



EPIDEMIOLOGY EDUCATIONAL CODE

[www.EpidemicCode.org](http://www.EpidemicCode.org)

# Equipe de desenvolvimento

**Professores/  
Pesquisadores:** Americo Cunha (UERJ)  
Lisandro Lovisolo (UERJ)  
Malú Grave (UFRJ)  
Rodrigo Burgos (UERJ)

---

**Estudantes:** Bruna Pavlack (IFMS) Leonardo de la Roca (UERJ)  
Diego Matos (UERJ) Lucas Chaves (UFU)  
Eber Dantas (UFRJ) Marcos Issa (UERJ)  
João P. Norenberg (UNESP) Michel Tosin (UERJ)  
Julio Basilio (UERJ) Roberto Luo (UERJ)

---

**Design Gráfico:** Amanda Cunha Guyt

---

**Contato:** Americo Cunha [americocunha@uerj.br](mailto:americocunha@uerj.br)

# O que é o EPIDEMIC?

O EPIDEMIC é uma ferramenta educacional para análises epidemiológicas (simulações, monitoramento, etc), desenvolvida por colaboradores de diversas instituições no âmbito da pandemia de COVID-19.

Missão:

- analisar indicadores de evolução de uma epidemia;
- instruir sobre simulações, análises, previsões epidemiológicas;
- fornecer uma ferramenta didática e intuitiva para o público interessado.



Neste tutorial, as informações dos gráficos estão em **inglês**, pois os códigos do EPIDEMIC estão no **padrão internacional**.



O EPIDEMIC está disponível em: [www.EpidemicCode.org](http://www.EpidemicCode.org).

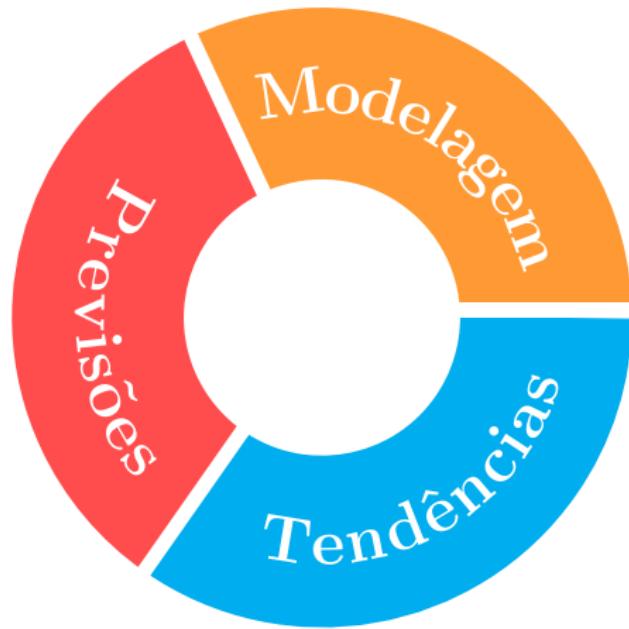
# Filosofia do EPIDEMIC

- Contribuir, de forma didática, para a compreensão de algumas das principais ferramentas de previsão epidemiológica.
- Disponibilizar, de maneira acessível, códigos computacionais para modelagem epidemiológica, seguindo os princípios de transparência, reproduzibilidade e validade.
- Fornecer um ambiente customizável em que o usuário possa ajustar os códigos de acordo com a sua situação-problema de interesse.
- Reforçar a importância da ciência no combate a uma epidemia.



Towards data science. Transparency, Reproducibility, and Validity of COVID-19 Projection Models, <[shorturl.at/IuK49](https://shorturl.at/IuK49)>.  
Acesso: 30/06/2020.

# Módulos do EPIDEMIC



# Módulos do EPIDEMIC

- **MODELAGEM:** modelos compartmentais descritos por equações diferenciais para simular a dinâmica de populações durante uma epidemia;
- **TENDÊNCIAS:** análise gráfica de tendências sobre o comportamento dinâmico de uma epidemia e seu progresso;
- **PREVISÕES:** regressor estatístico que utiliza dados para realizar previsões de curto prazo sobre quantidades de interesse.

## Software

O código EPIDEMIC foi elaborado no *software* livre GNU Octave<sup>©</sup>.



O EPIDEMIC também é compatível com a interface do *software* proprietário MATLAB<sup>©</sup>.

# Sumário

1 Modelagem

2 Tendências

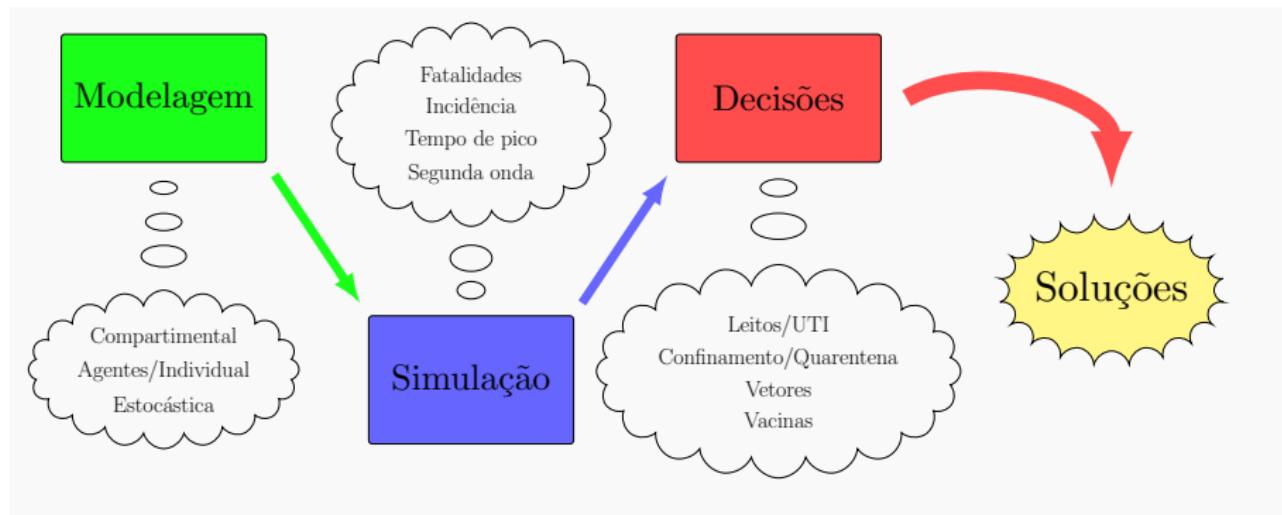
3 Previsões

4 Agradecimentos

5 Referências

# Modelagem

# A modelagem epidemiológica



## Modelos compartimentais

Num modelo compartmental a população de interesse é dividida em **compartimentos** de acordo com sua relação com a doença no tempo  $t$ : suscetíveis, infectados, hospitalizados, em quarentena, confinados etc.

O modelo compartmental mais simples é o **SIR**.

- **Suscetíveis**: podem se tornar infectados.
- **Infectados**: carregam o patógeno; podem se recuperar.
- **Recuperados**: se livraram do patógeno.



As setas ditam o fluxo de indivíduos de um compartimento para outro.

## Modelos compartmentais

Modelos compartmentais **determinísticos**: com um estado anterior pode-se prever todo o progresso da dinâmica.

Considerando  $N = S + I + R$  o número total da população, o número de indivíduos em cada compartimento no tempo  $t$  é determinado pela solução de sistemas de equações diferenciais.

$$\frac{dS}{dt} = -\beta S \frac{I}{N}, \quad \frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I.$$

- Os fluxos são proporcionais ao **tamanho dos compartimentos**.
- Transmissão ocorre de **I** para **S** a uma taxa  $\beta$ .
- Indivíduos em **I** se recuperam a uma taxa  $\gamma$ .

MODELAGEM → SIMULAR EPIDEMIAS → ANALISAR CARACTERÍSTICAS

# Modelos compartmentais: vantagens

- **Simples**: resumem a dinâmica como a ação de grupos (baixa dimensão).
- **Flexíveis**: podem ser adicionados
  - mais compartimentos;
  - mais de uma população;
  - classes por fatores socio-econômicos (e.g.: idade, renda);
  - efeitos climáticos/periódicos;
  - efeitos espaciais;
  - variáveis estocásticas.
- **Intuitivos**: parâmetros e número básico de reprodução bem claros.

# O que é o $\mathcal{R}_0$ e qual a sua importância?

$\mathcal{R}_0$ : número básico de reprodução.

**Definição clássica:** “o número médio de infecções secundárias produzidas quando um único indivíduo infectado é introduzido em uma população de hospedeiros onde todos são suscetíveis”.

**Na modelagem matemática:** número adimensional que determina se ocorrerá ou não uma epidemia. É um parâmetro do modelo que delimita o ponto de *bifurcação* do sistema epidêmico.

O  $\mathcal{R}_0$  de um modelo não necessariamente replica a definição clássica, porque depende das **hipóteses** do modelo.



# O que é o $\mathcal{R}_0$ e qual a sua importância?

Quando há um infectado apenas tem-se  $N \approx S$ , donde

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I \quad \rightarrow \quad \frac{dI}{dt} = (\beta - \gamma) I.$$

Nesse cenário, a curva de infectados segue uma exponencial da forma

$$I(t) = \exp(\gamma(\mathcal{R}_0 - 1)t), \quad \mathcal{R}_0 = \frac{\beta}{\gamma}.$$

- Se  $\mathcal{R}_0 < 1 \rightarrow$  **Não há epidemia.**
- Se  $\mathcal{R}_0 > 1 \rightarrow$  **Há epidemia.**

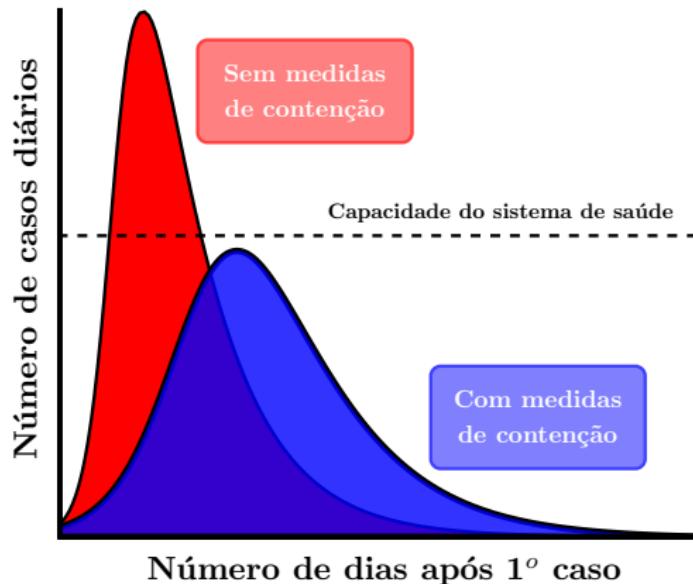
Em casos de epidemias:  $\mathcal{R}_0$  menores  $\rightarrow I(t)$  cresce mais devagar

A curva é “achatada”!

# Por que “achatar” a curva de infectados?

“Achatar” a curva significa atrasar a disseminação do patógeno.

