda} = \frac{\frac{1}{\mu}}{1-p};$$

- Tempo médio de cada unidade na Fila: Seja $E(S) = \frac{1}{\mu}$ o tempo médio de atendimento, temos:

$$W_Q = W - E(S) = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)};$$

- Número médio de clientes na fila:

  Pela fórmula de Little (2.8.2),

$$L_Q = \lambda.W_Q = \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

2.10 Modelo M/M/c

Neste caso temos que o número de atendentes é igual a c. Cada servidor possui distribuição de tempo de serviço exponencial independente e identicamente distribuídas. Novamente o processo de chegada é assumido ser Poisson.

Sob essas colocações tem-se um processo de nascimento e morte, onde as razões de chegadas não dependem do estado do sistema, e as razões de saída mudam de acordo com o número de clientes. São definidas como

$\lambda_n = \lambda, n \geq 0;$

$$\mu_n = \begin{cases} n\mu, \ 0 \leq n < c; \\ c\mu, \ n \geq c \text{ (se todos os servidores estão ocupados).} \end{cases} \tag{2.10.1}$$

Aplicando as restrições definidas em (2.10.1) na equação (2.3.4), obtemos

$$P_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} P_0, & 1 \leq n < c, \\ \frac{\lambda^n}{c^{n-c}c!\mu^n}, & n \geq c. \end{cases} \tag{2.10.2}$$

Utilizando a restrição de que as probabilidades devem somar um, pode-se encontrar P_0

$$P_0 = \left(\frac{\sum_{n=0}^{c-1} \lambda^n}{n!\mu^n} + \frac{\sum_{n=c}^{\infty} \lambda^n}{(c^{n-c}c!\mu^n)} \right)^{-1}.$$

Utilizando as expressões para somade uma progressão geométrica, e fazendo $r = \lambda/\mu$ e $\rho = \lambda/c\mu$, podemos reescrever P_0 como

$$P_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)} \right)^{-1}, \ \rho < 1. \tag{2.10.3}$$

Para um estado de equilíbrio, a taxa média de chegada deve ser menor que a taxa média de serviço do sistema. Se aplicarmos $c = 1$ nas fórmulas de P_0 e P_n reduzimos as equações para o sistema M/M/1. Baseando-se nas probabilidades de estado de equlíbrio dada nas equações (2.10.2) e (2.10.3) podemos encontrar as medidas de desempenho desse modelo de maneira similar ao sistema M/M/1.

- Número médio de clientes na fila:

$$L_Q = \sum_{n=c+1}^{\infty} (n-c)P_n = \sum_{n=c+1}^{\infty} (n-c)\frac{r^n}{c^{n-c}c!}P_0 = \frac{r^c}{c!}P_0 \sum_{n=c+1}^{\infty} (n-c)\frac{r^{n-c}}{c^{n-c}} = \frac{(r^c\rho)}{c!(1-\rho)^2}P_0;$$

- Tempo médio de cada unidade na Fila:

  Pela fórmula de Little (2.8.2),

$$W_Q = \frac{L_Q}{\lambda};$$

- Número médio de clientes no sistema:

  Pela fórmula de Little (2.8.1),

$$L = \lambda W = L_Q + r;$$

- Tempo médio de cada unidade no sistema:

$$W = W_Q + \frac{1}{\mu}.$$

3 Leitos de UTI na rede pública de saúde do Distrito Federal

A Unidade de Terapia Intensiva foi criada no início do século XX a partir do aperfeiçoamento das "Salas de Recuperação Pós-Anestésica" para pacientes submetidos à Neurocirurgia no Hospital Americano Johns Hopkins, os pacientes eram classificados de acordo com o nível de dependência, de tal forma que os mais próximos à área de atividade da enfermagem fossem os pacientes mais graves, para vigília mais frequente e um melhor atendimento (SCHLINZ, 2016). No Brasil, a implantação da primeira UTI foi na década de 70, no Hospital Sírio Libanês em São Paulo.

As UTI's podem ser classificadas em Adulto, Pediátrica, Neonatal e as UTI's Especializadas, dentre elas destacam-se: Cardiológica ou Coronariana, Cirúrgica, Neurocirúrgica e Clínica.

A UTI Coronariana é voltada exclusivamente para o tratamento de pacientes com doenças cardíacas ou pacientes que passam por procedimentos como Coronariografia Digital, Marcapasso, Cirurgia de Revascularização do Miocardio, etc.

Existem na UTI médicos, enfermeiros, técnicos de enfermagem e profissionais de apoio, como nutricionistas, psicólogos, assistente social e farmacêuticos. Na UTI também é permitido a admissão de religiosos para assistência dos pacientes sob solicitação dos mesmos ou de seus parentes. Os Cuidados Intensivos são muito abrangentes, e destacam-se na atualidade para as internações causadas pelo vírus SARS-CoV2 que pode ser transmitido, principalmente, de pessoa para pessoa por meio de gotículas do nariz ou da boca que se espalham quando uma pessoa com COVID-19 tosse, espirra ou fala.

Pacientes com COVID-19 geralmente apresentam sintomas respiratórios, especialmente nas formas graves da doença, o que leva a um maior risco de morte. Durante a pandemia, a demanda por leitos de UTI cresceu de forma proporcional ao aumento do número de casos, levando a lotações e esgotamento dos bens hospitalares, principalmente dos respiradores mecânicos que tem a função de proporcionar a oxigenação dos pulmões, quando estes são gravemente afetados.

O primeiro caso da pandemia pelo novo coronavírus, SARS-CoV2, foi identificado em Wuhan, na China, no dia 31 de dezembro de 2019. Desde então, os casos começaram a se espalhar rapidamente pelo mundo. Em 28 de fevereiro de 2020, por meio do Decreto Nº 40.475, foi declarado cenário de emergência no Distrito Federal. O primeiro caso da doença pelo novo Coronavírus no DF foi confirmado em 5 de março de 2020. Em

11 de março de 2020, o contágio pelo SARS-CoV-2 passou a ser classificado pela OMS como pandemia. Em 20 de março de 2020 o Ministério da saúde anunciou o estado de transmissão comunitária da doença em todo o país, daí em diante o Brasil passou por 3 ondas de casos de Covid-19.

Figura 3: Novos casos de COVID-19 no DF por dia.

Fonte: Our World in Data, 2022

A taxa de ocupação dos leitos públicos das Unidades de Terapia Intensiva (UTI) para tratar pacientes com confirmação ou suspeita de covid-19 no final de janeiro de 2022 voltou a atingir níveis preocupantes no Distrito Federal. Segundo o portal InfoSaúde[1], da Secretaria de Saúde do DF (SES-DF), a taxa de ocupação estava em 97,62%, sendo que os leitos adultos e pediátricos atingiram 100% de ocupação no período.

