

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

WILLIAM MACHADO PANTALEÃO

**ANÁLISE DE ESCALA DO R_0 EM FUNÇÃO DO TAMANHO DA
POPULAÇÃO COM DADOS DA *COVID-19***

PORTO ALEGRE
2022

WILLIAM MACHADO PANTALEÃO

**ANÁLISE DE ESCALA DO R_0 EM FUNÇÃO DO TAMANHO DA
POPULAÇÃO COM DADOS DA *COVID-19***

Trabalho de Conclusão de Curso apresentado ao Instituto de Física da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Física.

Orientador: SEBASTIÁN GONÇALVES

PORTO ALEGRE
2022

RESUMO

A grande quantidade de dados disponíveis de uma mesma doença em diferentes locais e no mesmo período de tempo, devido à pandemia de *COVID-19*, contribui no estudo de parâmetros como o número básico de reprodução, R_0 . Este trabalho propõe-se, então, a analisar o R_0 para diferentes áreas urbanas funcionais (*functional urban areas - FUAs*), buscando um melhor entendimento sobre o comportamento da dinâmica de doenças infectocontagiosas. Para isso, foram obtidos dados a partir de séries temporais de casos acumulados, estimativas populacionais e tabelas com códigos para *FUAs* das cidades de diferentes países. Em seguida foram implementadas diferentes formas de calcular o número de reprodução para cada população, baseadas nos modelos SIR e SEIR e em métodos estatísticos comumente utilizados para obtenção do número de reprodução efetivo, R_t , durante epidemias. Com os ajustes dos valores obtidos, pelo método dos mínimos quadrados ordinários (*ordinary least squares - OLS*), os coeficientes de correlação de Pearson, ρ , confirmaram a existência de uma lei de escala entre o número básico de reprodução e o tamanho da população para unidades urbanas, independente do método empregado para análise.

Palavras-chave: Número Básico de Reprodução. Escalonamento Urbano. *COVID-19*.

ABSTRACT

The large amount of data available on the same disease in different locations and in the same period of time, due to the COVID-19 pandemic, contributes to the study of parameters such as the basic reproduction number, R_0 . This work proposes, therefore, to analyze the R_0 for different functional urban areas - FUAs, seeking a better understanding of the behavior of the dynamics of infectious diseases. For this, data were obtained from time series of accumulated cases, population estimates and tables with codes for FUAs of cities in different countries. Then, different ways of calculating the reproduction number for each population were implemented, based on the SIR and SEIR models and on statistical methods commonly used to obtain the effective reproduction number, R_t , during epidemics. With the adjustments of the values obtained, by the method of ordinary least squares - OLS, the Pearson correlation coefficients, ρ , confirmed the existence of a scaling law between the basic reproduction number and the population size for urban units, regardless of the method used for analysis.

Keywords: Basic Reproduction Number. Urban Scaling. COVID-19.

SUMÁRIO

1	INTRODUÇÃO	1
2	FUNDAMENTAÇÃO TEÓRICA	3
2.1	Modelos SIR e SEIR	3
2.1.1	Número Básico de Reprodução - R_0	5
2.2	Número Efetivo de Reprodução - R_t	6
2.2.1	Modelos Estatísticos	7
2.3	Método dos Mínimos Quadrados - <i>OLS</i>	7
2.3.1	Coeficiente de Correlação de Pearson - ρ	8
2.4	<i>Urban Scaling</i>	9
2.4.1	Áreas Urbanas Funcionais - <i>FUAs</i>	9
3	MATERIAIS E MÉTODOS	11
3.1	Análise dos Dados: Casos Acumulados por <i>FUAs</i>	11
3.2	Modelos SIR e SEIR: Taxas de Crescimento Exponencial	13
3.3	Método de Cori: Valores Médios de R_t	14
3.3.1	Aplicação em Dados Sintéticos	14
3.3.2	Estimativas para <i>FUAs</i>	20
4	RESULTADOS E DISCUSSÕES	22
5	CONCLUSÕES	27
	Referências	28
	Apêndices	31
	APÊNDICE A Teoria da Escala de Contatos	32

1 INTRODUÇÃO

Na data de 31 de dezembro de 2019, a Organização Mundial da Saúde (OMS) foi alertada sobre vários casos de pneumonia na cidade de Wuhan, província de Hubei, na República Popular da China. Tratava-se de um novo tipo de coronavírus, que não havia sido identificado antes em seres humanos. Esse novo coronavírus recebeu o nome de *SARS-CoV-2* e foi responsável por causar a doença *COVID-19* (PAHO, 2020).

Um mês depois, a OMS declarou que o surto do novo coronavírus constituía uma Emergência de Saúde Pública de Importância Internacional (ESPII), o mais alto nível de alerta da Organização. Essa decisão buscou aprimorar a coordenação, a cooperação e a solidariedade global para interromper a propagação do vírus. Contudo, em 11 de março de 2020, ao reconhecer a existência de surtos em vários países e regiões do mundo, a OMS caracterizou a *COVID-19* como uma pandemia ¹.

A necessidade de um maior entendimento acerca de questões fundamentais referentes a essa nova doença e a possibilidade de observar como uma epidemia se comporta em diferentes locais num mesmo período de tempo, contribuíram para que fossem analisadas e provadas questões fundamentais da teoria de epidemias, como a natureza da sua dinâmica.

Uma forma de analisar essa dinâmica é mensurando a velocidade de propagação de uma doença infectocontagiosa, a partir da estimativa do seu número básico de reprodução. Este número, denotado por R_0 , é definido como o número médio de casos secundários gerados a partir de um indivíduo infectado em uma população homogênea. O R_0 representa um indicativo de propagação de uma doença emergente, à medida em que, para que uma epidemia ocorra, ele deve assumir valores maiores que 1. Ou seja, para os casos em que $R_0 \leq 1$, a doença não se espalha.

A bibliografia sobre modelagem de epidemias é hoje extensa, continua crescendo rapidamente e não poderia ser abordada de forma completa nesta monografia (MURRAY, 2002). A primeira formulação matemática registrada é de Bernoulli (BERNOULLI; PETROPOLITANAE, 1760), quem em 1760 usou equações diferenciais ordinárias não lineares para modelar o efeito da vacinação na propagação da varíola. Mas é a formulação compartimental de Kermack e McKendrick (KERMACK; MCKENDRICK, 1927) a que se usa de forma extensa atualmente. Nesta formulação, a população é considerada em um dos três estados possíveis: suscetível (S), infectado (I) ou removido (R), pelo qual o modelo básico é chamado de SIR. Um conjunto de equações diferenciais ordinárias, que será apresentado na próxima seção, descreve a dinâmica dos compartimentos, que no modelo SIR, depende de dois parâmetros: a taxa de transmissão, β , e a taxa de recuperação, γ . No modelo SIR,

¹Uma pandemia é uma epidemia de uma doença infecciosa que se espalhou por uma grande região, como vários continentes ou em todo o mundo, afetando um número substancial de indivíduos.

o R_0 tem uma relação teórica precisa com esses dois parâmetros: $R_0 = \beta/\gamma$.

Em tese, o R_0 independe do tamanho ou de outra característica da população, e a propagação de uma doença seguirá um comportamento denominado transmissão dependente de frequência ou ação de massa (KEELING, 2008). Contudo, foi possível observar que o R_0 apresenta um comportamento de escala com o tamanho da população para áreas urbanas (CARDOSO; GONÇALVES, 2020), semelhante ao observado em quantidades socioeconômicas. Isso foi possível devido à grande quantidade disponível de dados da *COVID-19*, o que permitiu que se pudesse analisar a disseminação de uma mesma doença dentro de regiões que englobem cidades onde seus habitantes encontram-se relacionados. Essa relação pode ser devida a um mercado de trabalho em comum, por exemplo.

Para confirmar este comportamento de escala, pode-se fazer uma análise do número efetivo de reprodução, R_t , ao longo do tempo para cada região. Diferentemente do R_0 , o R_t pode ser calculado a qualquer momento de uma epidemia a partir de estimativas estatísticas. Um método comumente utilizado para fazer essas estimativas é o de Cori (CORI et al., 2013b), que possibilita avaliar a transmissão de uma doença em pontos específicos ou quase em tempo real.

A dependência dos valores de R_0 com o tamanho da população, observada tanto para estimativas feitas a partir de modelos compartimentais quanto para métodos estatísticos, pode ser explicada por uma teoria de escala de contatos. Segundo essa teoria, existe uma variação da taxa de contato entre indivíduos quando analisados diferentes tamanhos populacionais (CARDOSO; GONÇALVES, 2021) disponível no Apêndice A.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Modelos SIR e SEIR

Como mencionado na introdução, no modelo SIR para epidemias (KERMACK; MCKENDRICK, 1927), a população de tamanho N é compartimentada em três classes

$$N = S + I + R, \quad (2.1)$$

onde S se refere à parte suscetível, I à parte infectada e R à parte removida da população, seja por recuperação (sem a possibilidade de nova contaminação) ou morte.

Normalizando a equação (2.1) tem-se que

$$s + i + r = 1, \quad (2.2)$$

onde $s = S/N$, $i = I/N$ e $r = R/N$.

As interações entre essas diferentes classes podem ser descritas por um conjunto de equações diferenciais

$$\begin{aligned} \frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \\ \frac{dr}{dt} &= \gamma i, \end{aligned} \quad (2.3)$$

onde β é a taxa de infecção e γ é a taxa de recuperação. Essas duas taxas dependem da interação entre patógeno (vírus, bactéria, etc.) e hospedeiro, enquanto a primeira depende também da interação social entre hospedeiros.

As equações em (2.3) indicam que o número de indivíduos suscetíveis diminui e o de infectados aumenta, quando dois deles se encontram, a uma taxa β , enquanto que o número de infectados e removidos diminui e aumenta, respectivamente, de forma espontânea.

O modelo SEIR se trata de uma variante do modelo SIR onde é introduzida a classe E referente aos indivíduos que foram expostos à doença e se encontram em um período latente, ou seja, que foram infectados mas ainda não são infecciosos.

Logo, na forma normalizada, onde $e = E/N$, a relação de dependência entre as classes se dá na forma

$$s + e + i + r = 1, \quad (2.4)$$

e o conjunto de equações diferenciais para o modelo SEIR será, então, o conjunto do

modelo SIR acrescido da variação do número de expostos, ou seja

$$\begin{aligned}\frac{ds}{dt} &= -\beta si \\ \frac{de}{dt} &= \beta si - \sigma e \\ \frac{di}{dt} &= \sigma e - \gamma i \\ \frac{dr}{dt} &= \gamma i,\end{aligned}\tag{2.5}$$

onde σ é a taxa de incubação.

A propagação de uma doença pode ser entendida como uma transição de fase termodinâmica (TOMÉ; OLIVEIRA, 2020), com um parâmetro de ordem definido como a área sob a curva epidêmica

$$A = \int_0^{\infty} f dt,\tag{2.6}$$

onde $f = f(t)$ é a frequência de novos casos, dada por

$$f = \beta si.\tag{2.7}$$

Considerando que $\beta si = -ds/dt$ e que em $t = 0$, $s = 1$, a solução da equação (2.6) será

$$A = \int_0^{\infty} f dt = - \int_1^{s^*} ds = 1 - s^*\tag{2.8}$$

onde s^* é o valor de indivíduos suscetíveis quando $t \rightarrow \infty$.