Curva muito acentuada → colapso do sistema de saúde → **mais mortes**.



# Limitações da modelagem

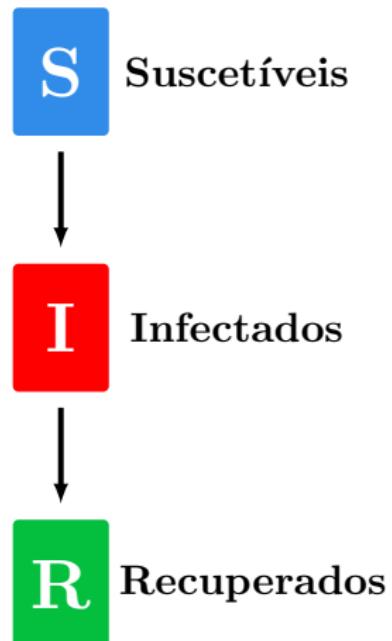
- Qualquer modelo é uma caricatura da realidade.
- Modelos compartmentais não consideram efeitos individuais, assumem por hipótese que a população é homogênea.
- Modelos precisam de dados para serem validados.
- Modelos muito simples têm dificuldade em reproduzir a realidade.
- Modelos muito complexos são de difícil utilização.

# Modelo SIR

- Uma população de tamanho constante  $N$  é dividida em **3 compartimentos**:

$$N = S(t) + I(t) + R(t).$$

- **Suscetíveis**: podem se tornar infectados.
- **Infectados**: carregam o patógeno; infectam suscetíveis; podem se recuperar.
- **Recuperados**: se livraram do patógeno; não transmitem; se tornaram imunes.
- Transmissão: contato direto entre **I** e **S**.



# Dinâmica do modelo SIR

- **Problema de Valor Inicial**, cujas EDOs são:

$$\frac{dS}{dt} = -\beta S \frac{I}{N} \quad (\text{taxa de suscetíveis}),$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I \quad (\text{taxa de infectados}),$$

$$\frac{dR}{dt} = \gamma I \quad (\text{taxa de recuperados}).$$

- Condições iniciais  $(S(0), I(0), R(0))$   $\xrightarrow[\text{das EDOs}]{\text{integração}} (S(t), I(t), R(t))$ .
- Séries temporais  $(S(t), I(t), R(t))$ : descrevem quantos indivíduos estão em cada compartimento no tempo  $t$ .

# Parâmetros do modelo SIR



- $N$ : total da população (indivíduos).
- $\beta$ : taxa de transmissão ( $\text{tempo}^{-1}$ ).
- $\gamma$ : taxa de recuperação ( $\text{tempo}^{-1}$ ).

## Número Básico de Reprodução — SIR

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

# Força de infecção $\lambda$

- Em geral tem-se  $\frac{dS}{dt} = -\lambda(t) S$ .
- Dependente da frequência:  $\lambda(t) = -\beta \frac{I(t)}{N}$ .
  - $\beta$ : taxa de transmissão constante.
  - $I(t)/N$ : probabilidade de contato com um indivíduo infectado.
- Outras modelagens existem:
  - $\beta$  dependente do tamanho populacional  $N$ ;
  - $\beta$  dependente do tempo  $t$ .



M. Begon, M. Bennett, R.G. Bowers, et al., A clarification of transmission terms in host-microparasite models: numbers, densities and areas, *Epidemiol. Infect.*, 129, 2002. DOI: <https://doi.org/10.1017/S0950268802007148>

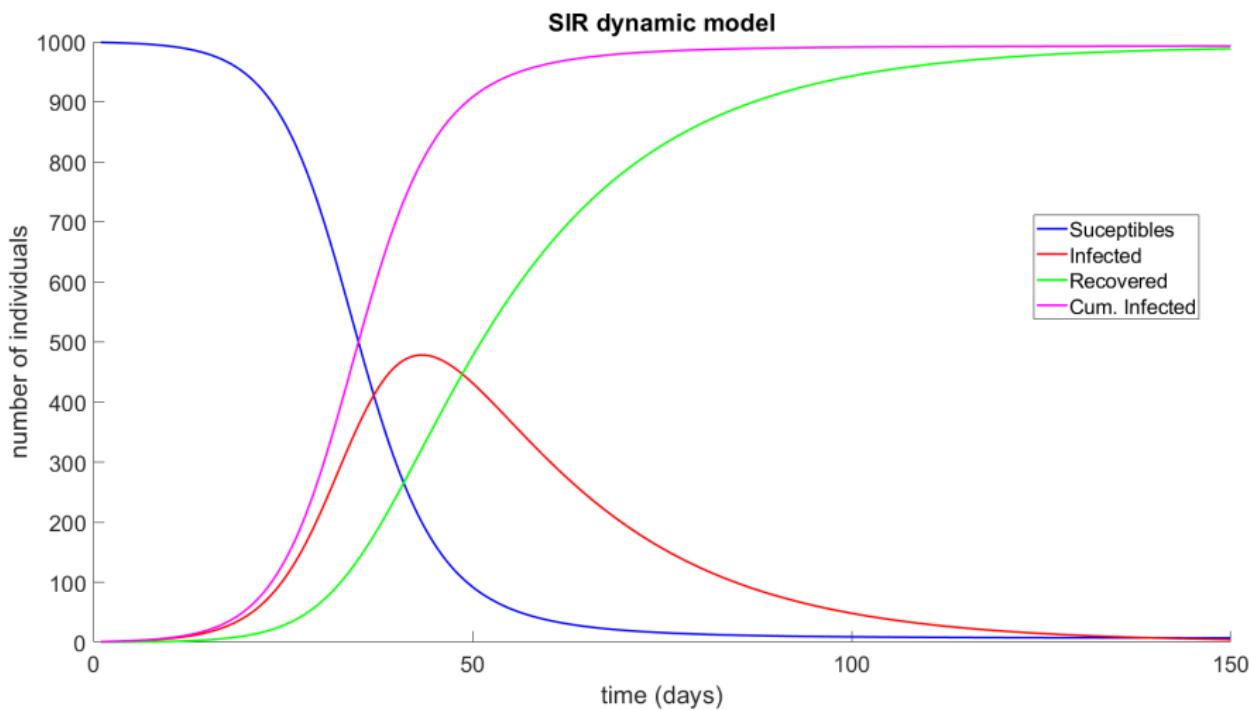
## Cumulativo de Infectados

- O compartimento  $I(t)$  informa quantos infectados existem no tempo  $t$ .
- O **cumulativo de infectados**  $C(t)$  informa quantos infectados existem até o tempo  $t$ .
- $C(t)$  pode ser calculado durante a integração, usando a parcela de entrada em  $I$ . Para o modelo SIR:

$$\frac{dC}{dt} = \beta S \frac{I}{N}.$$

- O **cumulativo de infectados** possibilita calcular medidas comuns da epidemiologia (**prevalência**) → comparação com dados.

# Séries temporais - Caso típico



## Séries temporais - Caso típico

```
-----  
+++++ SIR model +++++  
-----  
* population      = 1000  
  (individuals)  
* transmission rate = 0.25  
  (days^-1)  
* recovery rate    = 0.05  
  (days^-1)  
* R_nought        = 5  
  (dimensionless)  
-----
```

- $S$ : decaimento. O surto termina quando  $S$  estabiliza (“muito baixo”).
- $I$ : crescimento exponencial “ao início”.
- $R$ : compartimento de acumulação. Não interfere na dinâmica.
- **Cumulativo de infectados:** informa o número acumulado de infectados até o tempo  $t$ . Não interfere na dinâmica.

## Novos Casos

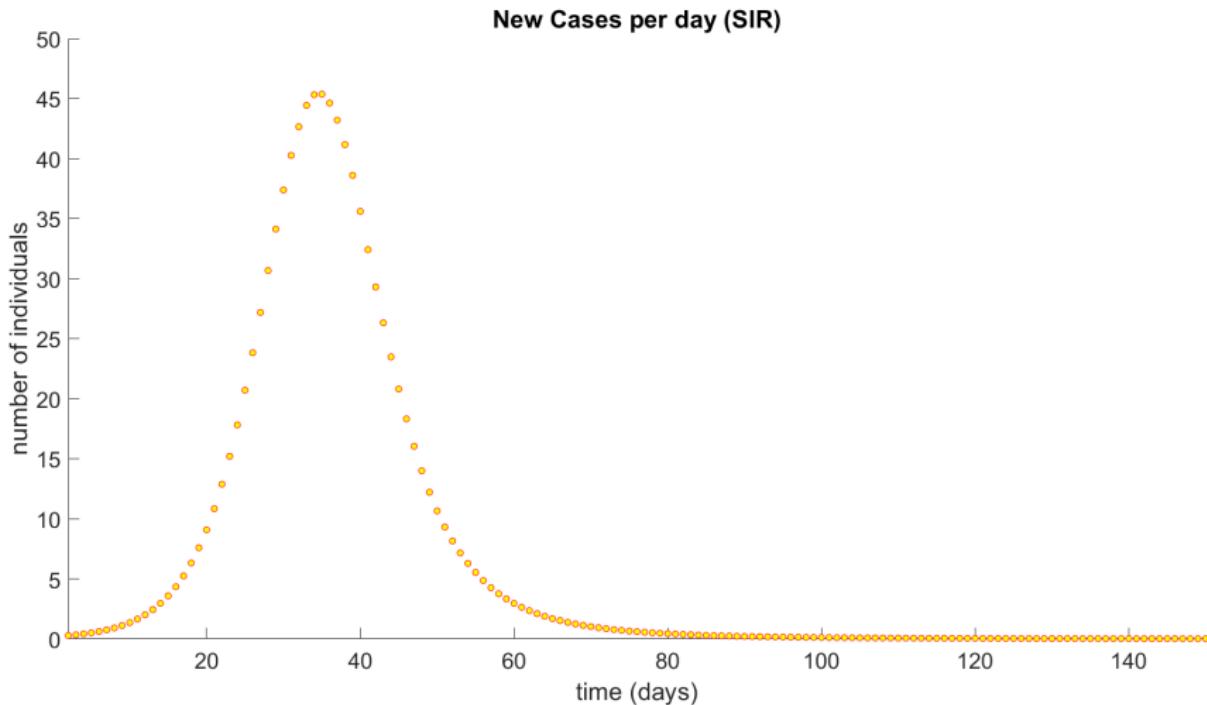
- A quantidade de novos casos por dia pode ser obtida a partir do **cumulativo de infectados**:

$$\text{NewCases}(t_i) = C(t_i) - C(t_{i-1}),$$

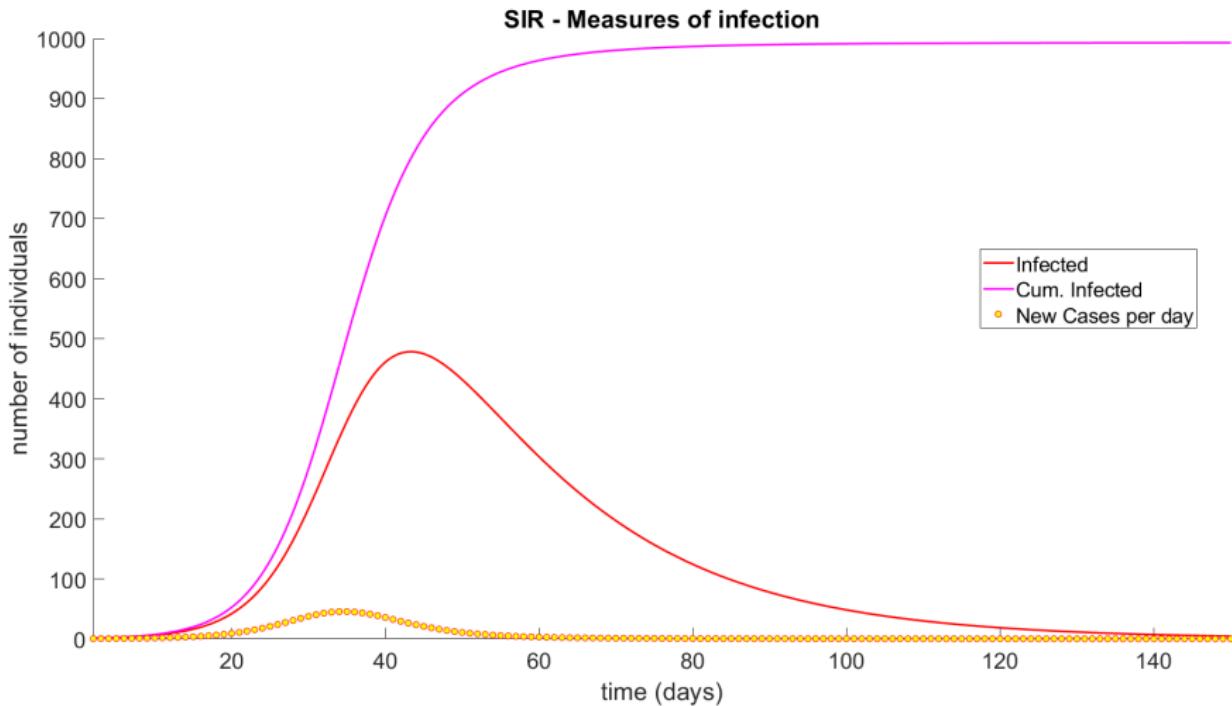
onde  $t_i$  é o instante selecionado para representar um certo dia.

- **NewCases** depende do intervalo: por dia, por semana, por mês, etc.
- **NewCases** → epidemiologia (**incidência**) → comparação com dados.
- Medidas de infecção (SIR):
  - **$I(t)$ , atual** número de infectados no tempo  $t$ ;
  - **$C(t)$ , total** de infectados **até** o tempo  $t$ ;
  - **$\text{NewCases}(t_i)$ , novos** infectados em  $t_i$  **em relação** a  $t_{i-1}$ .

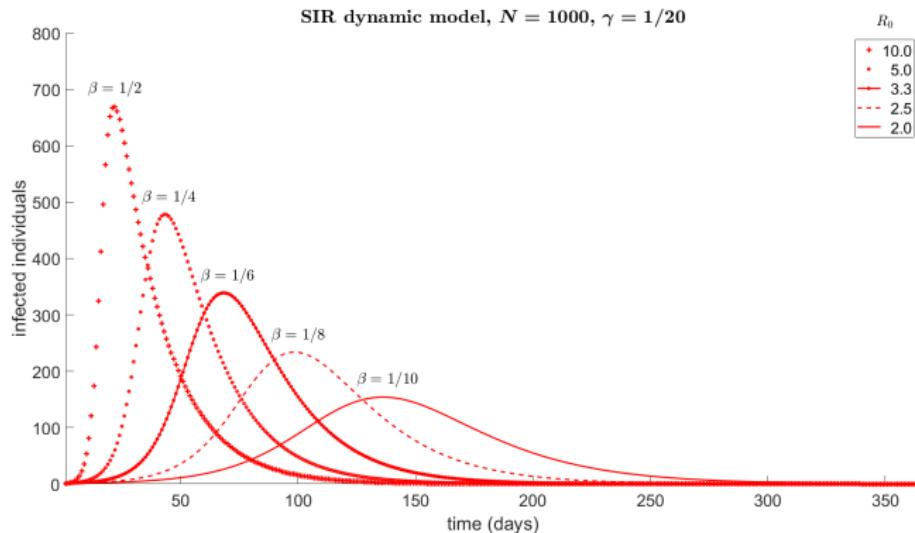
# Novos Casos



I × C × NewCases

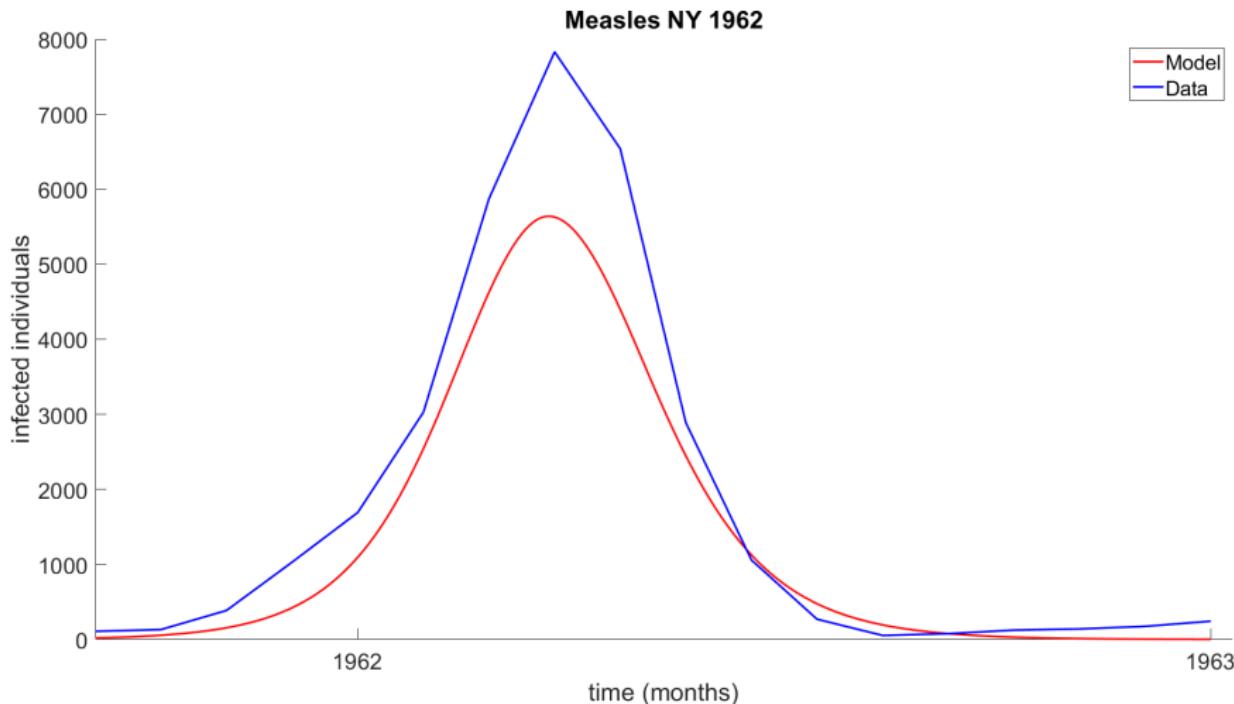


# $\mathcal{R}_0$ e controle



- “Ao início”,  $I(t) = \exp(\gamma(\mathcal{R}_0 - 1)t)$ .
- Reduzir  $\mathcal{R}_0 \rightarrow$  “achatar a curva”.
- Controle: Reduzir transmissão  $\rightarrow$  Reduzir  $\beta$ .

# Exemplo prático: Sarampo, Nova Iorque, 1962



F.Brauer, P. van den Driessche and J. Wu, **Mathematical Epidemiology**, Springer-Verlag, Berlin, 2008, p.5-11.  
DOI: <https://doi.org/10.1007/978-3-540-78911-6>

# Exemplo prático SIR: Sarampo, Nova Iorque, 1962

```
-----  
+++++++ SIR model ++++++  
-----  
* population      = 7781984  
  (individuals)  
* transmission rate = 3.6  
  (days^-1)  
* recovery rate    = 0.2  
  (days^-1)  
* R_nought        = 18  
  (dimensionless)  
* S0              = 0.065*N  
  (individuals)  
* I0              = 123*5/30  
  (individuals)  
-----
```

- Modelo simples → não reproduz perfeitamente os dados;
- Reprodução qualitativa do surto: tempo de pico, taxa de crescimento.



F.Brauer, P. van den Driessche and J. Wu, **Mathematical Epidemiology**, Springer-Verlag, Berlin, 2008, p.5-11.  
DOI: <https://doi.org/10.1007/978-3-540-78911-6>

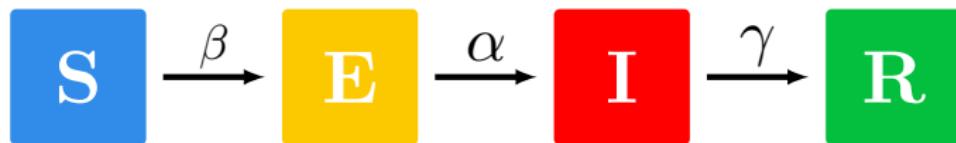
# Limitações do modelo SIR

- Não considera **incubação/latência**.
- Não considera **mortalidade** dos hospedeiros.
- Não considera **medidas de controle**: vacinas, hospitalização, quarentena, etc.
- Não considera fatores **espaciais**: assume uma população distribuída de forma homogênea.
- Não considera diferenças em **idade** ou **sexo biológico** dos hospedeiros.
- Não considera **migrações** nem efeitos **climáticos**.
- Assume apenas **uma via de transmissão**.
- Assume que todos os infectados adquirem **plena imunidade**.

# Modelo SEIR

- **4 compartimentos:** suscetíveis (**S**), expostos (**E**), infecciosos (**I**) e recuperados (**R**).
- Uma população de tamanho constante  $N$ :

$$N = S(t) + E(t) + I(t) + R(t)$$



# Modelo SEIR

- **Suscetíveis**: podem ser infectados, tornando-se expostos.
- **Expostos**: carregam o patógeno mas ainda não podem transmiti-lo; eventualmente tornam-se infecciosos.
- **Infecciosos**: carregam o patógeno e infectam suscetíveis; podem se recuperar.
- **Recuperados**: se livraram do patógeno e não transmitem; se tornaram imunes.
- Transmissão: contato direto de *I* com *S*.

## Dinâmica do modelo SEIR

- **Problema de Valor Inicial**, cujas EDOs são:

$$\frac{dS}{dt} = -\beta S \frac{I}{N} \quad (\text{taxa de suscetíveis}),$$

$$\frac{dE}{dt} = \beta S \frac{I}{N} - \alpha E \quad (\text{taxa de expostos}),$$

$$\frac{dI}{dt} = \alpha E - \gamma I \quad (\text{taxa de infectados}),$$

$$\frac{dR}{dt} = \gamma I \quad (\text{taxa de recuperados}).$$

- Condições iniciais  $\xrightarrow[\text{das EDOs}]{\text{integração}}$  séries temporais.