As unidades de terapia intensiva são importantes para o tratamento e supervisão de pacientes em estado grave a fim de promover a restauração de sua saúde. No entanto, caso não sejam atenciosamente administradas podem resultar em grande tempo de internação. Como consequência disso, há a produção de expensas excessivas para o hospital, a depreciação da saúde dos pacientes e desacolhimento de indivíduos em estado crítico devido a menor desocupação de leitos. Dessa forma, para que essa recepção seja de fácil acesso aos pacientes que dela necessitam é indispensável que se faça o acompanhamento do indicador de média de permanência de UTI e o número total de leitos, aspirando a prática de mediações para que a permanência do paciente seja desempenhada da maneira mais rápida e eficiente possível.

Em abril de 2021 houve a maior quantidade de leitos públicos de terapia intensiva (UTI) para atendimento à COVID-19, totalizando um número de 482 leitos. O primeiro plano de desmobilização de leitos COVID foi exposto no dia 05 de outubro de 2020

[1]https://info.saude.df.gov.br/leitospublicosutigeraissalasit/

e não concluiu todas as suas fases finais de ativação da desmobilização em função do aumento de casos do final do ano. Em dezembro de 2020 um segundo Plano foi proposto, com o intuito de organizar as ações a serem desenvolvidas pela SES/DF na mobilização dos leitos destinados ao suporte dos pacientes hospitalizados por COVID-19 por meio de um processo mais eficiente e adaptável, empenhando-se em se ajustar ao panorama epidemiológico para que se pudesse alternar a quantidade de leitos em períodos de alta ou de queda da disseminação da doença e do adoecimento dos residentes do DF (Plano de Mobilização de Leitos COVID-19 do Distrito Federal, 2021).

A SES-DF monitora em tempo real a situação das vagas em UTI no Distrito Federal e ativa as fases do Plano de Mobilização de leitos, conforme a necessidade de cada momento.

4 Metodologia

Este trabalho foi dividido em duas porções: na primeira foi realizado um estudo teórico do conteúdo acerca da Teoria de Filas e, num segundo momento, foi feita uma aplicação dessa teoria num conjunto de dados do sistema público de saúde do Distrito Federal.

Para a obtenção dos dados foram realizadas capturas de telas das informações e depois a conversão[2] dessas imagens para arquivos de texto, pois a opção de download dos dados era inexistente. Todos os cálculos e gráficos presentes neste trabalho produzidos pelos autores foram feitos 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 04 de fevereiro até 03 de março, período em que a terceira onda de casos da pandemia estava ocorrendo no DF. A atualização da informações é realizada periodicamente durante o dia pela Central de Regulação da Internação Hospitalar[3], a partir da atualização do estado clínico do paciente.

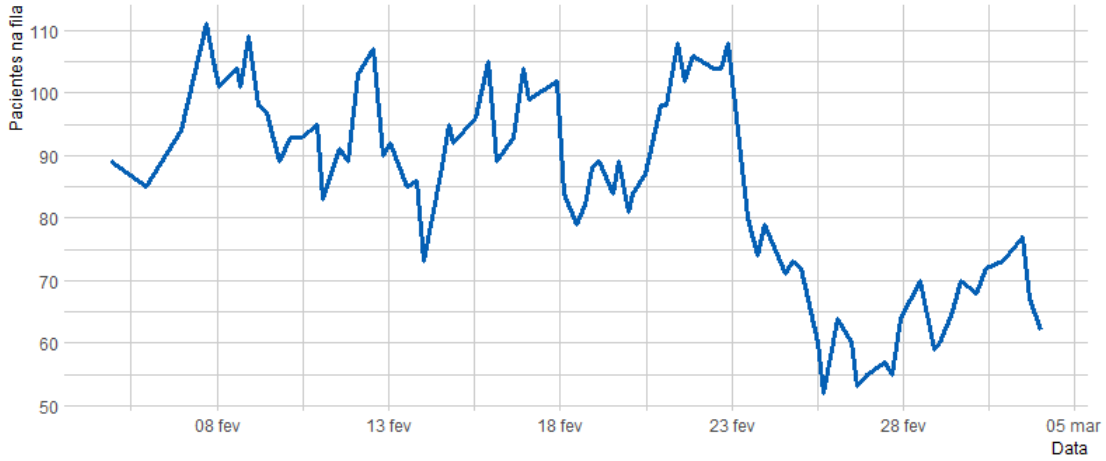
4.1 Conjunto de dados

Nesse estudo de caso, será analisado o processo de geração de filas por pacientes a espera por leitos de UTI na rede pública de saúde do Distrito Federal. O objetivo é observar o número de pessoas em espera e propor melhorias no desempenho do sistema para que o atendimento seja efetuado mais rápido e diminua a probabilidade de perdas e complicações para os pacientes em espera. 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.

Os dados tratam-se de tempos entre chegadas e de atendimento. Tendo em vista que a lista de pacientes em espera foi monitorada em intervalos, os dados coletados não estão completos, uma vez que, a lista é atualizada com uma peridiocidade pequena, irregular e sua versão anterior fica indisponível a cada nova atualização.

[2]https://www.invertexto.com/converter-imagem-em-texto
[3]https://info.saude.df.gov.br/lista-de-espera-por-leitos-de-uti/

Figura 4: Numéro de pacientes na fila de espera.



Para cada paciente tem-se o horário de chegada, o horário de saída estimado, o subtipo de leito (Coronariano, Neurocirúrgico, Clínico, Cirúrgico, Materno e Não Informado), se o paciente era Suspeito/Confirmado com COVID-19 e se o Suporte Dialítico seria necessário em seu leito. O posicionamento na lista de espera na UTI obedece aos critérios de prioridade (PRI) de acordo com o quadro clínico do paciente e do tipo de leito solicitado, ou seja, os que estão no topo da lista não serão, necessariamente os primeiros a serem direcionados aos leitos, pois dependem do tipo de leito disponível. Na tabela abaixo, para melhor vizualização dos dados temos 5 dos 1092 pacientes observados no período.

Tabela 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.

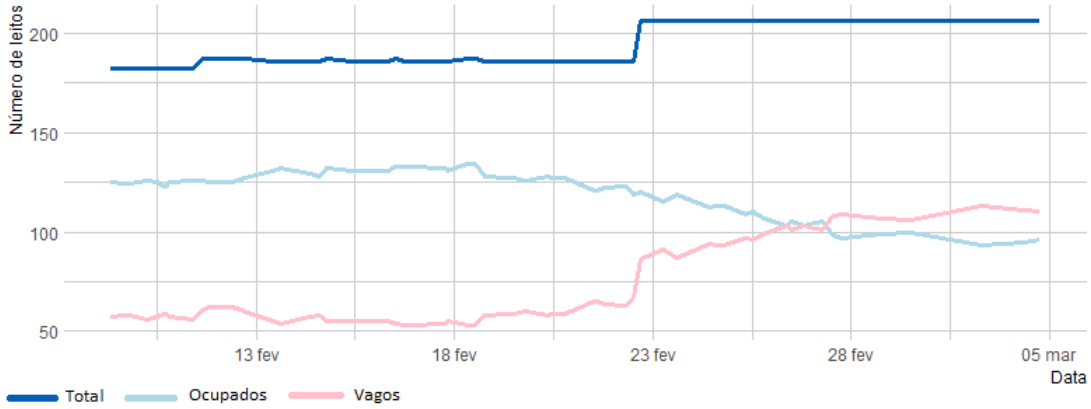
A informação da Taxa de ocupação Total dos leitos COVID também foi coletada no mesmo período, em que o método de Cálculo da taxa de ocupação total de leitos possui o somatório de pacientes internados no dia (total de leitos ocupados) como numerador e o somatório de Leitos Operacionais no dia (total de leitos COVID menos

bloqueados/aguardando liberação) como Denominador.