Para longos tempos, a fração de indivíduos expostos e infectados irá desaparecer e a soma dos valores finais s^* e r^* , referente a $r(t \rightarrow \infty)$, será 1.

Dividindo as equações diferenciais referentes às classes de suscetíveis e removidos nos modelos SIR (2.3) e SEIR (2.5), tem-se que

$$r = -\frac{\gamma}{\beta} \ln s.\tag{2.9}$$

Substituindo (2.9) em (2.8) e fazendo $s^* = 1 - r^*$, vê-se que

$$A = -\frac{\gamma}{\beta} \ln(1 - A),\tag{2.10}$$

independentemente do modelo epidemiológico considerado ser o SIR ou o SEIR. Isso quer dizer que a presença de indivíduos expostos não irá interferir no tamanho da epidemia, apenas na velocidade de sua propagação, conforme demonstrado na figura 1.

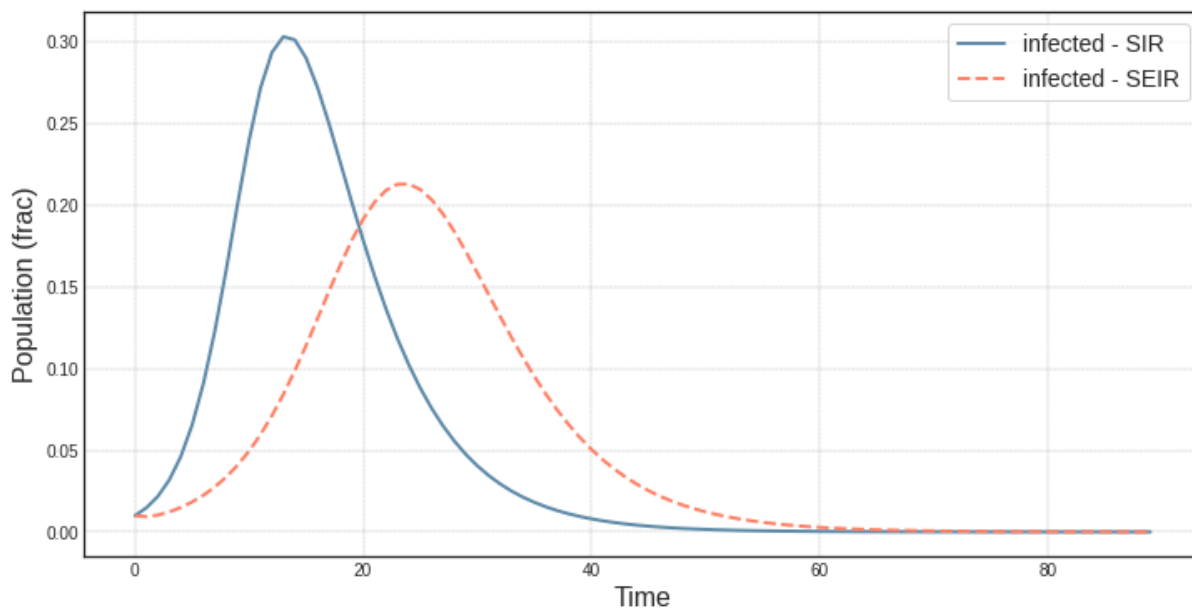


Figura 1 – Porção de indivíduos infectados em função do tempo para os modelos SIR e SEIR, considerando os mesmos parâmetros $\beta = 0.6$ e $\gamma = 0.2$ em ambos os casos e $\sigma = 0.5$ para o modelo SEIR.

2.1.1 Número Básico de Reprodução - R_0

O número básico de reprodução, denotado por R_0 , e definido como

$$R_0 = \frac{\beta}{\gamma} \quad (2.11)$$

é o número médio de casos secundários gerados a partir de um indivíduo infectado em uma população homogênea. Seu valor é calculado no início de uma epidemia, quando a quantidade de indivíduos suscetíveis à doença é aproximadamente o total da população (DIETZ, 1993).

O R_0 representa um indicativo de propagação de uma doença emergente, à medida em que, para que uma epidemia ocorra ele deve assumir valores maiores que 1. Para os casos em que $R_0 \leq 1$, a doença não se espalha.

Em sua fase inicial, uma epidemia normalmente irá apresentar um crescimento exponencial (MA, 2020). Considerando o modelo SIR, a taxa associada a este crescimento pode ser encontrada linearizando as equações (2.3) na situação em que $s = 1$, $i = 0$ e $r = 0$, ou seja,

$$\frac{di}{dt} = (\beta - \gamma)i, \quad (2.12)$$

onde

$$\beta - \gamma = \lambda_{SIR} \quad (2.13)$$

é a taxa de crescimento exponencial do modelo SIR.

Substituindo essa relação em (2.11), tem-se que

$$R_0 = \frac{\lambda_{SIR}}{\gamma} + 1. \quad (2.14)$$

Já no modelo SEIR, a taxa λ_{SEIR} pode ser encontrada quando $s = 1$, $e = 0$, $i = 0$ e $r = 0$ a partir das equações para e e i em (2.5), que se tornam

$$\frac{de}{dt} = -\sigma e + \beta i \quad (2.15)$$

e

$$\frac{di}{dt} = \sigma e - \gamma i. \quad (2.16)$$

A matriz Jacobiana referente a essas equações será

$$J = \begin{bmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{bmatrix} \quad (2.17)$$

que terá como maior autovalor real

$$\lambda_{SEIR} = \frac{1}{2} \left[-(\sigma + \gamma) + \sqrt{(\sigma + \gamma)^2 + 4\sigma(\beta - \gamma)} \right], \quad (2.18)$$

onde λ_{SEIR} é a taxa de crescimento exponencial do modelo SEIR. Percebe-se em (2.18) que para $\lambda_{SEIR} > 0$, ou seja para que se tenha um crescimento exponencial inicial, β deve ser maior que γ , isto é, deve ser satisfeita a condição $R_0 = \beta/\gamma > 1$. A mesma condição deve ser satisfeita para o modelo SIR em (2.11).

Outrossim, a partir da (2.18) se pode expressar R_0 em termos de λ_{SEIR} e das duas taxas, γ e σ , do modelo SEIR:

$$R_0 = \frac{(\lambda_{SEIR} + \sigma)(\lambda_{SEIR} + \gamma)}{\sigma\gamma}. \quad (2.19)$$

2.2 Número Efetivo de Reprodução - R_t

A transmissibilidade de uma doença pode ser quantificada com uma análise estatística do número de reprodução efetivo R_t . O R_t pode ser definido como o número esperado de novas infecções causadas por um indivíduo infeccioso em uma população que não é mais totalmente suscetível. Ou seja, diferentemente do R_0 , o R_t pode ser calculado a qualquer momento de uma epidemia.

As estimativas de R_t são comumente utilizadas para avaliar como mudanças nas políticas, imunidade da população e outros fatores afetaram a transmissão da doença em momentos específicos ou quase em tempo real. (GOSTIC et al., 2020)

2.2.1 Modelos Estatísticos

Cori et al. (2013b) assumem que um indivíduo infectado possui um perfil de infectividade dado por uma distribuição de probabilidade ω_s , denominada intervalo de geração ¹, que depende do tempo s desde a infecção.

A estimativa de R_t pode ser obtida pela razão

$$R_t = \frac{I_t}{\sum_{s=1}^t I_{t-s}\omega_s}, \quad (2.20)$$

onde I_t é o número de novas infecções no passo de tempo t e o somatório no denominador é a infecciosidade total de indivíduos infectados no tempo t .

O R_t em (2.20) pode ser interpretado como o número médio de casos secundários que cada indivíduo infectado infectaria se as condições permanecessem como estavam no tempo t (CORI et al., 2013b).

O método de Wallinga e Teunis (2004) estima a probabilidade de que o caso j , que foi infectado no tempo t_j , infecte o caso i em relação à probabilidade de que outro caso k infecte o caso i , ou seja

$$p_{ij} = \frac{\omega(t_i - t_j)}{\sum_{i \neq k} \omega(t_i - t_k)}, \quad (2.21)$$

onde $\omega(\tau)$ é a distribuição para o intervalo de geração, específico para cada doença infecciosa.

O R_t para o caso j será o somatório de (2.21) para todos os casos i :

$$R_j = \sum_i p_{ij}. \quad (2.22)$$

Já Bettencourt e Ribeiro (2008) derivam uma relação aproximada entre o R_t e a taxa de crescimento exponencial da epidemia:

$$I_{t+1} = I_t e^{\frac{R_t - 1}{g}}, \quad (2.23)$$

onde g é o tempo de geração médio.

A equação (2.23) é derivada do modelo SIR e assume, portanto, uma distribuição exponencial para o intervalo de geração.

Os métodos de Wallinga e Teunis (2004) e Cori et al. (2013b) usam abordagens que alavancam informações sobre o intervalo serial de uma doença.

O método (2.20) estima com maior precisão o R_t em tempo real, usando apenas dados passados e suposições paramétricas mínimas (GOSTIC et al., 2020).

2.3 Método dos Mínimos Quadrados - OLS

O método dos mínimos quadrados, ou *OLS* na sigla em inglês, consiste em encontrar os parâmetros de uma função modelo que façam com que essa função se ajuste da melhor forma a um conjunto de dados.

¹Intervalo de geração é o tempo entre as incidências de infecção em um par infector-infectado.

Seja um conjunto de dados simples (x_i, y_i) , onde x_i é uma variável independente e y_i é uma variável dependente cujo valor é encontrado por observação, a função modelo será

$$f(x, \beta), \quad (2.24)$$

onde β é um vetor que contém m parâmetros ajustáveis.

O ajuste do modelo será feito então a partir do seu resíduo

$$r_i = y_i - f(x_i, \beta), \quad (2.25)$$

e o método dos mínimos quadrados encontrará os valores ideais para os parâmetros minimizando a soma dos quadrados destes resíduos.

Ou seja, para o caso de $y_i = a + bx_i + e_i$, o método dos mínimos quadrados procura minimizar $\sum_{i=1}^n e_i^2$.

Substituindo e_i por $y_i - a - bx_i$, tem-se que

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (2.26)$$

cuja minimização pode ser obtida igualando as derivadas em relação a a e b a zero

$$\begin{aligned} \frac{\partial S}{\partial a} &= -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial S}{\partial b} &= -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0. \end{aligned} \quad (2.27)$$

Distribuindo e dividindo a primeira expressão em (2.27) tem-se que a estimativa para o parâmetro constante será dada por

$$a = \bar{y} - b\bar{x}, \quad (2.28)$$

onde \bar{y} e \bar{x} são as médias amostrais de y e x , respectivamente.