# Parâmetros do modelo SEIR

- $N$ : total da população (indivíduos).
- $\beta$ : taxa de transmissão ( $\text{tempo}^{-1}$ ).
- $\alpha$ : taxa de latência ( $\text{tempo}^{-1}$ ).
- $\gamma$ : taxa de recuperação ( $\text{tempo}^{-1}$ ).

## Número Básico de Reprodução — SEIR

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

# Expostos

**Momento de Exposição:** quando o patógeno encontra um suscetível.

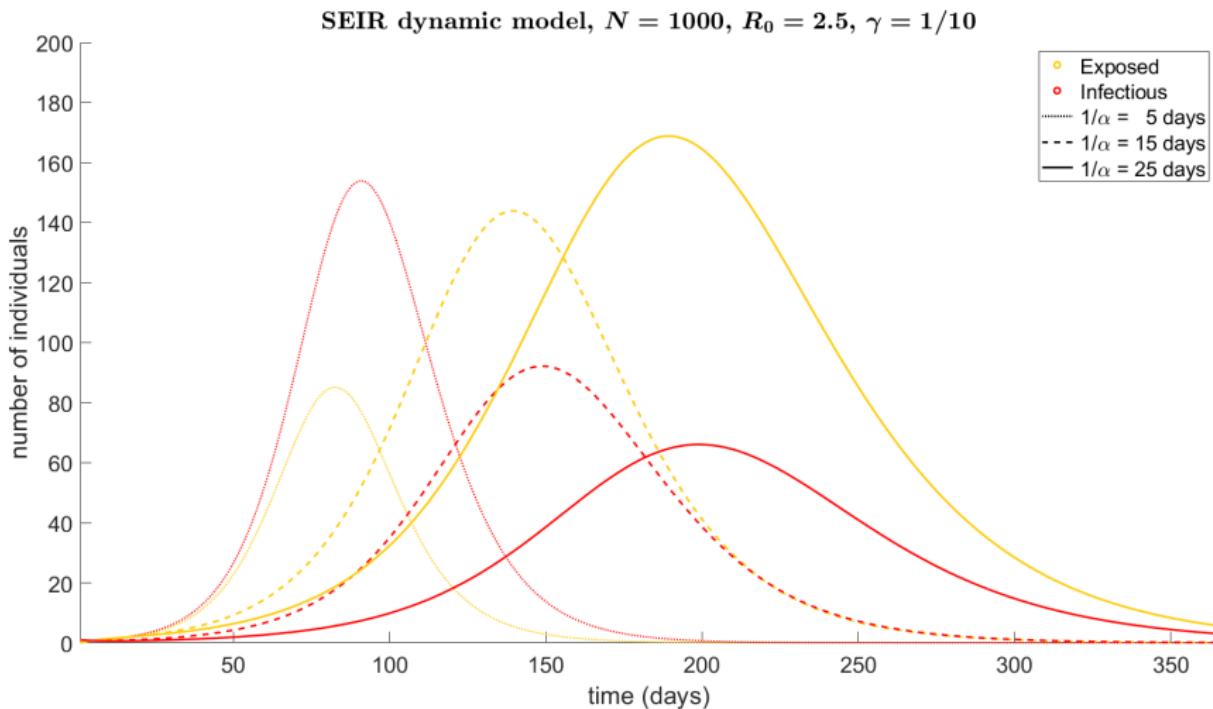
**Atrasos** (*time delays*) referentes ao momento de exposição.

- Período de incubação: atraso até apresentar sintomas.
- Período de latência: atraso até poder transmitir (**infeccioso**).

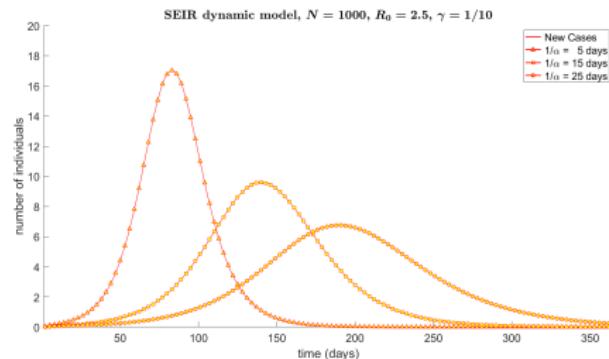
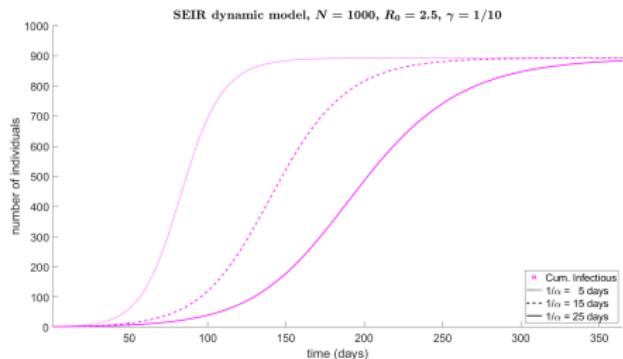
O comportamento **expostos** modela a latência com o parâmetro  $\alpha$ .

- **Outras modelagens existem!** Exemplo: EDOs com atraso,  $I(t - \tau)$ .
- **Limitação:** dados dependem do surgimento dos sintomas (incubação).

# Taxa de latêncio $\alpha$ e atraso



# Taxa de latência $\alpha$ e atraso



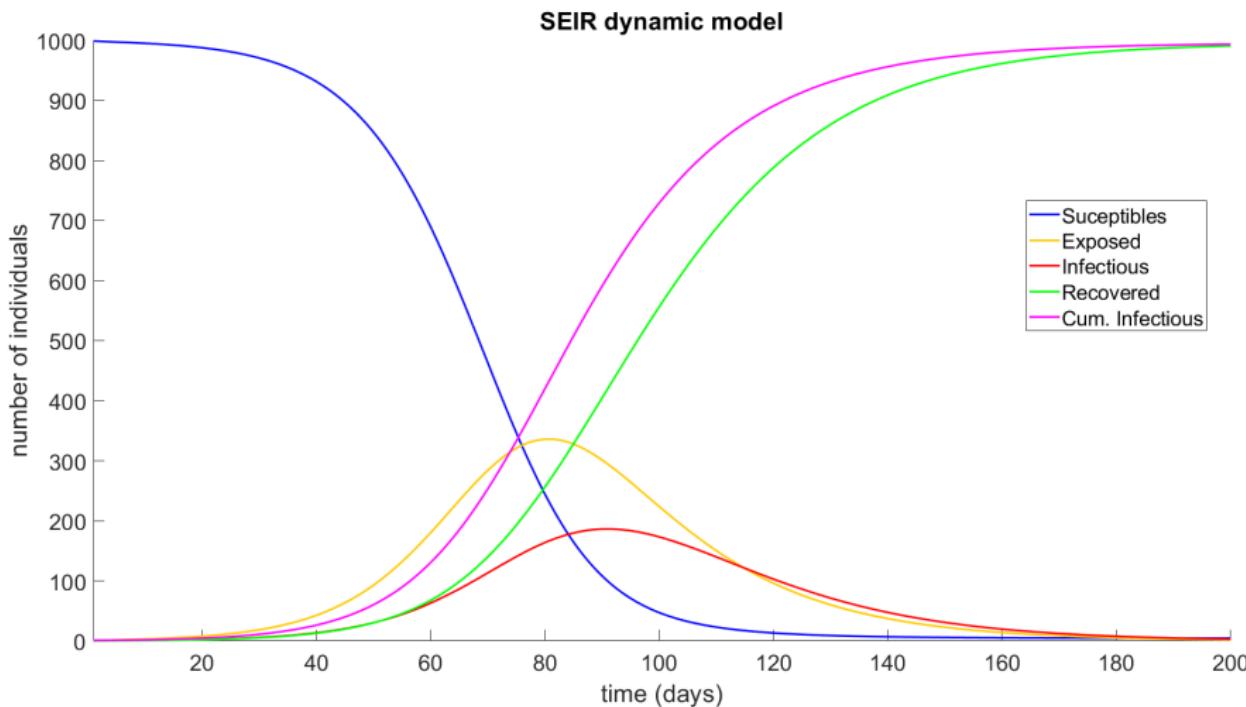
- O atraso da latência afeta o número de infectados **simultâneos**:
  - ao final, mesmo total de infectados ( $C$ );
  - não afeta transmissão → mesmo  $R_0$ ;
  - baixa latência → muitos **infecciosos em um mesmo  $t$**  ( $I$ , *NewCases*).
- **Limitação:**  $E$  muito cheio → difícil detecção em **tempo real**.

# Cumulativo de Infecciosos

- Infectado: teve contato com o patógeno. No SEIR:  $E + I$ .
- Infeccioso: infectado capaz de transmitir. No SEIR:  $I$ .
- Comparação com dados → **cumulativo de infecciosos**:

$$\frac{dC}{dt} = \alpha E .$$

# Séries temporais



# Parâmetros

```
-----
+-----+ SEIR model +-----+
-----
* population      = 1000
  (individuals)
* transmission rate = 0.458
  (days^-1)
* latent rate      = 0.052632
  (days^-1)
* recovery rate     = 0.083333
  (days^-1)
* R_nought         = 5.496
  (dimensionless)
-----
```

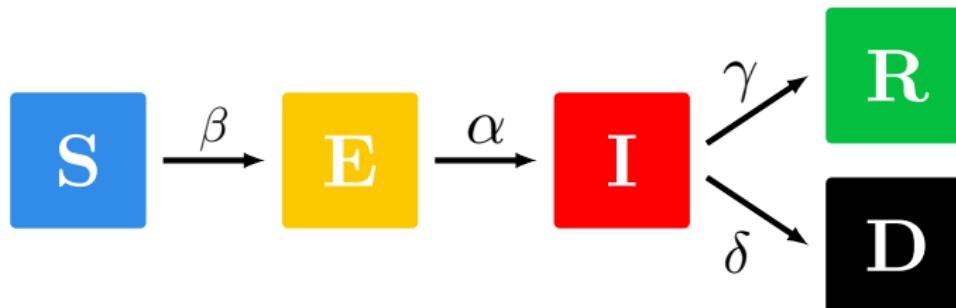
## Limitações do modelo SEIR

- Não considera **incubação/latência**.
- Não considera **mortalidade** dos hospedeiros.
- Não considera **medidas de controle**: vacinas, hospitalização, quarentena, etc.
- Não considera fatores **espaciais**: assume uma população distribuída de forma homogênea.
- Não considera diferenças em **idade** ou **sexo biológico** dos hospedeiros.
- Não considera **migrações** nem efeitos **climáticos**.
- Assume apenas **uma via de transmissão**.
- Assume que todos os infectados adquirem **plena imunidade**.

# Modelo SEIRD

- **5 compartimentos:** suscetíveis (**S**), expostos (**E**), infecciosos (**I**), recuperados (**R**) e mortos (**D**<sup>1</sup>).
- Uma população de tamanho inicial  $N_0$ :

$$N_0 = S(t) + E(t) + I(t) + R(t) + D(t)$$



<sup>1</sup>Do inglês *deceased*.