Figura 5: Taxa de ocupação total dos leitos COVID.

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

Figura 6: Número de leitos COVID por dia.



No dia 22 de fevereiro houve um 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, o que é justificável devido a uma menor quantidade de novos casos diários de COVID-19 no DF e o aumento do 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.

5 Simulação para o Modelo M/M/1

Neste capítulo serão realizadas simulações com números aleatórios para o modelo M/M/1 com o intuito de obter estimativas para o valor do tempo médio de espera em fila, o tamanho médio da fila e a taxa de ocupação do único servidor do modelo.

Essas simulações serão realizadas através do software R (versão 4.0.5) e com base nos códigos adaptados disponíveis em anexo (ARAÚJO, 2015). Também são obtidos os intervalos de confiança para as estimativas e comparação com os resultados das fórmulas teóricas do modelo.

5.1 Estimação

As simulações serão feitas de acordo com a tolerância do erro estipulado. O valor do número de simulações será calculado a partir da fórmula que maximiza a variância pelo método conservativo:

$$n = \left(\frac{z_{95\%} \times \sigma}{\epsilon} \right)^2 = 751.5398,$$

onde $z_{95\%}$ é o quantil da normal com 95% de confiança, $\epsilon = 0.015$ é o erro de precisão, $\sigma = \frac{1}{\lambda^2} = 0.25$ a variância e n o número total de simulações. Com isso temos que o sistema será simulado 752 vezes.

Neste caso foram gerados números aleatórios e independentes em que a distribuição do tempo entre novas chegadas ao sistema segue uma distribuição exponencial ($\lambda = 2$) e o tempo necessário para a realização do serviço também segue uma distribuição exponencial ($\mu = 4$). O tempo total de simulação será 400, logo o número médio de clientes será $2 \times 400 = 800$. Quanto maior o número médio de clientes mais próximas serão as medidas empíricas das teóricas.

Tabela 3: Estimativas da simulação e valor teórico obtido através da fórmula.

Estimativa	Fórmula	Teórica	Simulada
Tempo de espera médio na fila	$\frac{\lambda}{\mu(\mu-\lambda)}$	0,25	0,2471
Taxa de ocupação média	$\frac{\lambda}{\mu}$	0,5	0,4997
Tamanho médio da fila	$\frac{\lambda^2}{\mu(\mu-\lambda)}$	0,5	0,4954

Fonte: Autores, 2022.

Vemos que as estimativas resultantes das simulações são bem próximas das teóricas.

5.2 Intervalos de Confiança

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 serão calculados usando a fórmula

$$IC = \left[\bar{x} - z_{\alpha/2} \times \sqrt{\sigma^2},\ \bar{x} + z_{\alpha/2} \times \sqrt{\sigma^2}\right], \tag{5.2.1}$$

onde $\bar{x}$ representa a respectiva estimativa ponderada pelo número de simulações. As variâncias do tempo de espera médio na fila, taxa de ocupação média e tamanho médio da fila são respectivamente, 0.0451^2, 0.0251^2 e 0.098^2.

Tabela 4: Intervalos de Confiança para as estimativas médias.

Estimativa	Limite Inferior	Limite Superior
Tempo de espera médio na fila	0,15859	0,33564
Taxa de ocupação média	0,45040	0,54912
Tamanho médio da fila	0,30331	0,68768

Fonte: Autores, 2022.

Uma vez que a variância do tamanho médio da fila era a maior entre as estimativas esperava-se que para seu intervalo de confiança a amplitude do intervalo seria maior, o que de fato ocorreu. Por fim, todos os intervalos contém o valor real das estimativas.

6 Resultados

6.1 Seleção do modelo

Para os últimos 10 pacientes dos 1092 que ingressaram na Fila, os horários de chegada e o intervalo entre a chegada do Paciente atual e o anterior são descritos na Tabela 5.

Tabela 5: Pacientes,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

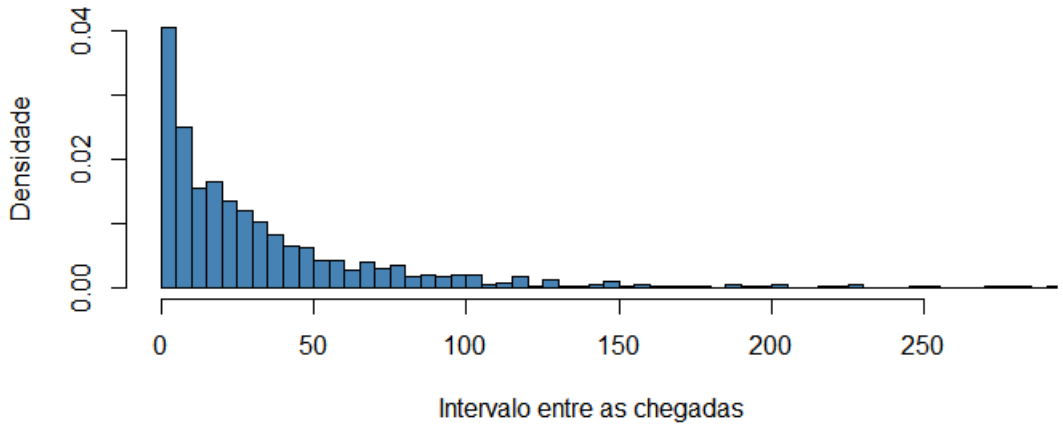
O número de pacientes que chegam por minuto e o intervalo médio entre chegadas, para os 27 dias e meio analisados, são respectivamente,

$$\hat{\lambda} = \frac{\text{Total de pacientes}}{\text{Tempo analisado (em minutos)}} = 0,02589 \quad e \quad \bar{x}_1 = \frac{\text{Soma dos intervalos}}{\text{Total de pacientes}} = 38,8.$$

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.