Substituindo (2.28) na segunda expressão de (2.27), encontra-se então a estimativa para o coeficiente de x

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.29)$$

2.3.1 Coeficiente de Correlação de Pearson - ρ

O coeficiente de correlação de Pearson, denotado por ρ é uma ferramenta usada para quantificar a relação entre duas variáveis.

Seus valores situam-se entre -1 e 1 indicando, nesses limites, forte correlação negativa e positiva, respectivamente, e inexistência de correlação quando $\rho = 0$.

Formalmente, ρ irá medir a covariância de duas variáveis X e Y dividida pelo produto de seus desvios padrão, ou seja,

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (2.30)$$

onde $\text{cov}(X,Y)$ é a covariância e σ_X e σ_Y são os desvios padrão.

2.4 Urban Scaling

As mudanças dos atributos de cidades com os seus tamanhos pode ser explicada por uma função de lei de potência do tipo

$$Y(N) = Y_0 N^\beta \quad (2.31)$$

onde Y representa alguma quantidade socioeconômica total da cidade e Y_0 e β são constantes para o tamanho da população, denotado por N (KEUSCHNIGG; MUTGAN; HEDSTRÖM, 2019).

Essa lei de escala pode ser encontrada, por exemplo, para o PIB (STRANO; SOOD, 2016), mobilidade (NOULAS et al., 2012), taxas de criminalidade (BETTENCOURT et al., 2010), entre outros.

A relação (2.31) decorre diretamente da rede de interações humanas, que estão subjacentes a fenômenos sociais como disseminação de doenças (SCHLÄPFER et al., 2014).

Para entender melhor como o R_0 se comporta em relação ao tamanho da população e, conseqüentemente, às diferentes redes de interações humanas, devem ser consideradas regiões que se enquadrem na definição de área urbana funcional, pois ela fornece a escala correta para que essa análise possa ser feita.

2.4.1 Áreas Urbanas Funcionais - FUAs

Segundo Milego et al. (2019) as áreas urbanas funcionais (*Functional Urban Areas* - FUAs), referem-se a núcleos urbanos centrais inter-relacionados a outros municípios por fluxos de pessoas, bens e serviços.

Essas áreas são uma ferramenta para comparar tendências socioeconômicas e espaciais nas cidades e para projetar políticas de desenvolvimento urbano, sendo mais adequadas do que as áreas administrativas, como municípios, para capturar economias de aglomeração² e abrangem toda a extensão do mercado de trabalho da cidade (DIJKSTRA; POELMAN; VENERI, 2019).

As FUAs podem ser usadas também para orientar governos nacionais e municipais no planejamento de infraestrutura, transporte, habitação, escolas e espaços de cultura e

²A economia de aglomeração pode ser entendida como um agrupamento de várias indústrias em uma única região geográfica, provocando uma aglomeração da produção. Essas indústrias não necessariamente precisam pertencer ao mesmo setor econômico.

lazer, fornecendo a escala certa para abordar questões que afetam tanto a cidade quanto a zona de deslocamento ao redor. A figura 2 ilustra como uma área urbana funcional engloba uma cidade e a zona de deslocamento ao seu redor.

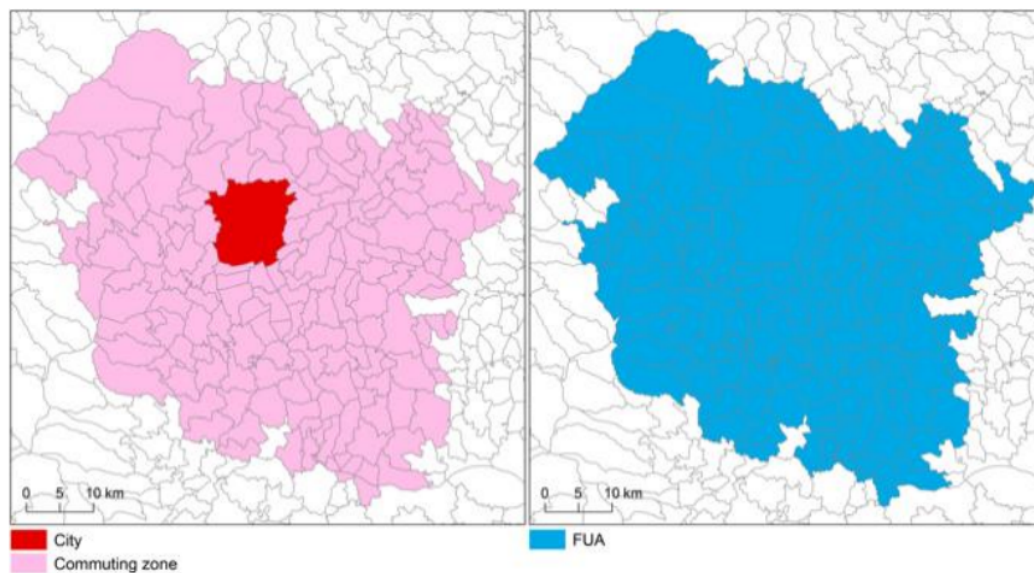


Figura 2 – À esquerda estão delimitadas a cidade de Graz, na Áustria, em vermelho e sua zona de deslocamento, em rosa. À direita a área urbana funcional compreendendo as duas regiões (DIJKSTRA; POELMAN; VENERI, 2019).

3 MATERIAIS E MÉTODOS

3.1 Análise dos Dados: Casos Acumulados por *FUAs*

As regiões consideradas para analisar os valores de R_0 foram as regiões geográficas imediatas (RGIs) do Brasil e as áreas estatísticas metro e micropolitanas (*metropolitan and micropolitan statistical areas - MSAs*) dos Estados Unidos. Essas regiões possuem definições que se assemelham, portanto, à de áreas urbanas funcionais *FUAs*.

Segundo o IBGE (2017), as RGIs são regiões estruturadas a partir de centros urbanos próximos para a satisfação das necessidades imediatas da população como compras de bens de consumo duráveis e não duráveis, busca de trabalho, procura por serviços de saúde e educação e prestação de serviços públicos. Ainda, as *MSAs* são definidas como áreas contendo um núcleo populacional substancial, juntamente com comunidades adjacentes com alto grau de integração econômica e social com esse núcleo (CENSUS, 2020).

Para analisar como uma epidemia se propaga nestes níveis urbanos, foram obtidas as séries temporais de casos acumulados de *COVID-19* para os municípios de cada país em Brasil.io (2020) e CSSE (2020). Em seguida foram obtidas tabelas em sites governamentais com as estimativas populacionais e os códigos referentes às áreas estatísticas de interesse. Com o auxílio da biblioteca *Pandas*, disponível na linguagem de programação *Python*, estes dados foram então mesclados e agrupados, resultando em um *dataframe*¹ de casos acumulados por área urbana funcional para cada país. A figura 3 mostra parte do *dataframe* construído para possibilitar a análise dos números de reprodução para as RGIs do Brasil.

¹*Dataframe* é uma estrutura de dados bidimensional com os valores alinhados de forma tabular em linhas e colunas, mutável em tamanho e potencialmente heterogênea.

	RGI	POP	2020-02-25	2020-02-26	2020-02-27	...	2021-02-20	2021-02-21	2021-02-22	2021-02-23
0	110001	666953	0.0	0.0	0.0	...	62417.0	62702.7	62993.0	63282.4
1	110002	242279	0.0	0.0	0.0	...	17735.1	17849.3	17965.3	18106.0
2	110003	122216	0.0	0.0	0.0	...	7819.9	7868.8	7926.8	7989.1
3	110004	313023	0.0	0.0	0.0	...	17995.6	18165.9	18338.7	18516.4
4	110005	308988	0.0	0.0	0.0	...	20629.5	20747.5	20866.8	20995.8
...
505	520019	722640	0.0	0.0	0.0	...	27976.6	28032.7	28088.7	28155.6
506	520020	506355	0.0	0.0	0.0	...	14076.9	14108.4	14137.0	14168.4
507	520021	127197	0.0	0.0	0.0	...	2689.0	2699.1	2712.3	2729.4
508	520022	57410	0.0	0.0	0.0	...	845.1	847.0	849.1	850.7
509	530001	3094325	0.0	0.0	0.0	...	251392.4	252045.0	252718.0	253381.1

510 rows x 367 columns

Figura 3 – Parte do *dataframe* construído com o auxílio da biblioteca *Pandas* com as médias móveis dos casos acumulados para *RGI*s do Brasil.

Pôde-se observar desses dados que o início da epidemia apresentou comportamento distinto dependendo da área urbana funcional, de acordo com as diferentes taxas de crescimento para cada curva de casos a partir do primeiro relato. Isso pode ser visto na figura 4.

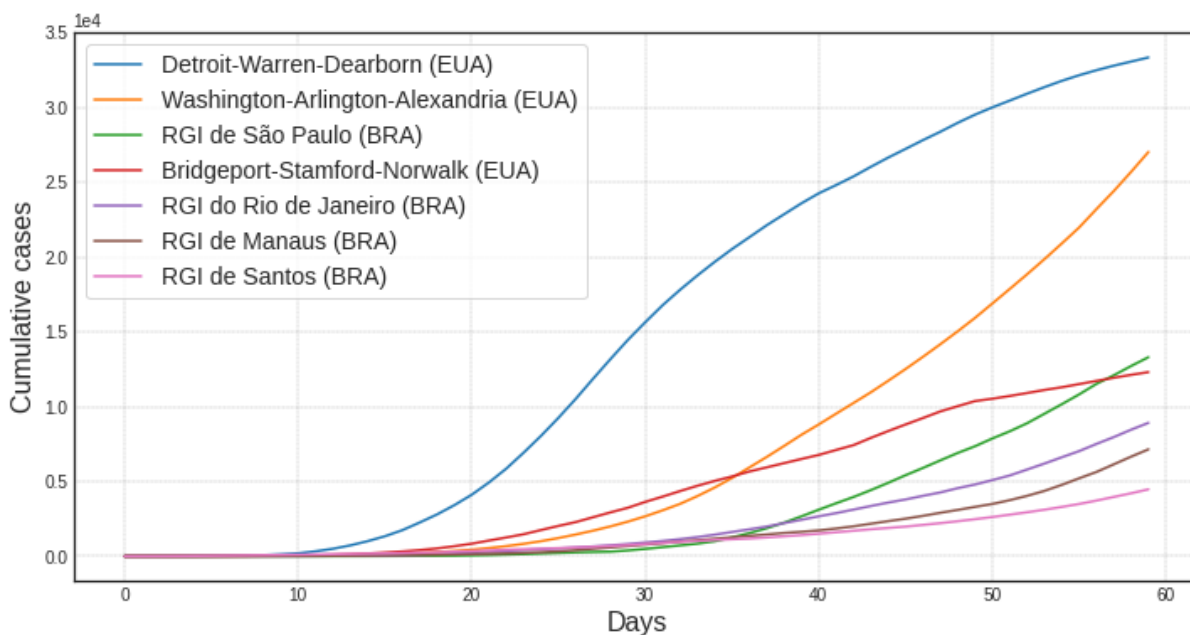


Figura 4 – Curvas de médias móveis de casos acumulados a partir do primeiro caso confirmado para as diferentes áreas urbanas funcionais analisadas. Nota-se uma diferença entre as taxas de crescimentos.