# Modelo SEIRD

- **Suscetíveis**: podem ser infectados, tornando-se expostos.
- **Expostos**: carregam o patógeno mas ainda não podem transmiti-lo; eventualmente tornam-se infecciosos.
- **Infecciosos**: carregam o patógeno e infectam suscetíveis; podem se recuperar ou morrer devido ao patógeno.
- **Recuperados**: se livraram do patógeno e não transmitem; se tornaram imunes.
- **Mortos**: faleceram por causa do patógeno.
- Transmissão: contato direto entre **I** e **S**.

# Dinâmica do modelo SEIRD

- **Problema de Valor Inicial**, cujas EDOs são:

$$\frac{dS}{dt} = -\beta S \frac{I}{N}, \quad \frac{dE}{dt} = \beta S \frac{I}{N} - \alpha E,$$

$$\frac{dI}{dt} = \alpha E - (\gamma + \delta) I, \quad \frac{dR}{dt} = \gamma I,$$

$$\frac{dD}{dt} = \delta I, \quad \text{onde } N(t) = N_0 - D(t).$$

- Condições iniciais  $\xrightarrow[\text{das EDOs}]{\text{integração}}$  séries temporais.

# Parâmetros do modelo SEIRD

- $N_0$ : população inicial (indivíduos).
- $\beta$ : taxa de transmissão ( $\text{tempo}^{-1}$ ).
- $\alpha$ : taxa de latência ( $\text{tempo}^{-1}$ ).
- $\gamma$ : taxa de recuperação ( $\text{tempo}^{-1}$ ).
- $\delta$ : taxa de mortalidade ( $\text{tempo}^{-1}$ ).

## Número Básico de Reprodução — SEIRD

$$\mathcal{R}_0 = \frac{\beta}{\gamma + \delta}$$

## Mortos

O compartimento **mortos** acumula quantos faleceram até o tempo  $t$  devido ao patógeno.

- População inicial  $N_0$ : constante; distribuída entre os compartimentos.

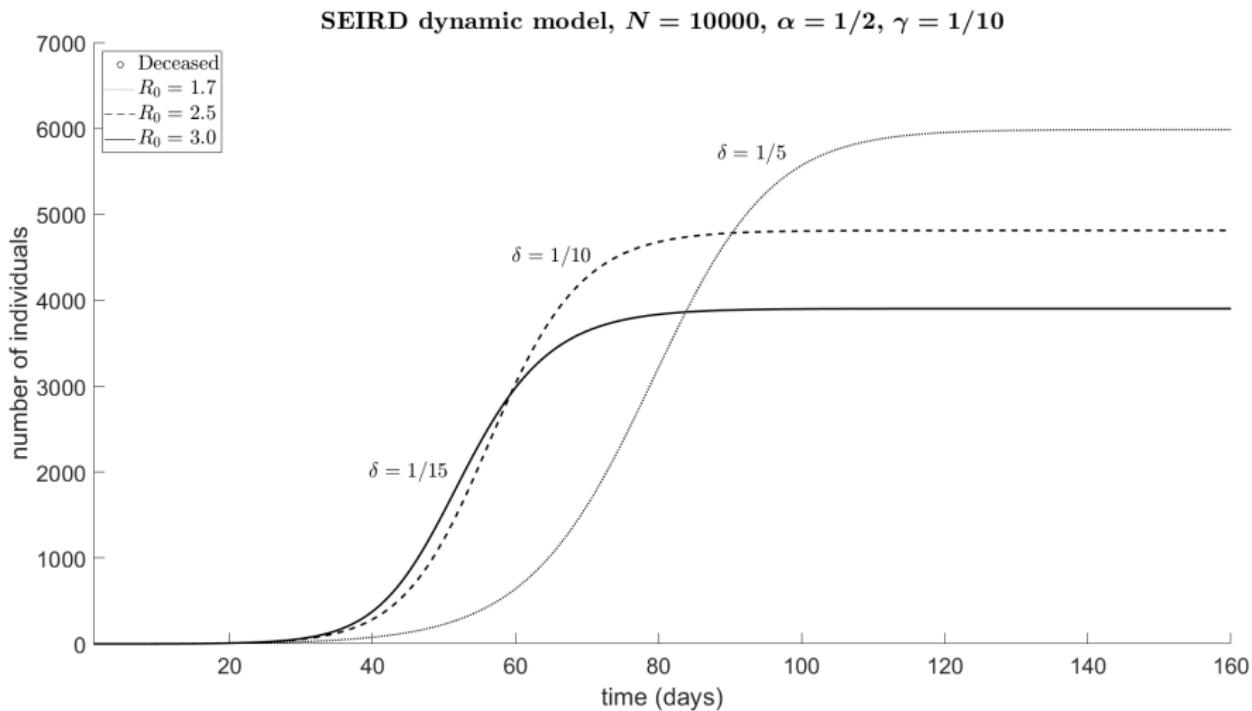
$$N_0 = S(t) + E(t) + I(t) + R(t) + D(t).$$

- População viva  $N$ : relevante para o termo da força de infecção.

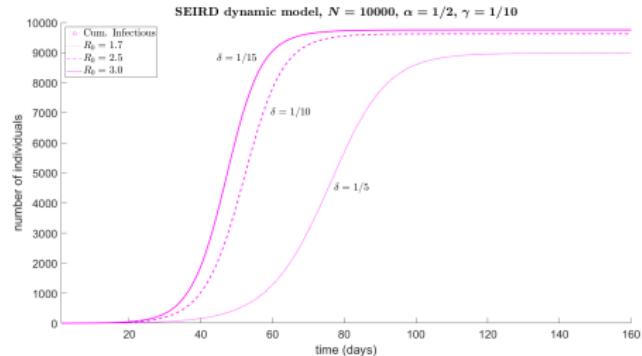
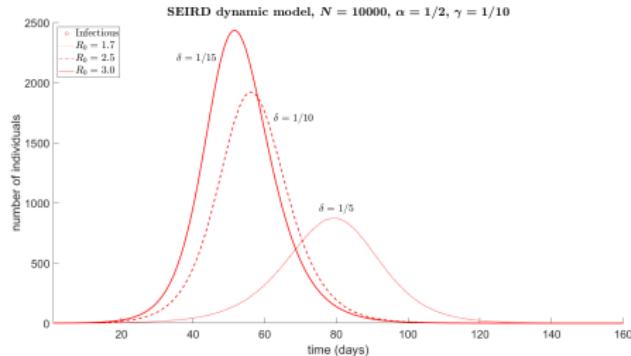
$$N(t) = S(t) + E(t) + I(t) + R(t) \rightarrow N(t) = N_0 - D(t).$$

- Em geral, medidas epidemiológicas referentes aos **mortos** são mais abundantes (**letalidade, mortalidade**) → comparação com dados.

# Taxa de mortalidade $\delta$

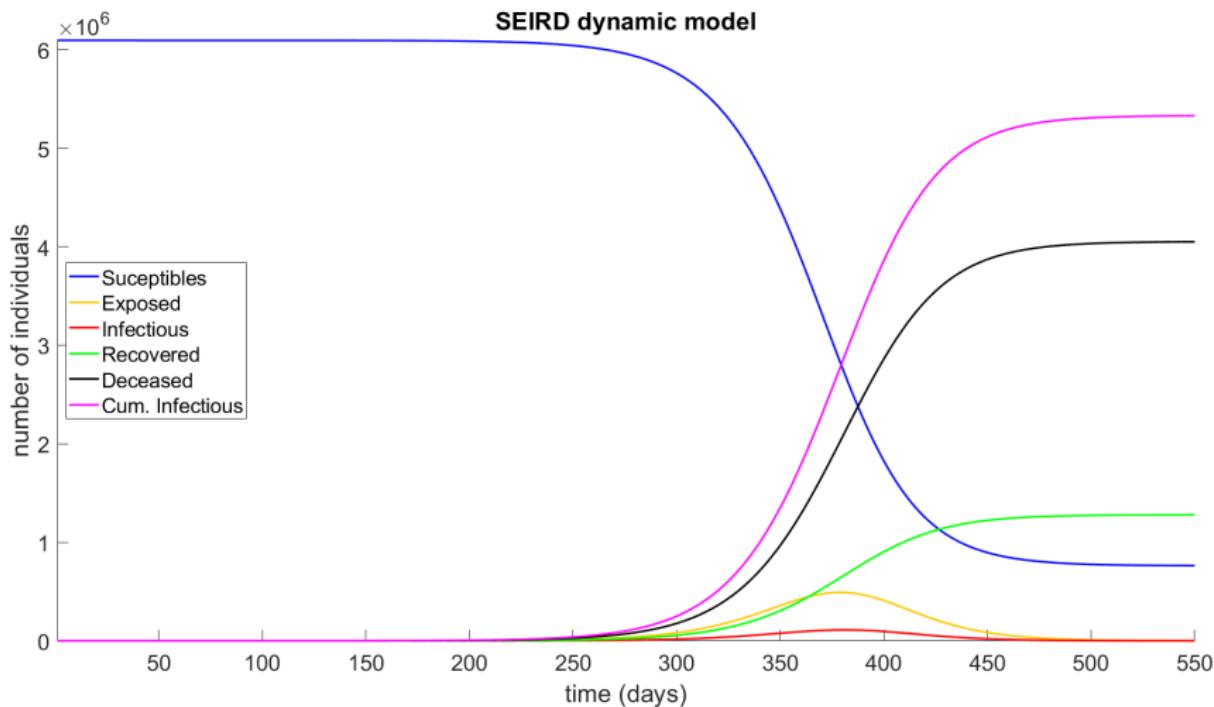


# Taxa de mortalidade $\delta$



- A mortalidade importa para a transmissão:
  - ao final, impacta o total de infectados  $C$ ;
  - muito letal  $\rightarrow$  reduz infecciosos vivos **em um mesmo**  $t$  ( $I$ ,  $NewCases$ );
  - alta taxa de mortalidade  $\rightarrow$  menor tempo médio em  $I \rightarrow$  menor  $R_0$ ;

# Séries temporais



# Parâmetros

```
-----  
+++++ SEIRD model +++++  
-----  
* initial population = 6092000  
  (individuals)  
* transmission rate   = 0.7124  
  (days^-1)  
* latent rate         = 0.11111  
  (days^-1)  
* recovery rate       = 0.11832  
  (days^-1)  
* death rate          = 0.37468  
  (days^-1)  
* R_nought            = 1.445  
  (dimensionless)  
-----
```

Alguns parâmetros provindos da epidemia de Ebola de 2014 em Serra Leoa.