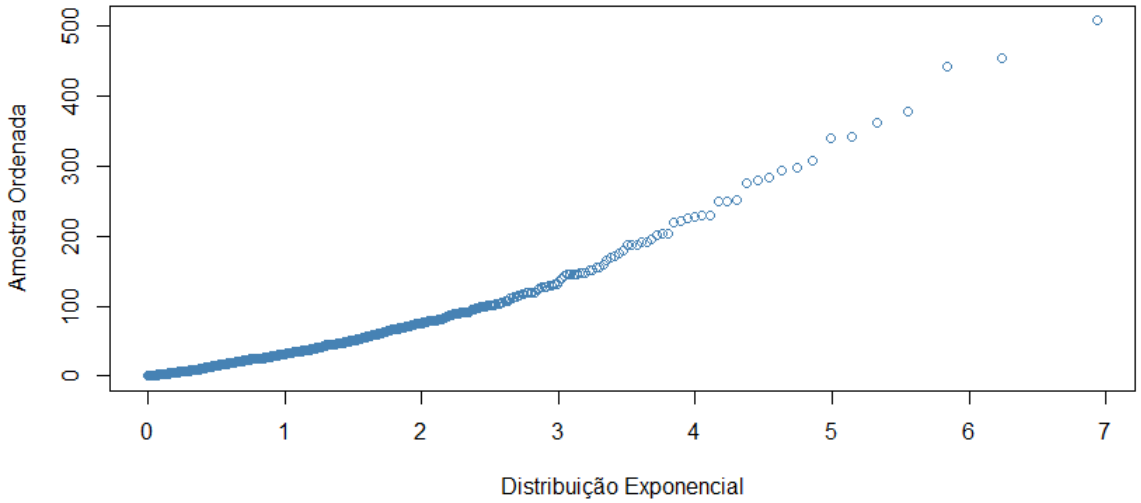
Antes do início da coleta de dados já haviam 63 pacientes na fila com horários de chegadas disponíveis, entretanto foram desconsiderados nos cálculos acima pois seus intervalos entre chegadas estavam imcompletos, uma vez que a maior parte dos pacientes que entraram na fila em períodos próximos já haviam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

Figura 7: Histograma dos intervalos entre chegadas.



Para verificar se o intervalo entre as chegadas possui distribuição Exponencial, foi realizada uma análise gráfica por meio de um gráfico Q-Q.

Figura 8: Gráfico Q-Q da distribuição dos intervalos entre chegadas.



Como as observações estão distribuidas linearmente podemos assumir que os intervalos entre as chegadas seguem uma distribuição Exponencial.

O número médio de pacientes atendidos por minuto e o intervalo médio entre atendimentos é de:

$$\hat{\mu} = \frac{\text{Pacientes atendidos}}{\text{Tempo analisado (em minutos)}} \quad e \quad \bar{x}_2 = \frac{\text{Tempo analisado (em minutos)}}{\text{Pacientes atendidos}}.$$

$$\hat{\mu} = \frac{1092 - 62}{27,5} \times \frac{1}{24} \times \frac{1}{60} = 0,02601 \quad e \quad \bar{x}_2 = \frac{27,5 \times 24 \times 60}{1030} = 38,45.$$

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 $= 0,1076$ obtido a um nível de significância de 5% aceita-se que os dados possam ser representados pelo modelo Poisson com média 37,45 e iremos assumir que o tempo de atendimento possui uma distribuição Exponencial com taxa $\hat{\mu} = 0,02601$.

O modelo M/M/1 será usado nesse estudo com $\hat{\lambda} = 0,02589$ número de chegadas por minuto e $\hat{\mu} = 0,02601$ número de atendimentos por minuto.

6.2 Modelo M/M/1

Para a aplicação desse modelo tem-se que a capacidade máxima do sistema e o tamanho da população são considerados infinitos, além disso, há apenas um único servidor e a ordem de atendimento segue a disciplina FIFO. A probabilidade do sistema estar vazio e a proporção do tempo em que há n pacientes na lista de espera, são, respectivamente,

$$P_0 = 1 - \frac{\hat{\lambda}}{\hat{\mu}} = 1 - 0,9954 = 0,0046;$$

$$P_n = \left(1 - \frac{\hat{\lambda}}{\hat{\mu}}\right)\left(\frac{\hat{\lambda}}{\hat{\mu}}\right)^n = (1 - 0,9954)(0,9954)^n = (0,0046)(0,9954)^n;$$

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

- Taxa de ocupação média:

$$\hat{U} = \frac{\hat{\lambda}}{\hat{\mu}} = 0,9954;$$

- Número médio de clientes no sistema:

$$\hat{L} = \frac{p}{(1-p)} = 216,84;$$

- Probabilidade de se ter n=100 ou mais clientes no sistema:

$$P_{\geq n} = p^n = p^{100} = 0,631;$$

- Tempo médio de cada unidade no sistema (minutos):

$$\hat{W} = \frac{\frac{1}{\hat{\mu}}}{1-p} = 8375,172;$$

  ou seja, 139,5862 horas ou 5 dias, 19 horas e 35 minutos.

- Tempo médio de cada unidade na Fila:

$$\hat{W}_Q = \frac{\hat{\lambda}}{\hat{\mu}(\hat{\mu}-\hat{\lambda})} = 8336,726;$$

  o que equivale a 138,9454 horas ou 5 dias, 18 horas e 57 minutos.

- Número médio de clientes na fila:

  Pela fórmula de Little (2.8.2),

$$\hat{L}_Q = \hat{\lambda}.\hat{W}_Q = \frac{\hat{\lambda}^2}{\hat{\mu}(\hat{\mu}-\hat{\lambda})} = 216.$$

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 usando a fórmula (5.2.1).Para o cálculo do intervalo de confiança do tempo médio de cada paciente no sistema foi usado o desvio padrão igual à $\frac{1}{(\hat{\lambda}+\hat{\mu})^2}$ (equação (2.5.3)), considerando que o tempo no sistema segue uma distribuição Poisson$(\hat{\lambda}+\hat{\mu})$ que é a soma das distribuições do tempo de atendimento e do tempo em fila..

Tabela 6: Intervalos de Confiança para as estimativas médias 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

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 da depreciação da saúde do paciente, maximizando o atendimento, uma parametrização em que existam 2 servidores é proposta.

6.3 Modelo M/M/2

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 de direito público ou convênio, tendo preferências 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

$$P_0 = \left(1 + 0,995 + \frac{0,995^2}{2(1-0.498)}\right) = 0,3354;$$

$$P_n = \begin{cases} \frac{(0,02589)^n}{n!(0,02601)^n} \times 0,3354, & 1 \leq n < 2, \\ \frac{(0,02589)^n}{2^{n-2}2!(0,02601)^n}, & n \geq 2. \end{cases}$$

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

- Número médio de clientes na fila:

$$\hat{L}_Q = \frac{(r^2\rho)}{2!(1-\rho)^2} P_0 = 0,328;$$

- Tempo médio de cada unidade na Fila (minutos):

$$\hat{W}_Q = \frac{\hat{L}_Q}{\hat{\lambda}} = 12,66;$$

- Número médio de clientes no sistema:

$$\hat{L} = \lambda \hat{W} = \hat{L}_Q + r = 1,323;$$

- Tempo médio de cada unidade no sistema (minutos):

$$\hat{W} = \hat{W}_Q + \frac{1}{\hat{\mu}} = 51,106.$$

Os desvios padrões utilizados no cálculo dos intervalos de confiança são os mesmos do modelo anterior.

Tabela 7: Intervalos de Confiança para as estimativas médias 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

Esses resultados comprovam que o modelo M/M/2 otimizaria o atendimento e reduziria 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 há 20 minutos (com 95% de confiança), o que no modelo M/M/1 era superior a 5 dias.