3.2 Modelos SIR e SEIR: Taxas de Crescimento Exponencial

Os valores de R_0 foram obtidos considerando-se as relações (2.14) e (2.19), sendo as taxas de crescimento λ calculada como

$$C(t) = C(t_0)e^{\lambda t} \Rightarrow \lambda = \frac{\ln C(t) - \ln C(t_0)}{t}, \quad (3.1)$$

onde $C(t)$ é o número de casos acumulados t dias após t_0 , relativo ao primeiro dia em que $C \neq 0$, e $C(t_0)$ é o número de casos acumulados em t_0 .

Substituindo (3.1) em (2.14), tem-se que

$$R_0^{SIR} = \frac{\ln C(t) - \ln C(t_0)}{\gamma t} - 1. \quad (3.2)$$

Da mesma forma, substituindo (3.1) em (2.19), tem-se que

$$R_0^{SEIR} = \frac{1}{\sigma \gamma t} \{[\ln C(t) - \ln C(t_0) + \sigma t] [\ln C(t) - \ln C(t_0) + \gamma t]\}. \quad (3.3)$$

Os valores de σ e γ foram obtidos a partir dos períodos de incubação $\tau_e = 1/\sigma = 5.1$ dias e infecção $\tau = 1/\gamma = 11.5$ dias para a *COVID-19* (LAUER et al., 2020).

Na figura 5 pode-se ver o ajuste feito aos casos acumulados no início da epidemia para a região geográfica imediata de Bento Gonçalves.

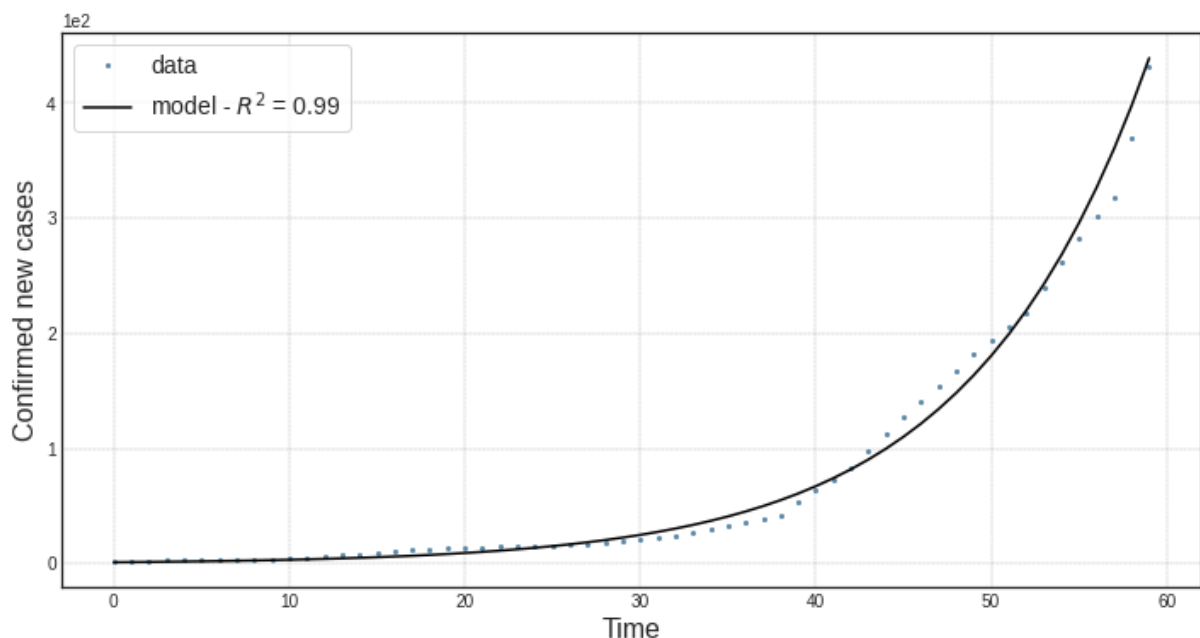


Figura 5 – Ajuste exponencial à curva de casos acumulados no momento inicial da epidemia para a RGI de Bento Gonçalves. O valor da taxa de crescimento associado a este ajuste foi usado para estimar o R_0 tanto no modelo SIR quanto no SEIR.

3.3 Método de Cori: Valores Médios de R_t

Para determinar o R_0 para cada unidade urbana a partir da equação (2.20) foi desenvolvida uma função em *Python* com o uso da biblioteca *Epyestim* (HILFIKER, 2020), que adapta parâmetros da *Covid-19* ao pacote *EpiEstim* (CORI et al., 2013a), disponível na linguagem de programação R. Este método é utilizado para estimar o valor de R_t , que varia ao longo do tempo, e devem ser considerados parâmetros como a janela de tempo utilizada para análise dos casos, o intervalo de geração, que irá ocorrer entre uma infecção primária e secundária, e o tempo de atraso entre uma infecção e o caso relatado.

3.3.1 Aplicação em Dados Sintéticos

Para entender melhor o funcionamento do método, foram gerados dados sintéticos com o modelo SEIR a fim de comparar as estimativas de R_t de casos reportados com as de dados onde o tempo exato da infecção é conhecido. Esses dados foram obtidos a partir dos valores de 2.00, 0.80 e 1.10 para o número de reprodução ao longo de 30 dias cada, num total de 90 dias, considerando um tempo de geração de quatro dias.

Para simular a série temporal de casos confirmados, o tempo de atraso entre uma infecção e uma notificação seguiu uma distribuição de probabilidade levando em conta um período de incubação médio de cinco dias. Isso foi feito com o auxílio da biblioteca *Epyestim* (HILFIKER, 2020).

Para estimar R_t a partir desses dados, deve-se considerar uma janela de tempo adequada. Na figura 6 pode-se ver como se comportam as estimativas de R_t para infecções e casos confirmados, frente aos valores exatos do número de reprodução, para diferentes janelas de tempo de análise das infecções ou casos relatados. A precisão dos resultados obtidos em cada situação pode ser vista na tabela 1.

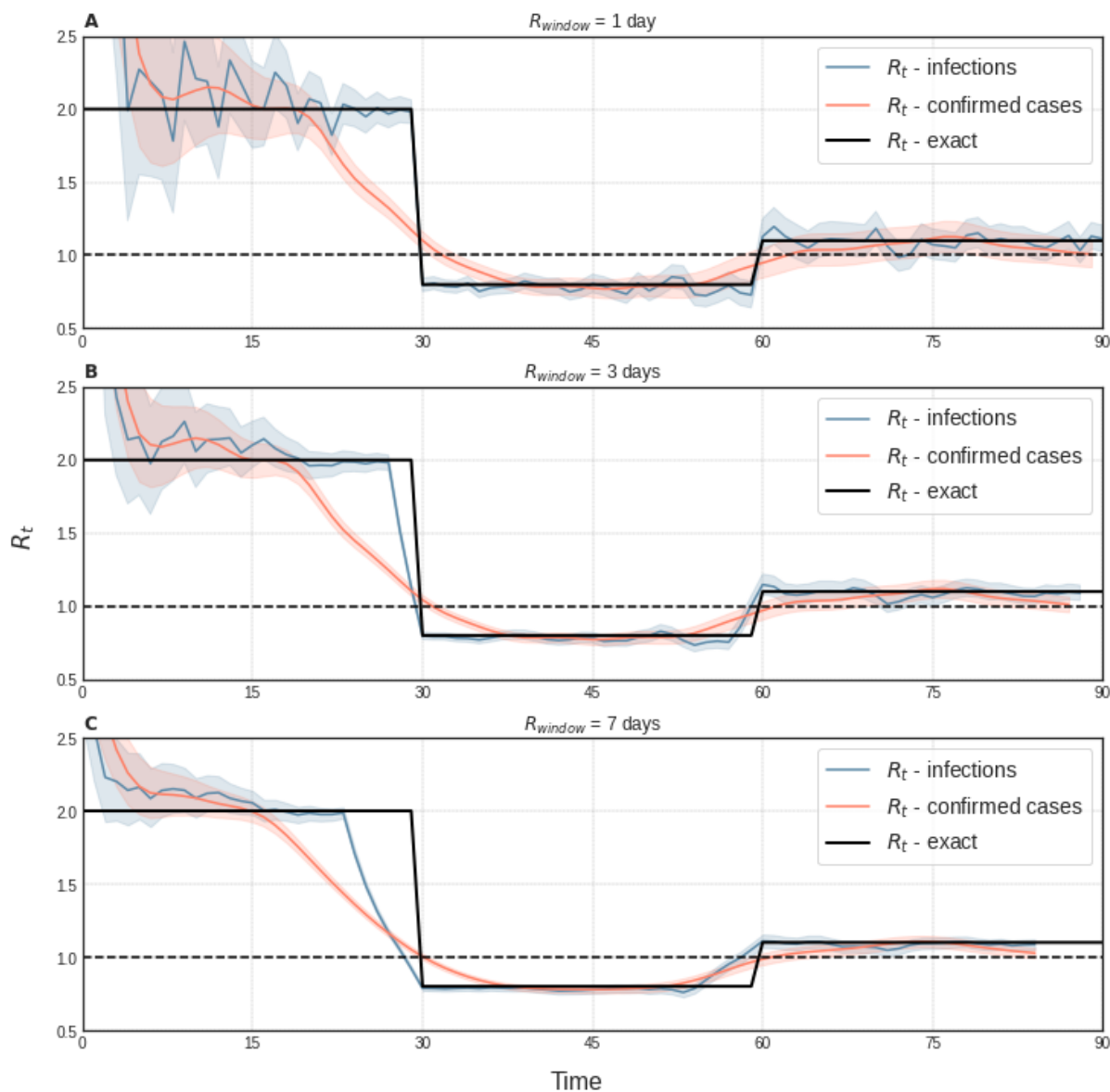


Figura 6 – Estimativas de R_t para dados sintéticos de infecções e casos reportados em comparação com o valor exato para diferentes janelas de tempo: A) Um dia; B) Três dias e C) Sete dias.

Tabela 1 – Tabela com valores de R_0 para dados sintéticos, considerando o valor médio de R_t nos primeiros 30 dias, onde o valor exato de R_0 é 2.00.

Dados	Janela de tempo (dias)	R_0	Erro relativo
Infecções	1	3.01	0.51
	3	2.51	0.25
	7	1.97	0.02
Casos relatados	1	2.26	0.13
	3	2.12	0.06
	7	1.91	0.04

O tempo de geração entre as infecções também possui um papel importante nos resultados. Tempos muito longos levam a um R_t superestimado quando $R_t > 1$ e subestimado em $R_t < 1$ (GOSTIC et al., 2020). Isso reforça a importância de usar aproximações precisas para as distribuições de tempo de geração, a partir dos intervalos seriais². Ressalta-se ainda que essas distribuições podem variar ao longo de uma epidemia, e que neste trabalho isso não foi considerado.