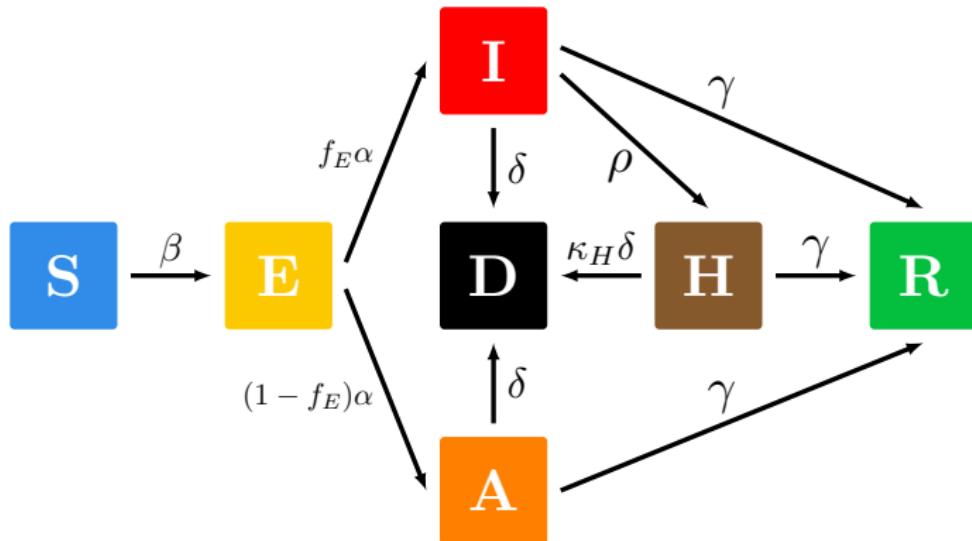
P. Diaz, P. Constantine, K. Kalmbach, E. Jones and S. Pankavichm. **A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation**, Applied Mathematics and Computation, 324(1): 141-155, 2018.  
DOI: <https://doi.org/10.1016/j.amc.2017.11.039>

# Limitações do modelo SEIRD

- Não considera **incubação/latência**.
- Não considera **mortalidade dos hospedeiros**.
- Não considera **medidas de controle**: vacinas, hospitalização, quarentena, etc.
- Não considera fatores **espaciais**: assume uma população distribuída de forma homogênea.
- Não considera diferenças em **idade** ou **sexo biológico** dos hospedeiros.
- Não considera **migrações** nem efeitos **climáticos**.
- Assume apenas **uma via de transmissão**.
- Assume que todos os infectados adquirem **plena imunidade**.

# Modelo SEIAHRD

- 7 compartimentos: suscetíveis (S), expostos (E), infecciosos sintomáticos (I), recuperados (R), infecciosos assintomáticos (A), hospitalizados (H) e mortos (D)



# Modelo SEIAHRD

- Uma população de tamanho inicial  $N_0$ :

$$N_0 = S(t) + E(t) + I(t) + A(t) + H(t) + R(t) + D(t)$$

- **Suscetíveis**: podem ser infectados, tornando-se expostos.
- **Expostos**: carregam o patógeno mas ainda não podem transmiti-lo; eventualmente tornam-se infecciosos.
- **Infecciosos sintomáticos**: carregam o patógeno e infectam suscetíveis; apresentam sintomas; podem se recuperar, ser hospitalizados ou morrer devido ao patógeno.
- **Infecciosos assintomáticos**: carregam o patógeno e infectam suscetíveis; não apresentam sintomas; podem se recuperar ou morrer devido ao patógeno.

# Modelo SEIAHRD

- **Hospitalizados**: carregam o patógeno e infectam suscetíveis; podem se recuperar ou morrer devido ao patógeno.
- **Recuperados**: se livraram do patógeno e não transmitem; se tornaram imunes.
- **Mortos**: faleceram por causa do patógeno.
- Transmissão: contato direto de *S* com *I* ou *A* ou *H*.

## Dinâmica do modelo SEIAHRD

- **Problema de Valor Inicial**, cujas EDOs são:

$$\frac{dS}{dt} = -\beta S \frac{I + A + \epsilon_H H}{N},$$

$$\frac{dE}{dt} = \beta S \frac{I + A + \epsilon_H H}{N} - \alpha E,$$

$$\frac{dI}{dt} = f_E \alpha E - (\gamma + \rho + \delta) I,$$

$$\frac{dA}{dt} = (1 - f_E) \alpha E - (\gamma + \delta) A,$$

$$\frac{dH}{dt} = \rho I - (\gamma + \kappa_H \delta) H,$$

$$\frac{dR}{dt} = \gamma (I + A + H),$$

$$\frac{dD}{dt} = \delta (I + A + \kappa_H H), \quad \text{onde } N(t) = N_0 - D(t).$$

- Condições iniciais  $\xrightarrow[\text{das EDOs}]{\text{integração}}$  séries temporais.

# Parâmetros do modelo SEIAHRD

- $N_0$ : população inicial (indivíduos).
- $\beta$ : taxa de transmissão ( $\text{tempo}^{-1}$ ).
- $\epsilon_H$ : fator de infecciosidade durante hospitalização (adimensional).
- $\alpha$ : taxa de latência ( $\text{tempo}^{-1}$ ).
- $f_E$ : fração sintomática (adimensional).
- $\gamma$ : taxa de recuperação ( $\text{tempo}^{-1}$ ).
- $\rho$ : taxa de hospitalização ( $\text{tempo}^{-1}$ ).
- $\delta$ : taxa de mortalidade ( $\text{tempo}^{-1}$ ).
- $\kappa_H$ : fator de mortalidade durante hospitalização (adimensional).

## Sintomas

O comportamento **infecciosos assintomáticos** modela os indivíduos que são capazes de transmitir o patógeno mesmo **sem sintomas**.

- Parâmetro  $f_E$  ( $0 < f_E < 1$ ): fração de infecciosos **com sintomas**.
- **Limitação:** muitos assintomáticos → difícil detecção precisa.
- **Outras modelagens existem!**
  - Mais compartimentos: de acordo com tipo e nível dos sintomas.
  - Mais parâmetros: taxas de transmissão diferentes para **sintomáticos** e **assintomáticos**.

## Hospitalizados

O compartimento hospitalizados modela os indivíduos **infecciosos sintomáticos** que estão em **tratamento hospitalar**.

- Parâmetro  $\rho$ : modela o quanto os **sintomáticos** são hospitalizados.
- Parâmetro  $\epsilon_H$  ( $0 < \epsilon_H < 1$ ): modela a redução da transmissão do patógeno pelos **infecciosos** hospitalizados.
- Parâmetro  $\kappa_H$  ( $0 < \kappa_H < 1$ ): modela a redução da mortalidade dos **infecciosos** devido a estarem **hospitalizados**.
- **Outras modelagens existem!**
  - Mais compartimentos: quarentena, UTI, etc.
  - Mais parâmetros: taxa de diagnosticação, fração de teste positivos, etc.
- Hospitalizados → **leitos e testes** → comparação com dados.

## Número básico de reprodução e controle

A hospitalização de **infecciosos** é uma estratégia de controle da epidemia. Um número adimensional análogo a  $\mathcal{R}_0$  pode ser definido ao supor que a estratégia de controle esteve ativa desde que se teve  $N \approx S$ .

### Número de Reprodução com Controle — SEIAHRD

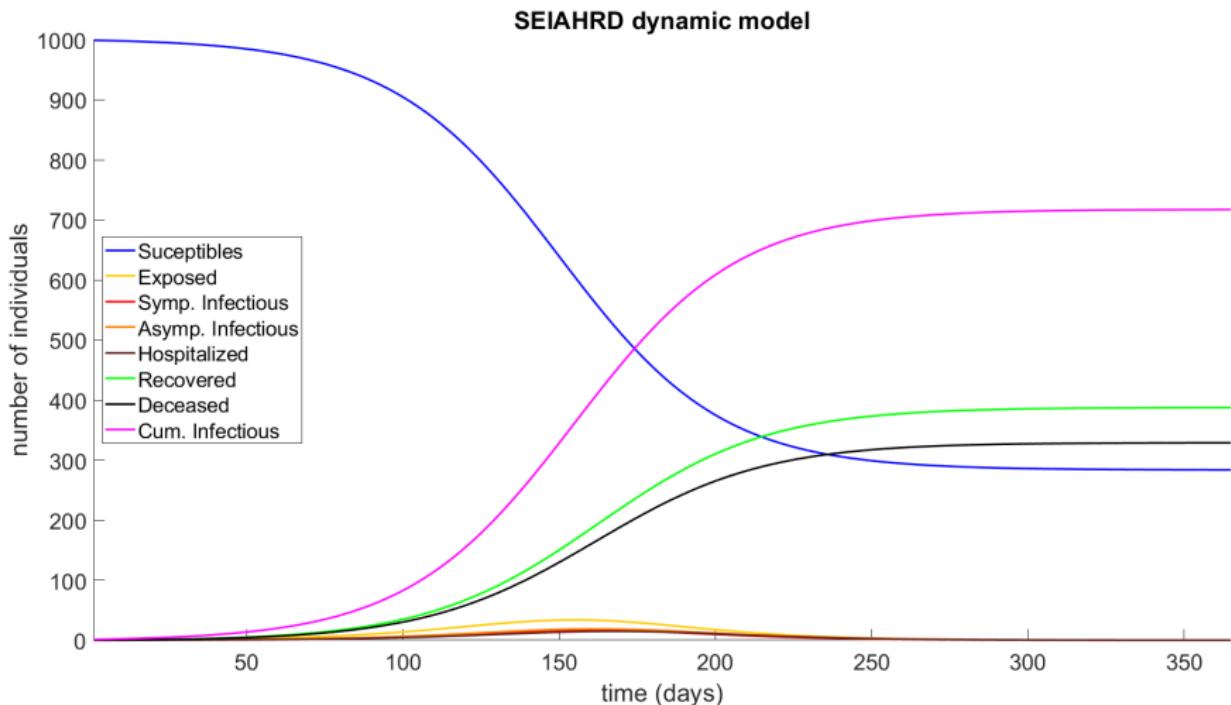
$$\mathcal{R}_C = \frac{f_E \beta}{\gamma + \rho + \delta} + \frac{(1 - f_E) \beta}{\gamma + \delta} + \frac{\rho f_E \epsilon_H \beta}{(\gamma + \rho + \delta)(\gamma + \kappa_H \delta)}$$

Na ausência ou atraso das estratégias de controle ( $\rho = \epsilon_H = \kappa_H = 0$ ), o modelo SEIAHRD possui um  $\mathcal{R}_0$  convencional.