A necessidade dos pacientes de um tratamento intensivo nos leitos oferecidos pelo sistema único de saúde seria atendido com sucesso e sem filas. O sistema de filas com 2 servidores aumentaria a efetividade dos serviços prestados, reduzindo o tempo médio de espera na fila e, consequentemente, evitando complicações na saúde dos pacientes que necessitassem de tratamento imediato. Do ponto de vista do servidor, a ausência de filas indica que em 33,5% do tempo os funcionários ficariam inativos, o que pode justificar a não inserção de um segundo guichê devido ao custo de um novo atendente, o seu tempo ocioso em serviço e os custos de uma nova equipe de profissionais da saúde para o monitoramento dos pacientes nos novos leitos.

6.4 Modelo M/M/3

Para confirmar que a adição de um segundo servidor é suficiente, ou seja, verificar se a aplicação de um terceiro servidor causaria melhorias significativas em relação ao modelo M/M/2 o modelo M/M/3 foi aplicado. Para o modelo com 3 servidores obtemos as seguintes estimativas para a probabilidade de atendimento imediato e a proporção do tempo em que existam n pacientes no sistema,

$$P_0 = \left(1 + 0,995 + 0,995^2/2 + \frac{0,995^3}{6(1-0.498)}\right) = 0,3654;$$

$$P_n = \begin{cases} \frac{(0,02589)^n}{n!(0,02601)^n} \times 0,3654, & 1 \leq n < 3, \\ \frac{(0,02589)^n}{3^{n-3}3!(0,02601)^n}, & n \geq 3. \end{cases}$$

O resultado de P_0 nos indica que em 36,54% do tempo o sistema fica vazio, e os atendententes dos guichês ficam ociosos um aumento de apenas 2,96% em relação ao modelo anterior. As medidas de desempenho para esse novo modelo são

- Número médio de clientes na fila:

$$\hat{L}_Q = \frac{(r^3\rho)}{3!(1-\rho)^2}P_0 = 0,0893;$$

- Tempo médio de cada unidade na Fila (minutos):

$$\hat{W}_Q = \frac{\hat{L}_Q}{\hat{\lambda}} = 3,448;$$

- Número médio de clientes no sistema:

$$\hat{L} = \lambda\hat{W} = \hat{L}_Q + r = 1,085;$$

- Tempo médio de cada unidade no sistema (minutos):

$$\hat{W} = \hat{W}_Q + \frac{1}{\hat{\mu}} = 41,895.$$

Tabela 8: Intervalos de Confiança para as estimativas médias 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

É perceptível que um modelo em que existam 3 servidores disponíveis não é mais aconselhável do que o modelo com apenas 2, uma vez que as estimativas pontuais e os intervalos de confiança foram bem próximos com diferenças de poucos minutos entre eles. O número médio de pacientes no sistema diminuiu de 1,32 para 1,08 e o máximo possível para essa média de acordo com o intervalo de 95% de confiança teve uma diferença de apenas 3 clientes a menos no sistema.

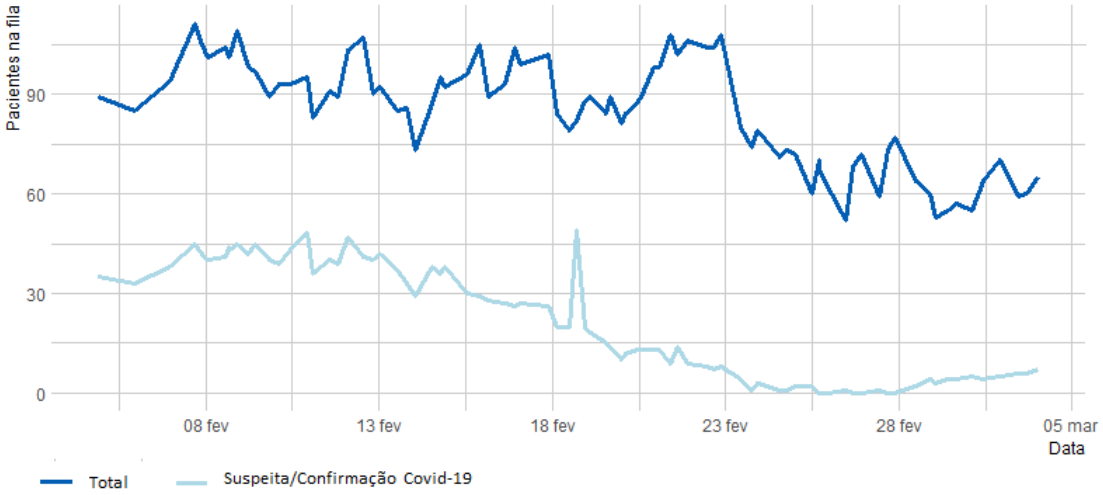
6.5 Pacientes com suspeita/confirmação de covid-19

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 necessidade de leitos.

A internação por Síndrome Respiratória Aguda Grave em leitos de UTI é feita quando existem indícios de que 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 por covid-19. Sabendo que existe uma população de risco, com maior predisposição a demanda de suporte ventilatório, se faz necessária a internação em UTI, visando o cuidado periódico destas vidas (F. CARVALHO; ELIAS; T. CARVALHO, 2021). No período observado cerca de 23,6% dos pacientes aguardando em fila possuiam suspeita ou confirmação do vírus SARS-CoV2.

Figura 9: Numéro de pacientes na fila total e com suspeita ou confirmação de covid-19.



Pela Figura 9 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 depacientes em espera por leitos de UTI, no período do dia 04 até 18 de fevereiro a proporção de pacientes com confirmação ou suspeita da doença era maior e a medida em que os casos foram diminuindo o tamanho da fila também diminuiu.

O número de pacientes com suspeita/confirmação de covid-19 que chegam por minuto e o intervalo médio entre suas chegadas, para os 27 dias e meio analisados, são respectivamente,

$$\hat{\lambda}_{covid-19} = \frac{258}{27,5} \times \frac{1}{24} \times \frac{1}{60} = 0,0065 \quad e \quad \bar{x}_{covid-19} = \frac{27,65}{257} = 0,108.$$

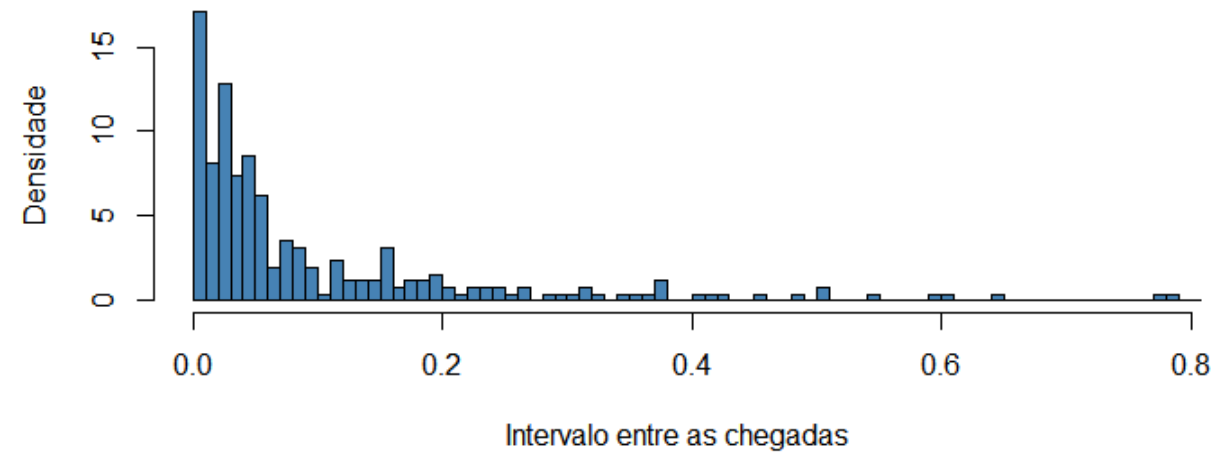
No cálculo da média a soma dos intervalos entre chegadas está em dias (27,65) e o último paciente registrado com suspeita/confirmação de covid-19 foi desconsiderado por não haver uma chegada após ele para calculo do intervalo.