A figura 7 mostra como o número de reprodução varia quando adota-se um tempo médio igual a sete dias para geração entre os casos reportados, ou seja, um tempo maior do que o valor verdadeiro de quatro dias entre as infecções.

Os resultados obtidos para R_0 , considerando diferentes intervalos de geração para infecções e casos relatados, podem ser vistos na tabela 2.

²Intervalo serial é o tempo entre o início dos sintomas em um par infectador-infectado.

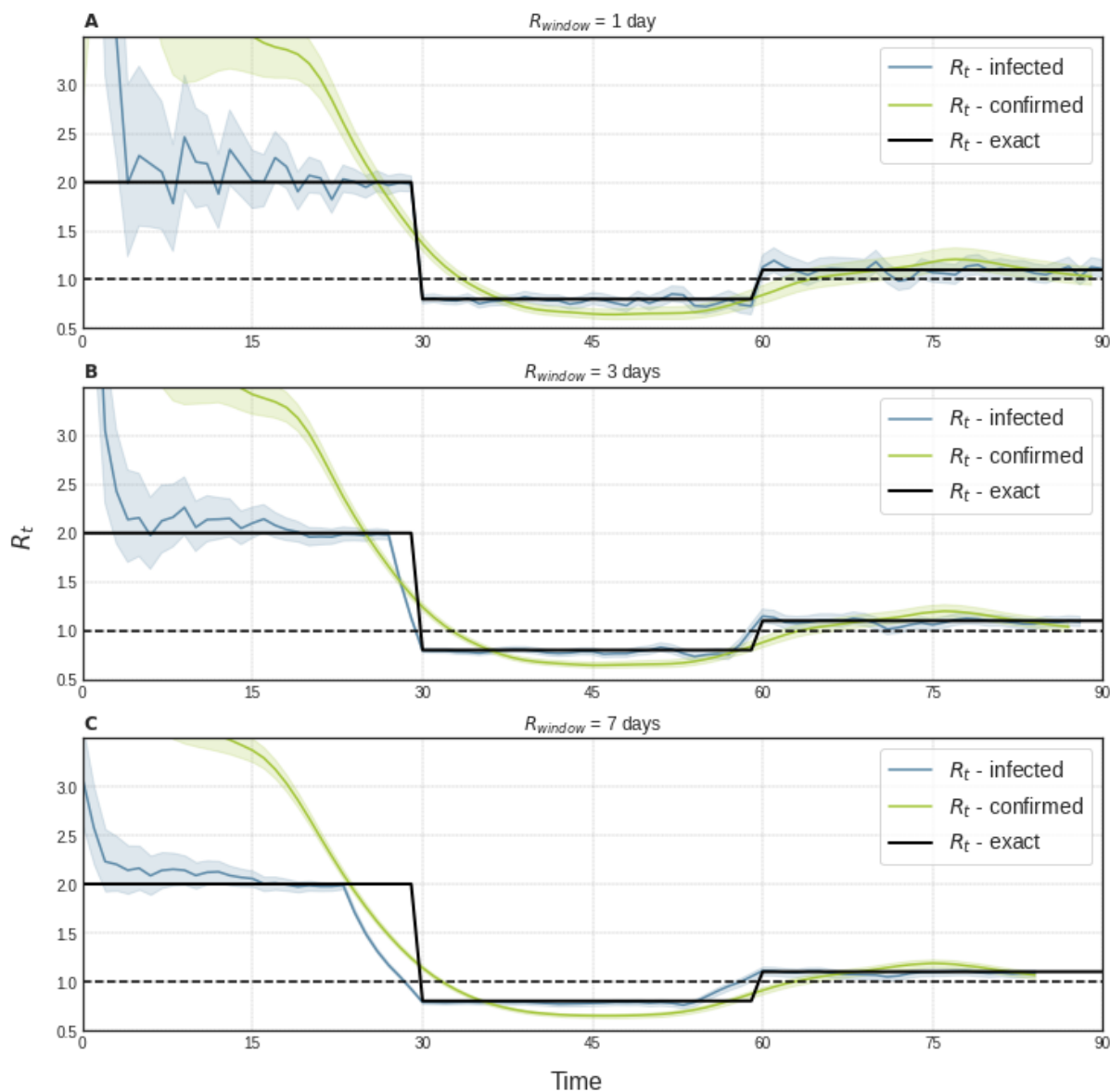


Figura 7 – R_t para dados sintéticos de infecções e casos reportados em comparação com o valor exato (em preto), considerando um intervalo de geração com média de 7 dias entre um caso primário e secundário (curvas verdes) ante à mesma média de 4 dias entre infecções primárias e secundárias (curvas azuis). Observe-se uma sensibilidade maior nos resultados para as mesmas janelas de tempo consideradas na figura 6, de: A) Um dia; B) Três dias e C) Sete dias.

Tabela 2 – Tabela com valores de R_0 para dados sintéticos com intervalos de geração com média de 4 dias para infecções e 7 dias para relatos. O R_0 foi considerando como sendo o valor médio de R_t nos primeiros 30 dias, onde o valor exato de R_0 é 2.00.

Dados	Janela de tempo (dias)	Tempo de geração (dias)	R_0	Erro relativo
Infecções	1	4	3.01	0.51
	3		2.51	0.25
	7		1.97	0.02
Casos relatados	1	7	3.97	0.98
	3		4.00	1.00
	7		3.50	0.76

Outro parâmetro que pode ser testado com a biblioteca *EpyEstim* (HILFIKER, 2020) é a distribuição do tempo de atraso entre infecções e casos relatados. Aumentando este tempo para dez dias em média, pôde-se notar que os valores de R_t são menos sensíveis a uma especificação incorreta do tempo de atraso do que do tempo de geração, conforme observado na figura 8 e na tabela 3.

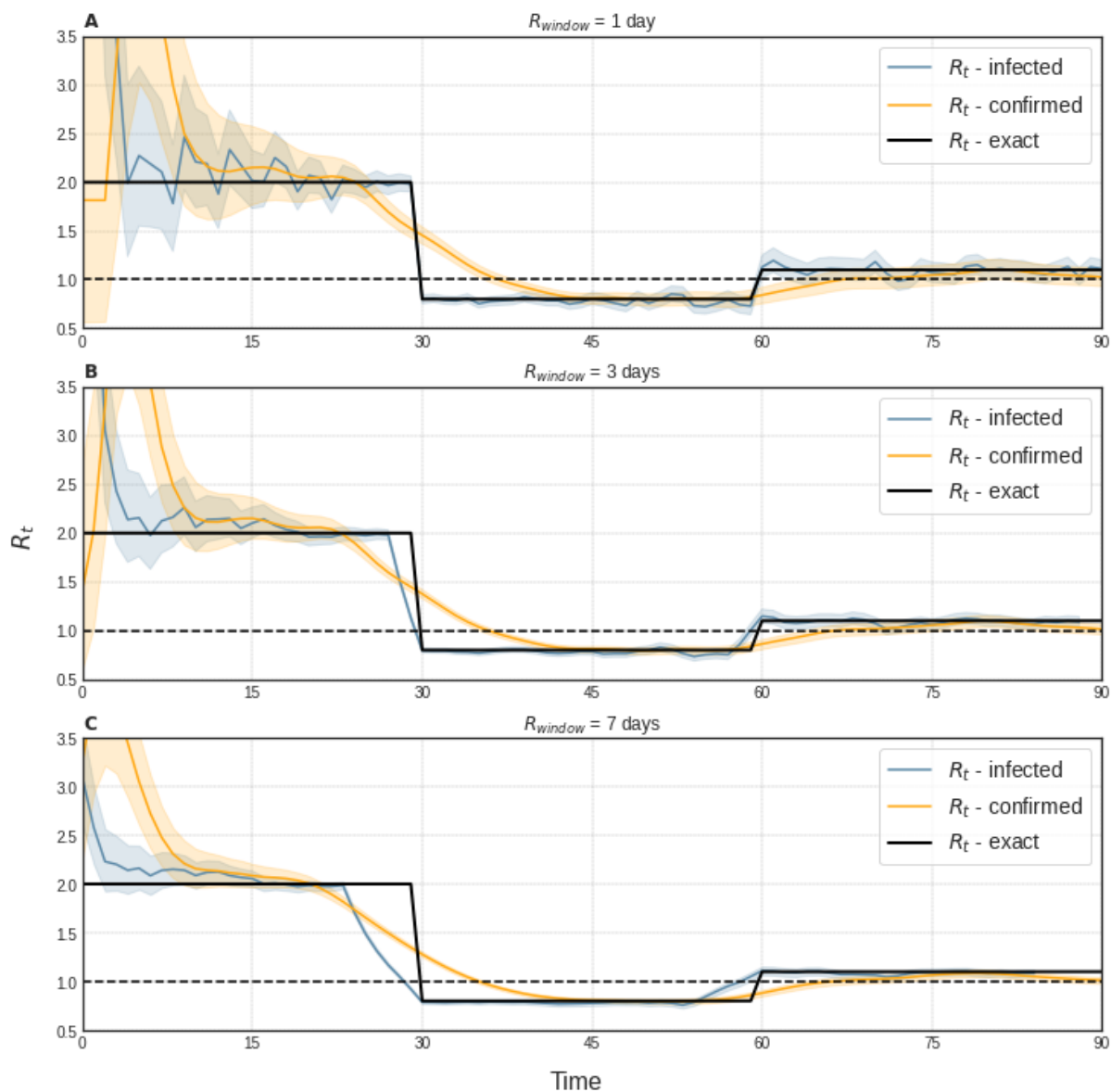


Figura 8 – R_t para dados sintéticos de infecções e casos reportados em comparação com o valor exato (em preto) considerando um tempo de atraso com média de 10 dias (curvas laranjas) para as mesmas janelas de tempo consideradas na figura 6, de: A) Um dia; B) Três dias e C) Sete dias.

Tabela 3 – Tabela com valores de R_0 para dados sintéticos com tempo médio de 10 dias para atraso entre uma infecção e um caso relatado. O R_0 foi considerado como sendo o valor médio de R_t nos primeiros 30 dias, onde o valor exato de R_0 é 2.00.

Dados	Janela de tempo (dias)	Tempo de atraso (dias)	R_0	Erro relativo
Infecções	1	0	3.01	0.51
	3		2.51	0.25
	7		1.97	0.02
Casos relatados	1	10	2.41	0.21
	3		2.38	0.19
	7		2.31	0.15

3.3.2 Estimativas para FUAs

Para aplicar o método de Cori para as populações das áreas urbanas funcionais, foi considerada uma janela de tempo para análise de sete dias, de acordo com os melhores resultados obtidos para os dados sintéticos (figura 6). O tempo de geração foi aproximado por uma distribuição gama para o intervalo serial, com média de quatro dias (FLAXMAN et al., 2020). Para o tempo de atraso entre uma infecção e um caso relatado, foi considerado um período de cinco dias, relativo ao período de incubação da *COVID-19* (BRAUNER et al., 2021).