### Número Básico de Reprodução — SEIAHRD

$$\mathcal{R}_0 = \frac{\beta}{\gamma + \delta}$$

## Séries temporais



# Parâmetros

```
-----  
+++++ SEIAHRD model +++++  
-----  
* initial population      = 1000  
  (individuals)  
* transmission rate       = 0.25  
  (days^-1)  
* hosp. infectivity       = 0.22  
  (dimensionless)  
* latent rate              = 0.2  
  (days^-1)  
* symptomatic fraction     = 0.6  
  (dimensionless)  
* recovery rate             = 0.071429  
  (days^-1)  
* hospitalization rate     = 0.1  
  (days^-1)  
* death rate                = 0.07  
  (days^-1)  
* hospitalization recovery = 0.55  
  (dimensionless)  
* R_nought                  = 3.5  
  (dimensionless)  
* R_control                 = 1.4527  
-----
```

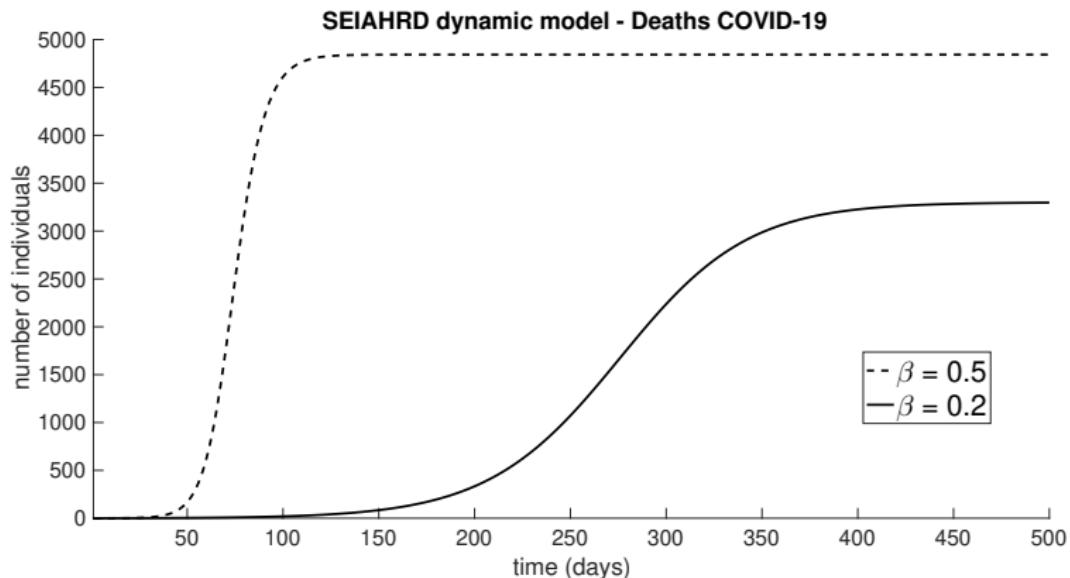
Alguns parâmetros usados na epidemia de Sars-CoV-2 de 2019 no Brasil.



W. Lyra, J. Nascimento, J. Belkhiria, L. de Almeida, P. P. Chrispim and I. Andrade. **COVID-19 pandemics modeling with SEIR(+CAQH), social distancing, and age stratification. The effect of vertical confinement and release in Brazil.**, medRxiv, 2020. DOI: <https://doi.org/10.1101/2020.04.09.20060053>.

# Comparativo de mortes para COVID-19 em relação ao $\beta$

- $N = 10.000$  indivíduos;
- Para  $\beta = 0,5 \Rightarrow \mathcal{R}_0 = 7$  e para  $\beta = 0,2 \Rightarrow \mathcal{R}_0 = 2,8$ .



## Limitações do modelo SEIAHRD

- Não considera **incubação/latência**.
- Não considera **mortalidade dos hospedeiros**.
- Não considera **medidas de controle**: vacinas, **hospitalização**, quarentena, etc.
- Não considera fatores **espaciais**: assume uma população distribuída de forma homogênea.
- Não considera diferenças em **idade** ou **sexo biológico** dos hospedeiros.
- Não considera **migrações** nem efeitos **climáticos**.
- Assume apenas **uma via de transmissão**.
- Assume que todos os infectados adquirem **plena imunidade**.

# Arquivos do código computacional

Para cada modelo, o EPIDEMIC utiliza dois arquivos. Exemplo, SIR:

- **[main\_SIR.m]** Arquivo principal. Define os parâmetros, condições iniciais e intervalo de tempo para a integração do sistema de EDOs. Também calcula os números de reprodução e plota os resultados das séries temporais.
- **[rhs\_SIR.m]** Define o sistema de EDOs utilizado pelo arquivo principal.

# Arquivo main: definição dos parâmetros

O usuário define os parâmetros:

```
% parameters and initial conditions [USER INPUT]
% _____
% population size (number of individuals)
N = 1000;

% transmission rate (days^-1)
beta = 1/4;

% recovery period (days)
Tgamma = 10;

% recovery rate (days^-1)
gamma = 1/Tgamma;
```

Por praticidade, alguns parâmetros são definidos através de seus inversos, por exemplo, `gamma`:

$$\gamma : \text{taxa de recuperação} = (\text{período de recuperação})^{-1}$$

## Arquivo main: condições iniciais

O usuário define as condições iniciais (CIs).

- As CIs são os valores dos compartimentos no tempo t0;
- Comentário: sugestão para CIs no início (ideal) da epidemia;
- Para computar o cumulativo de infectados, C0 também é requerido.

```
% initial conditions
%
% — Set the initial number of infected.
% — The number of susceptible will be the remaining population.
% — For an invasion scenario, set initial infected to 1.

R0 = 0;           % initial recovered   (number of individuals)
I0 = 1;           % initial infected    (number of individuals)
S0 = N-I0-R0;    % initial susceptible (number of individuals)

% initial cumulative infected (number of individuals)
C0 = I0;
%
```

## Arquivo main: $\mathcal{R}_0$ e display

O arquivo automaticamente calcula o número básico de reprodução. O  $\mathcal{R}_0$  e os parâmetros definidos são expostos na tela de comandos após um cabeçalho do EPIDEMIC.

```
% computing the basic reproduction number R_nought
R_nought = beta/gamma;

disp(' ')
disp('_____')
disp(' EPIDEMIC - Epidemiology Educational Code ')
disp(' by A. Cunha, E. Dantas, et al. ')
disp(' ')
disp(' This is an easy to run educational toolkit ')
disp(' for epidemiological analysis. ')
disp(' ')
disp(' www.EpidemicCode.org ')
disp('_____')
```

Em cada modelo, quantidades diversas que dependem dos parâmetros são computadas e expostas nessa parte.

## Arquivo main: integração

O usuário pode alterar os detalhes do processo de integração.

- Os vetores `param` e `IC` reúnem os parâmetros e condições iniciais para o arquivo `rhs`. Atente-se para a ordem nesses vetores.
- O vetor `tspan` define o intervalo de análise segundo o tempo inicial `t0`, tempo final `t1` e o passo `dt`.

```
% parameters vector
param = [N beta gamma];

% initial conditions vector
IC = [S0 I0 R0 C0];

% time interval of analysis
t0 = 1;                                % initial time (days)
t1 = 365;                               % final time   (days)
dt = 0.1;                                % time steps   (days)
tspan = t0:dt:t1;                         % interval of analysis
Ndt = length(tspan);                     % number of time steps
```

## Arquivo main: integração

O solver ode45 utiliza o arquivo **rhs** como uma função para avaliar as derivadas em diversos instantes.

- Parâmetros e condições iniciais são informados segundo `param` e `IC`.
- O método Runge-Kutta 4(5) é empregado pelo solver para a integração numérica no intervalo `tspan`.
- Saída: a matriz `y` fornece as séries temporais segundo o vetor `time`.

```
% ODE solver Runge-Kutta45
[time, y] = ode45(@(t,y)rhs_SIR(t,y,param),tspan,IC);

% time series
S = y(:,1); % susceptible (number of individuals)
I = y(:,2); % infected (number of individuals)
R = y(:,3); % recovered (number of individuals)
C = y(:,4); % cumulative infected (number of individuals)
```

Por praticidade, as séries temporais são separadas em vetores.

# Arquivo main: pós-processamento

A quantidade *NewCases* é computada para intervalos de dias.

```
% NewCases ( per day) computation
tu = 1; % time unit (days)
%
% --- tu/dt must be an integer
% --- tu = 1 defines a 'per day' computation
% --- For a 'per week' computation, use tu = 7 and t0 = 7
NewCases = zeros(floor((Ndt-1)/(tu/dt)),1);
for n = 1:length(NewCases)
    NewCases(n) = C((n)*tu/dt +1) - C((n-1)*tu/dt +1);
end
```

Gráficos são gerados com as séries temporais e *NewCases* (discreto).

```
% plot all compartments of SIR model
figure(1)
hold on
figS = plot(time,S,'DisplayName','Susceptibles','Color','b');
figI = plot(time,I,'DisplayName','Infected','Color','r');
figR = plot(time,R,'DisplayName','Recovered','Color','g');
figC = plot(time,C,'DisplayName','Cum. Inf.','Color','m');
hold off

% plot NewCases (per day) of SIR model
figure(2)
fig2 = scatter([1:length(NewCases)]+1,NewCases);
```

As demais linhas contêm detalhes para personalização das figuras.

## Arquivo rhs: entrada e saída

O arquivo define uma *function* para o GNU Octave\ MATLAB com as entradas:

- t: vetor de tempo;
- y: vetor de estados;
- param: conjunto de parâmetros.

```
function dydt = rhs_SIR(t, y, param)
```

A saída da *function* é o vetor dydt que informa as derivadas a cada tempo.

## Arquivo rhs: argumentos de entrada

Os argumentos t e y são organizados automaticamente pelo ode45 no arquivo principal, segundo:

- vetor tspan → t;
- vetor IC → y.

O argumento param é ajustado manualmente no arquivo principal para ser aberto no arquivo rhs. É importante se atentar para a ordem no vetor.

```
% model parameters: param = [N beta gamma]
N      = param(1); % population size    (number of individuals)
beta   = param(2); % transmission rate (days^-1)
gamma  = param(3); % recovery rate     (days^-1)
```

# Arquivo rhs: equações diferenciais

A parte principal do arquivo define as equações diferenciais.

```
% SIR dynamic model:  
%  
%     y = [S I R C]           is the state vector  
%     dydt = [dSdt dIdt dRdt dCdt] is the evolution law  
%  
% dSdt - rate of susceptible      (number of individuals/days)  
% dIdt - rate of infected        (number of individuals/days)  
% dRdt - rate of recovered       (number of individuals/days)  
% dCdt - rate of cumulative infected (number of individuals/days)  
  
[S I R C] = deal(y(1),y(2),y(3),y(4));  
  
dSdt = - beta*S.* (I/N);  
dIdt = beta*S.* (I/N) - gamma*I;  
dRdt = gamma*I;  
dCdt = beta*S.* (I/N);  
dydt = [dSdt; dIdt; dRdt; dCdt];
```

A ordem do vetor IC no arquivo principal é importante para o vetor de estados y e para a saída dydt da *function*.

# Personalização (SIR → SEIR)

Modelos compartimentais podem ser facilmente modificados.

Exemplo: SIR → SEIR

- **Altere as EDOs no rhs.** Adicione o compartimento *E*. Ajuste os termos de entrada (positivos) e saída (negativos) nos compartimentos.

```
[S I R C] = deal(y(1),y(2),y(3),y(4)); [S E I R C] = deal(y(1),y(2),y(3),y(4),y(5));
```

```
dSdt = - beta*S.* (I/N);  
dIdt = beta*S.* (I/N) - gamma*I;  
dRdt = gamma*I;  
dCdt = beta*S.* (I/N);  
dydt = [dSdt; dIdt; dRdt; dCdt];
```

```
dSdt = - beta*S.* (I/N);  
dEdt = beta*S.* (I/N) - alpha*E;  
dIdt = alpha*E - gamma*I;  
dRdt = gamma*I;  
dCdt = alpha*E;  
dydt = [dSdt; dEdt; dIdt; dRdt; dCdt];
```

Note que *C* é alterado para continuar registrando a entrada de *I*.