Esses resultados indicam que chegam em média 0,0065 pacientes por minuto, ou seja, 9 pacientes por dia, ou seja, 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 haviam 34 pacientes na fila com horários de chegadas disponíveis, entretanto foram desconsiderados nos cálculos acima pois seus intervalos entre chegadas estavam imcompletos, uma vez que a maior parte dos pacientes

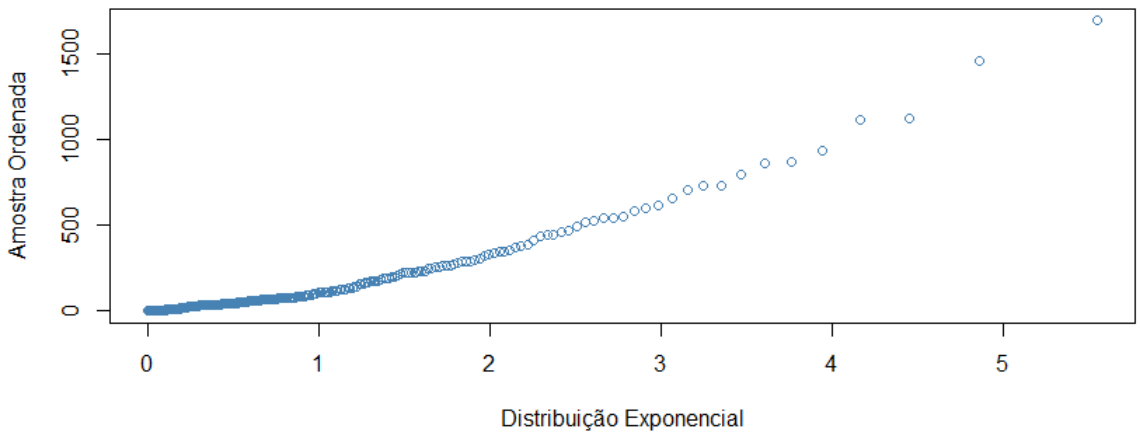
que entraram na fila em períodos próximos já haviam deixado a fila e seus dados estavam indísponíveis.

Figura 10: Histograma dos intervalos entre chegadas.



O teste Kolmogorov-Smirnov de Aderência foi realizado para verificarse o intervalo entre as chegadas possui distribuição Exponencial. Para um nível de significância de 99%, $(\alpha = 0,01)$, a estatística $D = max_x|S_n(x) - F_0(x)| = 0,183$ obtida é maior que o valor crítico tabelado do teste (0,102).

Figura 11: Gráfico Q-Q da distribuição dos intervalos entre chegadas.



Entretanto, ao analizar o gráfico Q-Q, como as observações se apresentam de forma linear podemos assumir que a distribuição dos intervalos é a mesmo da distribuição teórica utilizada, ou seja, iremos consentir que os intervalos entre as chegadas seguem

uma distribuição Exponencial.

O número médio de pacientes atendidos por minuto é de

$$\hat{\mu}_{covid-19} = \frac{292-7}{27,5} \times \frac{1}{24} \times \frac{1}{60} = 0,0072.$$

Para esse cálculo consideramos 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.

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 $= 0,14$ obtido a um nível de significância de 5% aceita-se que os dados possam ser representados pelo modelo Poisson com média 5,8 e iremos assumir que o tempo de atendimento possui uma distribuição Exponencial com taxa $\hat{\mu} = 0,0072$.

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, obtem-se as seguintes estimativas de desempenho:

Tabela 9: 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

Comparando as estimativas pontuais do modelo M/M/1 e M/M/2 há uma melhora significativa no tempo médio de espera 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 o que não é satisfatório o suficiente para se justificar a adição desse terceiro servidor, indicando a necessidade de apenas um segundo hospital de campanha.

Tabela 10: Intervalos de Confiança 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]

Ao se analisar os intervalos de confiança percebe-se que, o tempo de espera médio em fila que antes era superior à 21,3 horas e no modelo M/M/2 não supera 43,2 minutos e o tempo de espera médio no sistema que inicialmente era maior que 21,4 horas e no novo arquétipo com 2 servidores é menor que 5,3 horas, foram os que obtiveram as maiores diferenças entre os dois modelos.

Para confirmar que a adição de um segundo servidor para pacientes com suspeita/confirmação de COVID-19 é suficiente, ou seja, verificar se a aplicação de um terceiro servidor causaria melhorias significativas em relação ao modelo M/M/2 compara-se o desempenho de ambos os modelos.

O tempo médio de espera na fila que pelo intervalo de confiança era inferior a 44 minutos foi reduzido para menos de 17 minutos aproximadamento, o que não é tão significativo considerando a diferença de mais de 20 horas que um unico servidor a mais proporcionou.

É perceptível que um modelo em que existam 3 servidores disponíveis não é mais aconselhável do que o modelo com apenas 2, uma vez que as estimativas pontuais e os intervalos de confiança foram bem próximos com diferenças de poucos minutos entre eles. O número médio de pacientes no sistema diminuiu de 1,13 para 0,96 e o máximo possível para essa média de acordo com o intervalo de 95% de confiança teve uma diferença menor que clientes a menos no sistema.

7 Conclusão

O estudo da Teoria das Filas tem grande importância, uma vez que estão presentes em diversas situações do cotidiano. As Filas Markovianas estudadas neste trabalho possuem aplicações de modelos mais simples para a área da saúde, resultando em uma vasta área a ser explorada de modelos mais complexos.

A Teoria das Filas é uma ferramenta matemática que trata de eventos aleatórios e permite detalhar antecipadamente o comportamento de um sistema de filas encontrando soluções matemáticas que equilibrem tanto a aglomeração de clientes quanto a taxa de recepcionistas inativos, e através desse estudo propor sistemas que sejam eficazes economicamente para o provisor do atendimento e que atendam a procura do mercado.

Em aplicações, o estudo da Teoria das Filas tem como objetivo propor aperfeiçoamentos para a performance do sistema, ou seja, melhor aproveitamento dos recursos disponíveis considerando o ponto de vista do cliente e do servidor gerando assim um menor tempo de espera e um atendimento mais rápido e eficaz. As medidas de desempenho mais importantes de um sistema, como: número médio de clientes no sistema e na fila, comprimento médio da fila, tempo médio de espera no sistema e na fila e a taxa média de ocupação do serviço, auxiliam na tomada de decisão para redução das filas e do tempo de permanência dos clientes no sistema.