Conforme demonstrado na figura 9 para a região geográfica imediata de São Paulo e para a área estatística metropolitana de St. Louis, foi possível observar o comportamento do R_t no início da epidemia para cada área urbana funcional. Este início foi considerado, assim como nos modelos SIR e SEIR, como sendo os primeiros sessenta dias a partir do primeiro caso confirmado, considerando as médias móveis dos casos acumulados, a fim de conseguir aproximar os valores médios de R_t aos valores esperados de R_0 .

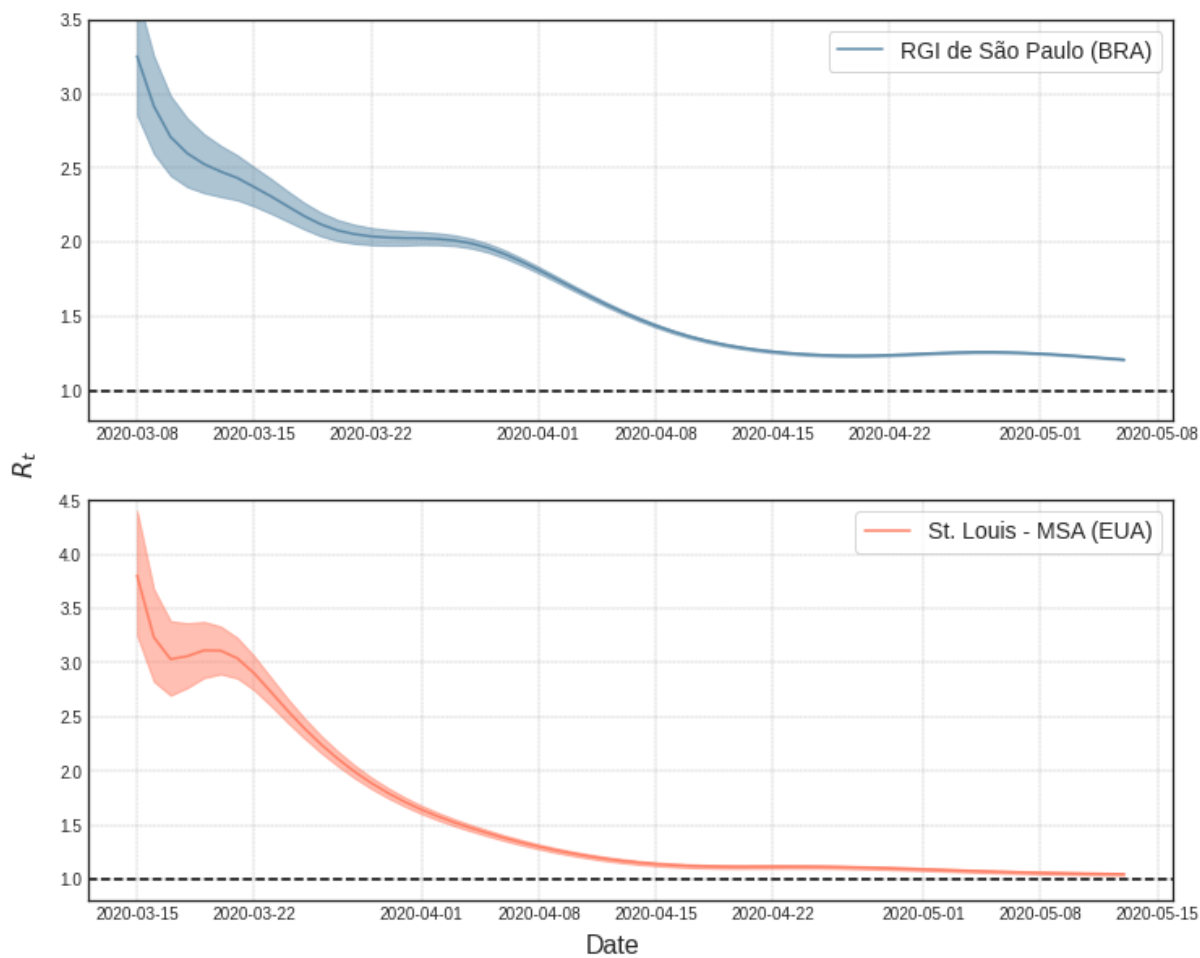


Figura 9 – Número de reprodução efetivo em função do tempo para RGI de São Paulo e MSA de St. Louis retornados pela função baseada no pacote *EpiEstim* (CORI et al., 2013a), para 60 dias a partir do primeiro caso confirmado.

4 RESULTADOS E DISCUSSÕES

Aos valores de R_0 retornados pelos modelos SIR e SEIR foram feitos ajustes pelo método dos mínimos quadrados. Nas figuras 10 e 11 pode-se ver a relação entre R_0 e o tamanho da população para as áreas urbanas funcionais do Brasil e dos Estados Unidos.

A estes valores foi feito um ajuste pelo método dos mínimos quadrados para

$$R_0 = a + b \ln(N) \quad (4.1)$$

sendo os valores dos parâmetros a e b , bem como os coeficiente de correlação indicando uma relação positiva entre as variáveis analisadas, indicados na tabela 4.

Tabela 4 – Tabela com valores dos parâmetros a e b e do coeficiente de correlação de Pearson ρ para os ajustes feitos pelo método dos mínimos quadrados para os modelos SIR e SEIR.

Método	FUAs	a	b	ρ
SIR	RGIs (BRA)	0.178	0.150	0.50
	MSAs (EUA)	-0.302	0.193	0.75
SEIR	RGIs (BRA)	-1.437	0.358	0.51
	MSAs (EUA)	-2.548	0.460	0.76

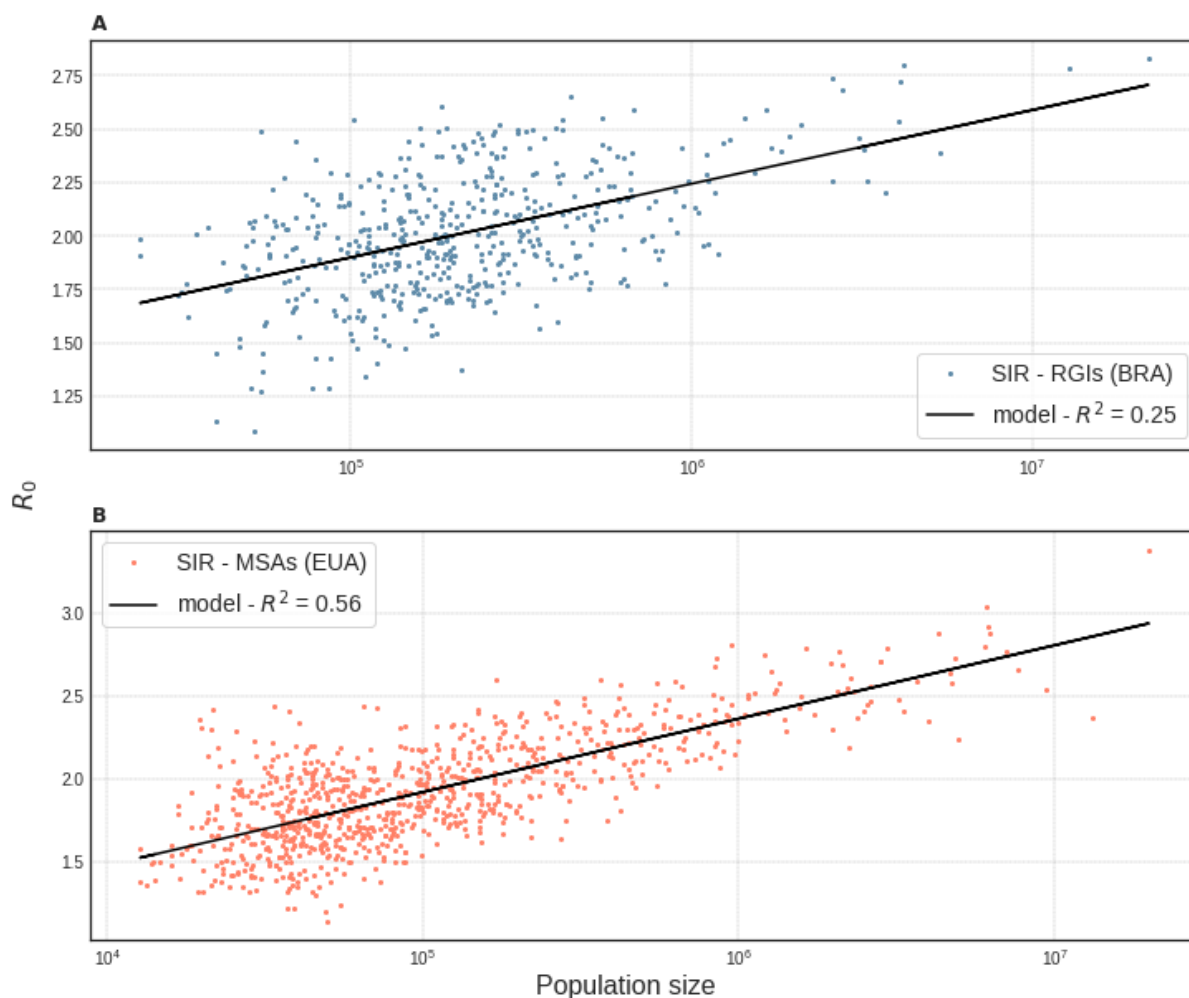


Figura 10 – Valores de R_0 obtidos pelo modelo SIR para: A) RGIs do Brasil com ajuste $R_0 = 0.178 + 0.150 \ln(N)$ e coeficiente de correlação $\rho = 0.50$; B) MSAs dos Estados Unidos com ajuste $R_0 = -0.302 + 0.193 \ln(N)$ e coeficiente de correlação $\rho = 0.75$.

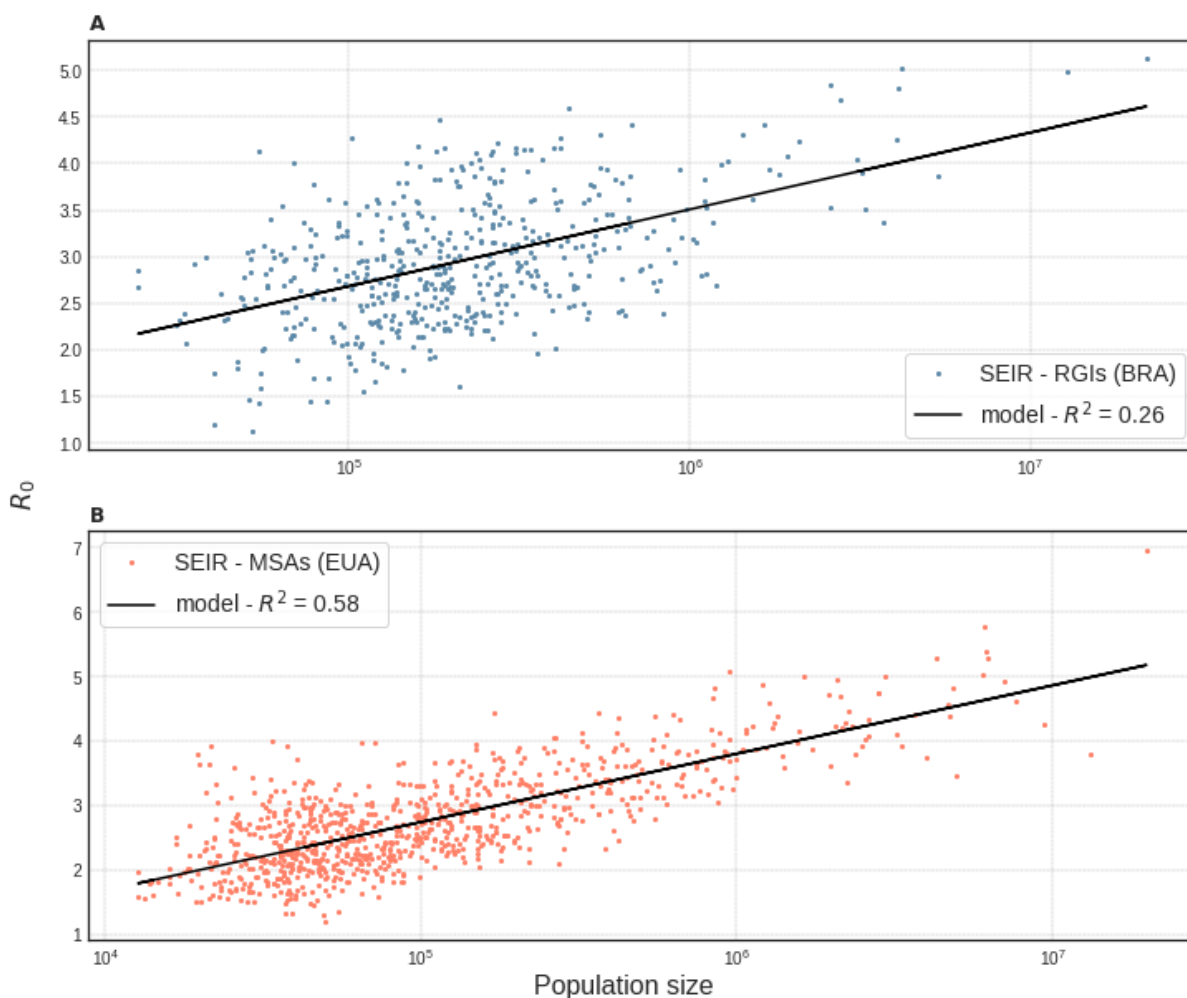


Figura 11 – Valores de R_0 obtidos pelo modelo SEIR para: A) RGIs do Brasil com ajuste $R_0 = -1.437 + 0.358 \ln(N)$ e coeficiente de correlação $\rho = 0.51$; B) MSAs dos Estados Unidos com ajuste $R_0 = -2.548 + 0.460 \ln(N)$ e coeficiente de correlação $\rho = 0.76$.

Os ajustes feitos aos valores médios de R_t no início da epidemia para cada região também mostraram um comportamento de escala de R_0 com o tamanho da população. Os valores dos parâmetros a e b e os coeficientes ρ podem ser vistos na tabela 5.

Tabela 5 – Tabela com valores dos parâmetros a e b e do coeficiente de correlação de Pearson ρ para o ajuste feito pelo método dos mínimos quadrados para o método de Cori.

Método	FUAs	a	b	ρ
CORI	RGIs (BRA)	0.895	0.039	0.33
	MSAs (EUA)	0.561	0.062	0.60

Na figura 12 pode-se ver que os mesmos procedimentos relativos aos ajustes nas figuras anteriores puderam ser repetidos, mantendo os valores positivos de ρ .

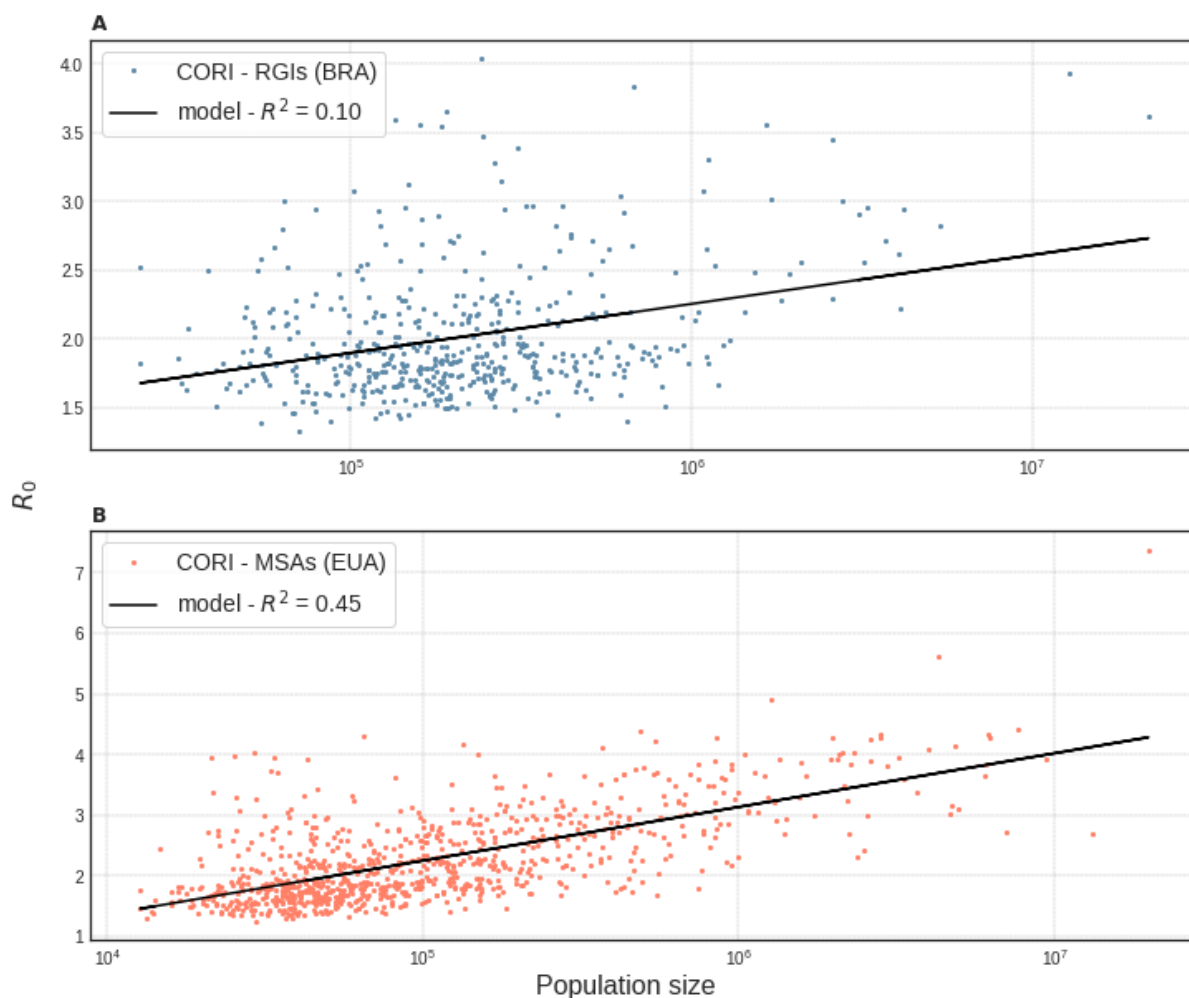


Figura 12 – Estimativas de R_0 obtidos pelo método de Cori para: A) RGIs do Brasil com ajuste $R_0 = 0.895 + 0.039 \ln(N)$ e coeficiente de correlação $\rho = 0.33$; B) MSAs dos Estados Unidos com ajuste $R_0 = 0.561 + 0.062 \ln(N)$ e coeficiente de correlação $\rho = 0.60$.

As populações das áreas urbanas analisadas foram também divididas em dez grupos em cada país, a fim de se observar o comportamento do R_0 a partir de seus valores médios e dos desvios-padrão dentro de cada região. Isso pode ser observado na figura 13, que mostra os valores de R_0 retornados a partir do modelo SEIR.

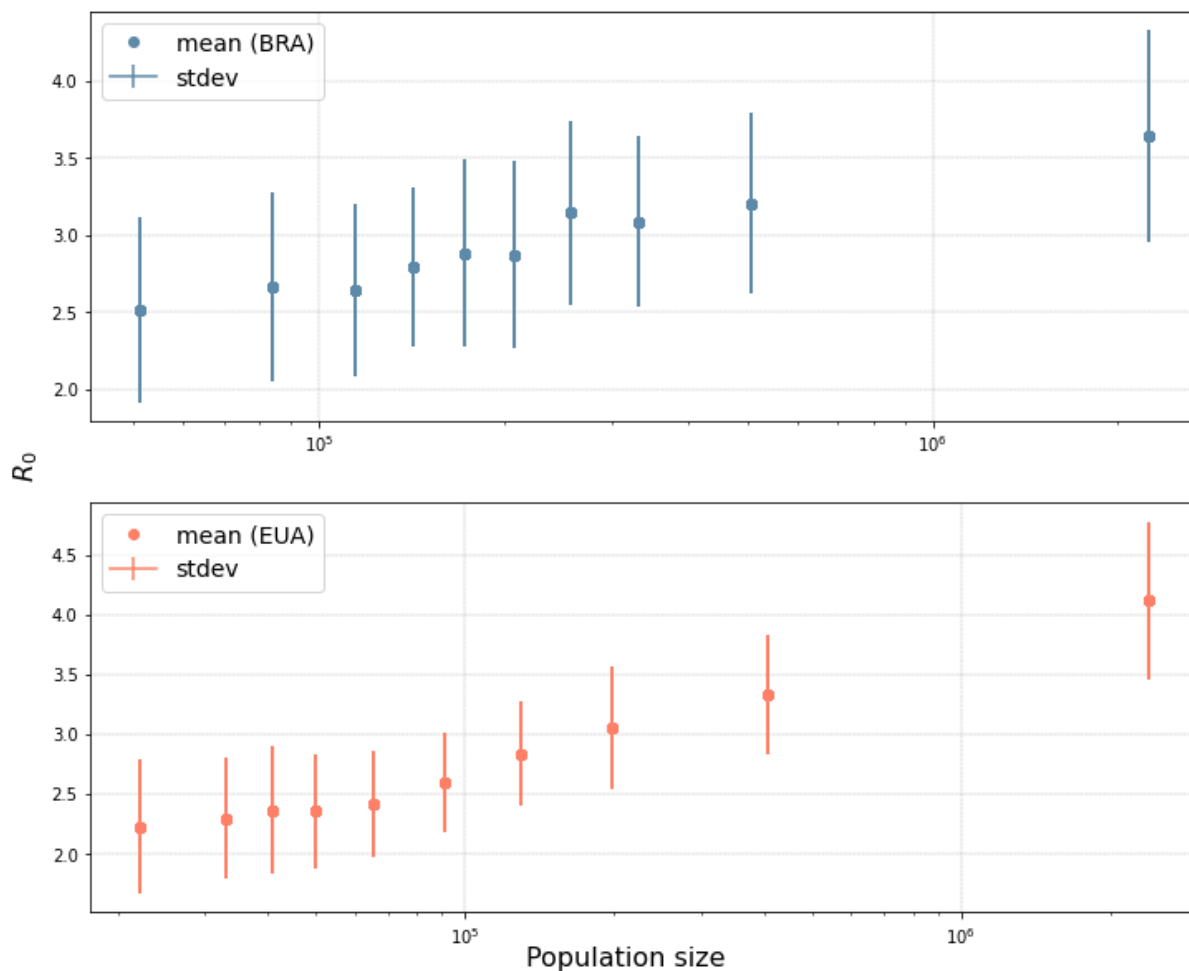


Figura 13 – Valores médios de R_0 e seus desvios-padrão para dez grupos de populações de áreas funcionais para cada país.