Ao analisar os resultados encontrados nas seções (6.2) e (6.4), observa-se que as taxas de ocupação médias ($U = \frac{\hat{\lambda}}{\hat{\mu}}$) são bastante elevadas (acima de 0,9) em ambos os casos. Como o sistema de filas é acumulativo ao longo do tempo, para que o tempo médio de espera por leito reduza para valores satisfatórios, o número médio de pacientes que chegam por minuto precisa ser suficientemente menor do que o número médio de pacientes atendidos por minuto, ou seja, a relação $\frac{\lambda}{\mu}$ precisa ser significantemente menor do que um.

Para tentar controlar o número médio de pacientes que esperam por leitos COVID também é importante a conscientização da população sobre a transmissão do vírus SARS-CoV-2 e orientações sobre distanciamento e uso de máscaras. Outros fatores que colaboram no aumento do tempo de internação é a falta de oxigênio e de medicamentos usados na entubação de pacientes graves de COVID-19, o que pode resultar na transferência de pacientes gerando grandes despesas para os hospitais e transtornos para os pacientes.

No que se estabelece, durante o desenvolvimento desse trabalho foi realizada a análise dos dados, identificando os fatores que limitam o desempenho do atendimento dos

pacientes e através da utilização da técnica, Teoria das Filas, foi indicada a implementação de novos servidores para otimizar o fluxo do processo de internação em leitos de UTI gerais e para pacientes com suspeita ou confirmação de COVID-19. Em ambos os casos abordados o emprego de um novo servidor gerou melhorias no desempenho do sistema e a adição de um terceiro apresentou melhoria pouco significante em relação ao modelo com 2 servidores.

A importância das Unidades de Terapia Intensiva é vista no tratamento e monitoramento de pacientes em estado grave de saúde com a finalidade da recuperação do vigor desses pacientes. No entanto, um elevado tempo de internação pode gerar complicações para a saúde dos internados devido a possíveis infecções, gastos adicionais para o hospital que poderiam ser evitados e o desacolhimento de pacientes em estado crítico em virtude de uma menor quantidade de leitos disponíveis. Dessa forma, para que esse serviço seja de fácil acesso, visto que no Sistema ùnico de Saúde (SUS) houve no máximo 206 leitos COVID no período analisado para todo o Distrito Federal, é necessário que se faça o acompanhamento do indicador de média de permanência de UTI, visando intervenções para que o aumento e redução do número de leitos seja executado da maneira mais segura e eficiente possível. A avaliação do quadro do paciente e diagnóstico da doença também devem ser otimizados para uma diminuição do tempo até a disponibilidade de uma vaga para transferência.

A análise das priorizações para a mobilização de novos leitos leva em consideração a oferta dos leitos de UTI no período, como também as repercussões assistencial e financeira. A mobilização de leitos de UTI em unidades subordinadas tem sua análise baseada nos parâmetros de performance assistencial da rede privada contratada.

A média de permanência na UTI pode ser empregada para apontar a contratação de novos leitos além de avaliar a efetividade dos serviços prestados. O tempo médio de permanência na fila dos leitos de UTI no DF no período analisado é de no mínimo 5 dias e para pacientes com necessidade de leitos para COVID-19 esse número é de 21,3 horas, podendo ser influenciado por fatores como a prioridade do paciente. Portanto, empregar medidas como a criação de protocolos mais eficientes para a admissão de novos pacientes ajudam a priorizar indivíduos em casos mais complexos, garantindo maiores chances de que existam leitos disponíveis para o atendimento da população.

É importante que se comente sobre o quanto o estudo da Teoria de Filas é complexo e sobre a dificuldade encontrada na coleta de dados. A Teoria das Filas possui grande importância na area da saúde, pois o estresse gerado por uma longa espera principalmente para pacientes em estados graves pode acarretar em uma deterioração da sua

saúde tanto física quanto mental e nos piores cenários causar a morte de pacientes na fila à espera de um leito de UTI.

8 Referências

A. K. ERLANG, The theory of probabilities and telephone conversations. Nyt Tidsskrift for Matematik, 1909.

I. I. PERDONÁ, R. V. NUNES, R. M. NEVES, S. C. NAIMER, e L. P. GODOY, "Sistema de manufatura: otimização de processos em uma unidade fabril de cimento através da teoria das filas", Exacta, 2017.

G. R. CAMELO, A. S. COELHO, R. M. BORGES, e R. M. SOUZA, "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.

V. MORAES, R. VENTURINI, M. TAVARES, M. de SÁ, e W. SALMER, "Boletim epidemiológico: Análise da média de permanência UTI adulto município de jataí", Observatório de Epdemiologia e Serviços de Saúde - EpiServ, 2018.

S. M. ROSS, Introduction to Probability Models. Elsevier, 2014.

E. F. B. JÚNIOR, "Estudo do sistema de filas de transplante de Órgãos no SUS", 2017. Monografia (Bacharel em Estatística), UNB (Universidade de Brasília), Brasília, Brasil.

J. F. de ARAÚJO, "Estudo da teoria de filas com aplicações", 2015. Monografia (Bacharel em Estatística), UNB (Universidade de Brasília), Brasília, Brasil.

G. BRESSAN, "Modelagem e simulação de sistemas computacionais - sistemas de filas simples", 2002. LARC-PCS/EPUSP.

K. SAMEJIMA, "Estatística não paramétrica", 2021. Disponível em: ⟨https://est.ufba.br/sites/est.ufba.br/files/kim/matd49-aula05-aderencia.pdf⟩.

W. J. CONOVER, Practical Nonparametric Statistics. New York: John Wiley, 1998.

D. G. KENDALL, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain", The Annals of Mathematical Statistics, 1953.

M. SCHLINZ, "O que é Unidade de Terapia Intensiva?", 2016. Disponível em: ⟨https://www.iespe.com.br/blog/o-que-e-unidade-de-terapia-intensiva/#: :text=A%20Unidade %

20de%20Terapia%20Intensiva,Florence%20Nightingale⟩.

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.

O. F. TORRES, "Elementos da teoria das Filas", Revista de Administração de Empresas, 1966.

G. C. F. CARVALHO, L. M. ELIAS, e R. T. CARVALHO, "Você sabe o que é uma UTI e sua importância na COVID-19?", 2021. Disponível em: ⟨https://coronavirus.saude.mg.gov .br/blog/61-o-que-e-uma-uti#: :text=As%20Unidades%20de%20Terapia%20Intensiva%20t%C3 %AAm%20sido%20uma%20ferramenta%20de,visando%20a%20manuten%C3%A7%C3 %A3o% 20destas%20vidas⟩.