A relação (4.1) pode ser explicada por uma teoria da escala de contatos (CARDOSO; GONÇALVES, 2021), disponível no Apêndice A.

5 CONCLUSÕES

A grande quantidade de dados gerados pela pandemia de *COVID-19*, contribuiu para demonstrar que hipóteses normalmente utilizada para explicar a dinâmica de epidemias, como a ação de massa, não são as mais adequadas.

Modelos compartimentais tradicionais como SIR e SEIR mostraram-se ferramentas eficazes para estimar o número básico de reprodução para diferentes populações a partir de dados de casos acumulados. Isso ocorre porque o início de uma epidemia apresenta um comportamento exponencial, cuja taxa de crescimento pode ser descrita em função dos parâmetros que são utilizados para o cálculo de R_0 (equações (2.13) e (2.18)).

Pelo método de Cori, essa estimativa também foi possível devido a suposições paramétricas mínimas sobre a epidemia e a capacidade de identificar mudanças abruptas no número efetivo de reprodução quando usados dados sintéticos.

Ainda, foi demonstrada a importância de considerar as distribuições para o intervalo de geração (figura 7) e para atrasos na observação de casos (figura 8) para uma correta observação do número efetivo de reprodução ao longo do tempo.

Finalmente, os ajustes feitos nos resultados encontrados para o número de reprodução, independentemente do método empregado ou do período analisado, indicam que a taxa de contatos entre indivíduos é impulsionada pelo tamanho da população para áreas urbanas funcionais.

Referências

- BERNOULLI, D.; PETROPOLITANAE, N. C. A. S. I. De duratione media matrimoniorum, pro quacunq̃ue coniugum aetate, aliisque quaestionibus affinibus. **Mém. de l'Acad. Royale des Sciences de Paris**, p. 1–45, 1760. Citado na página 1.
- BETTENCOURT, L. M. The origins of scaling in cities. **science**, American Association for the Advancement of Science, v. 340, n. 6139, p. 1438–1441, 2013. Citado na página 32.
- BETTENCOURT, L. M. et al. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. **PloS one**, Public Library of Science San Francisco, USA, v. 5, n. 11, p. e13541, 2010. Citado na página 9.
- BETTENCOURT, L. M.; RIBEIRO, R. M. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. **PloS one**, Public Library of Science San Francisco, USA, v. 3, n. 5, p. e2185, 2008. Citado na página 7.
- BRASIL.IO. **dataset: dataset covid19**. 2020. <<https://brasil.io/dataset/covid19/files/>>. Accessed: 2022. Citado na página 11.
- BRAUNER, J. M. et al. Inferring the effectiveness of government interventions against covid-19. **Science**, American Association for the Advancement of Science, v. 371, n. 6531, p. eabd9338, 2021. Citado na página 20.
- CARDOSO, B.-H. F.; GONÇALVES, S. Urban scaling of covid-19 epidemics. **arXiv preprint arXiv:2005.07791**, 2020. Citado na página 2.
- CARDOSO, B.-H. F.; GONÇALVES, S. Universal scaling law for human-to-human transmission diseases. **Europhysics Letters**, IOP Publishing, v. 133, n. 5, p. 58001, 2021. Citado 2 vezes nas páginas 2 e 26.
- CENSUS, U. S. **Metropolitan and Micropolitan**. 2020. <<https://www.census.gov/programs-surveys/metro-micro.html>>. Accessed: 2022. Citado na página 11.
- CORI, A. et al. **Package ‘EpiEstim’**. [S.l.]: Version, 2013. Citado 2 vezes nas páginas 14 e 21.
- CORI, A. et al. A new framework and software to estimate time-varying reproduction numbers during epidemics. **American journal of epidemiology**, Oxford University Press, v. 178, n. 9, p. 1505–1512, 2013. Citado 2 vezes nas páginas 2 e 7.
- CSSE, J. **COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University**. 2020. <<https://github.com/CSSEGISandData/COVID-19>>. Accessed: 2022. Citado na página 11.
- DIETZ, K. The estimation of the basic reproduction number for infectious diseases. **Statistical methods in medical research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 2, n. 1, p. 23–41, 1993. Citado na página 5.
- DIJKSTRA, L.; POELMAN, H.; VENERI, P. The eu-oecd definition of a functional urban area. OECD, 2019. Citado 2 vezes nas páginas 9 e 10.

FLAXMAN, S. et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. **Nature**, Nature Publishing Group, v. 584, n. 7820, p. 257–261, 2020. Citado na página 20.

GOSTIC, K. M. et al. Practical considerations for measuring the effective reproductive number, *r*. **PLoS computational biology**, Public Library of Science San Francisco, CA USA, v. 16, n. 12, p. e1008409, 2020. Citado 3 vezes nas páginas 6, 7 e 16.

HILFIKER, L. **Josi J. epyestim. Python package to estimate the time-varying effective reproduction number of an epidemic from reported case numbers. 2020.** 2020. Citado 2 vezes nas páginas 14 e 18.

HU, H.; NIGMATULINA, K.; ECKHOFF, P. The scaling of contact rates with population density for the infectious disease models. **Mathematical biosciences**, Elsevier, v. 244, n. 2, p. 125–134, 2013. Citado na página 32.

IBGE. Divisão regional do brasil em regiões geográficas imediatas e regiões geográficas intermediárias: 2017. **Coordenação de Geografia. Rio de Janeiro: IBGE**, 2017. Citado na página 11.

KEELING, M. Rohani. p.(2008). modeling infectious diseases in humans and animals. **Clin. Infect. Dis**, v. 47, p. 864–866, 2008. Citado na página 2.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character**, The Royal Society London, v. 115, n. 772, p. 700–721, 1927. Citado 2 vezes nas páginas 1 e 3.

KEUSCHNIGG, M.; MUTGAN, S.; HEDSTRÖM, P. Urban scaling and the regional divide. **Science advances**, American Association for the Advancement of Science, v. 5, n. 1, p. eaav0042, 2019. Citado na página 9.

LAUER, S. A. et al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. **Annals of internal medicine**, American College of Physicians, v. 172, n. 9, p. 577–582, 2020. Citado na página 13.

MA, J. Estimating epidemic exponential growth rate and basic reproduction number. **Infectious Disease Modelling**, Elsevier, v. 5, p. 129–141, 2020. Citado na página 5.

MILEGO, R. et al. Espon fuore: functional urban areas and regions in europe. ESPON, 2019. Citado na página 9.

MURRAY, J. D. **Mathematical biology: I. An introduction**. [S.l.]: Springer, 2002. Citado na página 1.

NOULAS, A. et al. A tale of many cities: universal patterns in human urban mobility. **PloS one**, Public Library of Science San Francisco, USA, v. 7, n. 5, p. e37027, 2012. Citado na página 9.

PAHO. **Pan American Health Organization (PAHO)**. 2020. <<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>. Accessed: 2022. Citado na página 1.

SCHLÄPFER, M. et al. The scaling of human interactions with city size. **Journal of the Royal Society Interface**, The Royal Society, v. 11, n. 98, p. 20130789, 2014. Citado na página 9.

STRANO, E.; SOOD, V. Rich and poor cities in europe. an urban scaling approach to mapping the european economic transition. **PloS one**, Public Library of Science San Francisco, CA USA, v. 11, n. 8, p. e0159465, 2016. Citado na página 9.

TOMÉ, T.; OLIVEIRA, M. J. d. Epidemic spreading. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 42, 2020. Citado na página 4.

WALLINGA, J.; TEUNIS, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. **American Journal of epidemiology**, Oxford University Press, v. 160, n. 6, p. 509–516, 2004. Citado na página 7.

Apêndices

APÊNDICE A – Teoria da Escala de Contatos

Considerando $N \gg 1$ indivíduos — representados como nós — distribuídos uniformemente em um espaço bidimensional com área de terra A , pode-se esperar que um indivíduo i forme uma ligação com um indivíduo j com probabilidade

$$P_{ij} = \frac{1}{\text{rank}_i(j)}, \quad (\text{A.1})$$

onde $\text{rank}_i(j)$ é o número de vizinhos mais próximos de i do que j .

Assumindo que a distância entre esses dois indivíduos é r , temos então

$$\text{rank}_i(j) = \rho\pi r^2 \Rightarrow P_{ij} = P(r) = \frac{1}{\rho\pi r^2}, \quad (\text{A.2})$$

com $\rho = N/A$ sendo a densidade populacional.

Esta relação de gravidade tem que ser limitada para preencher as condições para a probabilidade total e número de pessoas, então $r \in [r_0, r_1]$, com os limites definidos como segue: Primeiro, já que $0 \leq P \leq 1$, deve-se impor um raio de corte inferior r_0 tal que

$$P(r \leq r_0) \equiv 1 \Rightarrow r_0 = \frac{1}{\sqrt{\pi\rho}}; \quad (\text{A.3})$$

Segundo, é natural assumir um raio de corte superior $r_1 \propto \sqrt{A}$ para longas distâncias (BETTENCOURT, 2013).

Então, o grau médio da rede de interação social será

$$k = \frac{1}{2} \int_0^{r_1} dr (2\pi r \rho) P(r) = \frac{1}{2} \left[1 + 2 \int_{r_0}^{r_1} \frac{dr}{r} \right] = \left[1 + \ln \left(\frac{r_1^2}{r_0^2} \right) \right] \Rightarrow k = \ln(bN), \quad (\text{A.4})$$

onde o fator $b \equiv e\pi r_1^2/A$ dá a extensão da mobilidade média individual dentro da unidade demográfica.

Fatorando o número básico de reprodução como

$$R_0 = p\tau\mathcal{C} \quad (\text{A.5})$$

onde p é a probabilidade de transmissão, τ é o período infeccioso e \mathcal{C} é a taxa de contato *per capita*, espera-se que \mathcal{C} seja proporcional ao grau médio da interação social k (HU; NIGMATULINA; ECKHOFF, 2013).

Logo, tem-se que $R_0 \propto \mathcal{C} \propto k$.

Então, da Eq. A.4 chega-se na relação de escala:

$$R_0 \propto \ln(bN), \quad (\text{A.6})$$

onde b é uma constante para todos locais.