L. R. AMIDANI, "A teoria das filas aplicada aos serviços bancários", Revista de Administração de Empresas, 2013.

Anexo

9 Código da simulação

Lista de nomenclaturas utilizadas no código:

1.T: período total de simulação;

2.tci: instante de tempo em que ocorre a i-ésima chegada;

3.tai: instante de tempo em que ocorre o i-ésimo atendimento;

4.to: tempo total de ociosidade do servidor;

5.te: tempo total de espera na fila;

6.tme: tempo médio de espera na fila;

7.i: contador;

8.t: variável de controle (tempo).

Script:

```
sistema<- function(T,l,u)
{ #cria o vetor dos tempos de chegada dos clientes (tc)
  #a partir de uma exponencial de média 1/l
  #referente aos tempos entre as chegadas (tec).
  t=0
  i=1
  tc<- as.vector(rep(0,(1.3*T*l))) #criando vetor de zeros de tamanho 30% maior
  que o número médio de clientes = T*l.
  while (t<T){
    tec<- rexp(1,l)
    tc[i]<- t+tec
    t<- tc[i]
    i=i+1 }
  #arrumando o vetor tc.
  tc<- subset(tc,tc!=0) #retirando os zeros excedentes do vetor.
  n<- length(tc)-1
```

```
tc<- tc[1:(n)]
#criando o vetor de tempo de atendimento, ta, para os n clientes a partir de
uma exponencial de média 1/u
#e acumulando o tempo ocioso ou o tempo de espera na fila do sistema.
i=1
to=tc[1]
te=0
t=tc[1]
ta<- as.vector(rep(0,n))
while (i<n) {
  ta[i]<- rexp(1,u)
  if (t+ta[i]<tc[(i+1)])
    { to<- to + tc[i+1]-(t+ta[i])
    t<- tc[i+1] }
  else
    { te<- te + (t+ta[i])-tc[i+1]
    t<- t+ta[i] }
    i=i+1}
#fazendo o caso particular de i=n (pois não existe n+1).
ta[i]<- rexp(1,u)
if(t+ta[i]<T)
  { to<- to+(T-t-ta[i]) }

#cálculo teórico e empírico do tempo de espera médio do sistema
#e da taxa de ocupação do sistema.
te_medio<- te/n
te_medio_teorico<- l/(u*(u-l))
tx_ocup<- 1-(to/T)
tx_ocup_teorico<- l/u
#print(tc)
#print(ta)
#criando um vetor que indica em que tempo o cliente saiu do sistema,
#levando em consideração duas possibilidades:
#1) o cliente chega, é atendido e sai.
#2) o cliente chega, espera na fila, é atendido e sai.
ts<- as.vector(rep(0,n))
```

```
ts[1]<- tc[1]+ta[1]
for(i in 2:n){
  if(tc[i]<ts[i-1])
    { ts[i]<- ts[i-1]+ta[i] }
  else
    { ts[i]<- tc[i]+ta[i] } }
#print(ts)
#criando um data.frame com as variáveis:
#"CLIENTE": indicando o número do cliente no sistema;
#"SITUAÇÃO": indicando se o cliente está chegando ou saindo do sistema;
#"CÓDIGO_SITUAÇÃO": igual à 1 quando o cliente entra no sistema e -1 quando
sai;
#"TEMPO": indicando em que momento do tempo está a observação.

x<-as.data.frame(matrix(c(rep(1:n,2),
                          rep("chegada",n),
                          rep("saida",n),
                          rep(1,n),
                          rep(-1,n),
                          tc,
                          ts),nrow=(2*n),ncol=4,dimnames=list(c(1:(2*n)),
                          c("CLIENTE",
                          "SITUACAO",
                          "CODIGO_SITUACAO",
                          "TEMPO"))))
x$CODIGO_SITUACAO=as.numeric(as.character(x$CODIGO_SITUACAO))
#transformando o CÓDIGO_SITUAÇÃO

x$TEMPO=as.numeric(as.character(x$TEMPO))
#e o TEMPO em numérico para manipulação.

x<-x[order(x$TEMPO),] #ordenando os dados em relação ao TEMPO.
#calcular o tamanho da fila como a soma dos códigos até o momento.

for(k in 1:(2*n))
  { x$FILA[k]<- sum(x$CODIGO_SITUACAO[1:k])-1 } #(-1) que representa
```

```
    ociosidade no sistema.
  #calcular o tempo em que o sistema ficou em cada tamanho de fila.
  for(k in 2:(2*n))
    { x$PROPORCAO[1]<- x$TEMPO[2]
  x$PROPORCAO[k]<- x$TEMPO[k+1]-x$TEMPO[k] }
  #print(tail(x))

  #calculando o tamanho médio da fila ponderado pelo tempo:
  y<-subset(x,x$FILA>0) #separando os números positivos para o tamanho da fila,
  pois quando 0, a multiplicação é nula.

  y<-y[order(y$TEMPO),]
  teste<- sum(x$PROPORCAO[1:(length(x$PROPORCAO)-1)])
#print(teste)
  tamanho_medio_fila<- (sum(y$FILA * y$PROPORCAO))/teste
  tamanho_medio_da_fila_teorico<- ((l/u)^2)/(1-(l/u))
#imprimir os resultados!
  final<-list("Tempo médio de espera na fila"= te_medio,
              "Tempo médio de espera na fila (teorico)"= te_medio_teorico,
              "Taxa de ocupação"= tx_ocup,
              "Taxa de ocupação (teorico)"= tx_ocup_teorico,
              "Tamanho médio da fila ponderado"= tamanho_medio_fila,
              "Tamanho médio da fila (teórico)"= tamanho_medio_da_fila_teorico)
  return(final)}
```

```
# quanto maior o número médio de clientes (T*l) mais próximas serão as medidas
empíricas das teóricas.
(n<-(qnorm(.95)/(4*0.01))^2) #Definindo o tamanho de n
te_medio<-numeric(n)
tx_ocup<-numeric(n)
tamanho_medio_fila<-numeric(n)
```

```
for (i in 1:n){ #Gerando as simulações
  sim<-sistema(400,2,4)
  te_medio[i]<-sim[[1]]
  tx_ocup[i]<-sim[[3]]
  tamanho_medio_fila[i]<-sim[[5]]}
#Cálculo das médias
(te_medio_medio<- mean(te_medio))
(tx_ocup_medio<-mean(tx_ocup))
(tamanho_medio_medio<-mean(tamanho_medio_fila))
#Cálculo dos desvios padrões
(te_medio_sd<- sd(te_medio))
(tx_ocup_sd<- sd(tx_ocup))
(tamanho_medio_sd<- sd(tamanho_medio_fila))
#Intervalos de confiança
(ICinf1<- te_medio_medio - (qnorm(.975))*te_medio_sd)
(ICsup1<- te_medio_medio + (qnorm(.975))*te_medio_sd)
(ICinf2<- tx_ocup_medio - (qnorm(.975))*tx_ocup_sd)
(ICsup2<- tx_ocup_medio + (qnorm(.975))*tx_ocup_sd)
(ICinf3<- tamanho_medio_medio - (qnorm(.975))*tamanho_medio_sd)
(ICsup3<- tamanho_medio_medio + (qnorm(.975))*tamanho_medio_sd)
```