Certifique-se que a ordem no vetor *y* e *dydt* está coincidindo.

# Personalização (SIR → SEIR)

- Altere os parâmetros e CIs no main e rhs. Modifique os vetores param e IC no main. Certifique-se que a ordem nos vetores está coincidente entre os arquivos.

```
% parameters vector  
param = [N beta gamma];  
  
% initial conditions vector  
IC = [S0 I0 R0 C0];
```

```
% parameters vector  
param = [N beta alpha gamma];  
  
% initial conditions vector  
IC = [S0 E0 I0 R0 C0];
```

- Separe as séries temporais. Ajuste a separação das séries temporais de acordo com a ordem do vetor IC.

```
% time series  
S = y(:,1);  
I = y(:,2);  
R = y(:,3);  
C = y(:,4);
```

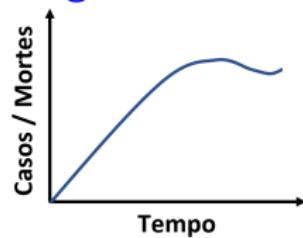
```
% time series  
S = y(:,1);  
E = y(:,2);  
I = y(:,3);  
R = y(:,4);  
C = y(:,5);
```

# Tendências

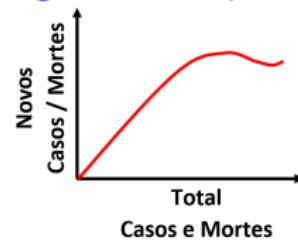
# Monitoramento

As análises de tendências em epidemiologia tem como objetivo acompanhar o comportamento de uma epidemia numa certa região geográfica (país, estado, cidade etc). No EPIDEMIC essas análises são realizadas utilizando duas estratégias básicas de visualização:

## Contágio e Mortalidade



## Progresso da epidemia



- Avalia a evolução temporal dos números de infectados e mortos. Avalia a incidência em função do tempo.

- Avalia a taxa de variação da epidemia relativamente ao totais de casos e de mortes.

# Especificidades dos gráficos

Cada gráfico de monitoramento gerado pelo código fornecerá diferentes tipos de análises. Por este motivo, compreender o contexto no qual o código será aplicado, bem como a representação gráfica gerada pelo mesmo, é de grande importância para obter a informação desejada.

## Análise temporal

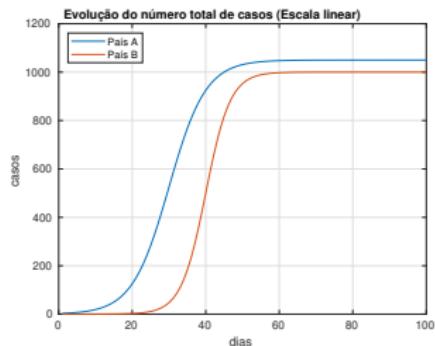
As análises gráficas em função do tempo são feitas em relação ao número de dias desde que uma quantidade “X” de casos ou mortes foi registrada. Assim, é possível realizar uma comparação temporal da incidência da doença entre diferentes locais mesmo que ela tenha ocorrido em período distintos.

## Incidência relativa

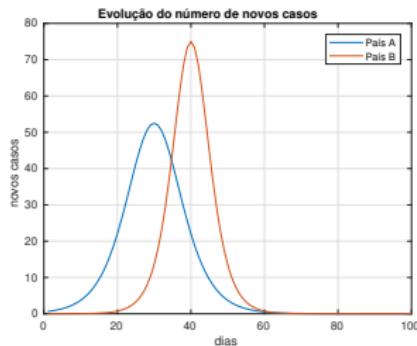
A avaliação da proporção de pessoas infectadas em um local de interesse é feita através da incidência relativa (IR) da doença. Utilizam-se as mesmas curvas da análise temporal, porém com o número de casos proporcional à população, por milhão (ou cem mil) de habitantes. Isso permite comparar o impacto da epidemia em cada região em termos relativos.

# Os gráfico e suas análises

Observando dois países hipotéticos após uma determinada epidemia; as curvas (à direita) desses países mostram que o país "B" teve um crescimento mais acentuado de casos em um curto período de tempo quando comparado ao país "A". Adicionalmente, as curvas geram a impressão de que o país "B" apresentou mais casos do que o país "A".



A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21<sup>a</sup> Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>



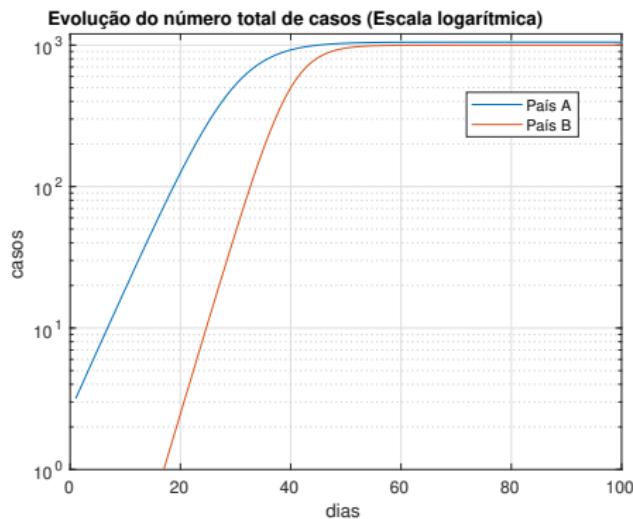
Contudo, pela análise do gráfico à esquerda, o país "B" apresentou um nº de casos ligeiramente menor do que o país "A". Logo, a informação desejada está intimamente relacionada ao tipo de gráfico analisado, neste caso, se deseja-se saber qual país tem mais casos, o gráfico com o valor acumulado é mais útil.

# Uso da escala logarítmica

Durante a epidemia, enquanto as curvas ainda estão sendo construídas, como saber se o pico está próximo?

Uma forma para obter indícios sobre o pico da epidemia é gerar gráficos utilizando **escala logarítmica**.

Na escala logarítmica o crescimento exponencial aparece como uma “reta” crescente. Assim, fica mais fácil observar quando a taxa de crescimento tem uma queda, conforme pode ser observado no gráfico ao lado.



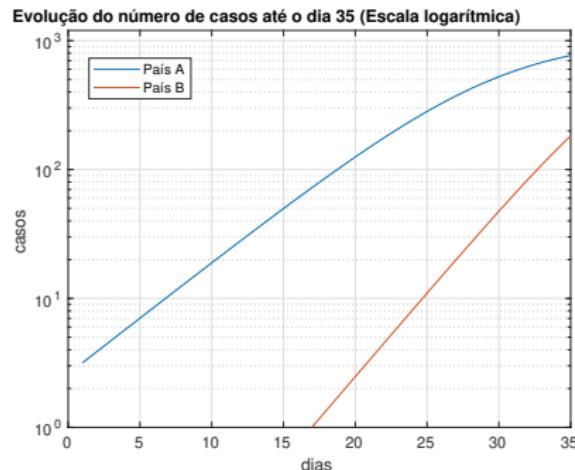
A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21<sup>a</sup> Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

# Uso da escala logarítmica

Durante a epidemia, enquanto as curvas ainda estão sendo construídas, como saber se o pico está próximo?

Observando o gráfico ao lado, dos cenários dos países “A” e “B” até o 35º dia de análise.

É possível verificar o comportamento exponencial das curvas na maior parte do tempo, porém a curva do país “A”, após o 25º dia, apresenta certa inclinação/curvatura. Fato que aponta para a estabilização dos casos, além de indicar que o pico da epidemia pode estar próximo.



\*Vale ressaltar que estas são curvas hipotéticas, bem-comportadas e geradas sem ruído.



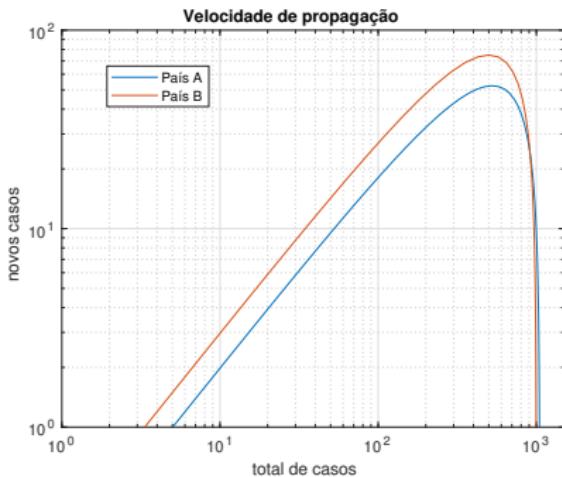
A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21ª Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

# Mudança no progresso da epidemia

Uma mudança no progresso da epidemia pode ser melhor avaliada no gráfico do número de novos casos em função do total de casos.

Neste gráfico o eixo horizontal corresponde ao total de casos, de modo que, o tempo fica implícito na avaliação.\*

É possível ver a mudança de comportamento em ambas as curvas, ou seja, a “queda”.



\*Não explicitar o tempo permite comparar o progresso da epidemia em diferentes locais em diferentes momentos.

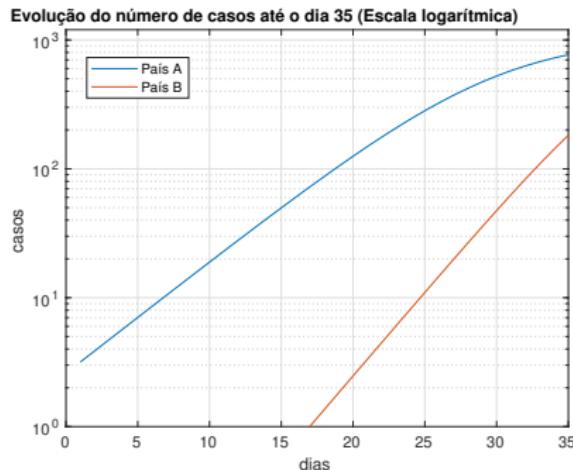


A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21<sup>a</sup> Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

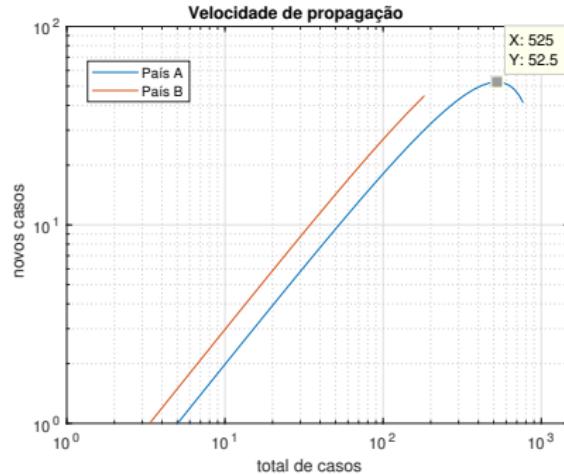
# Mudança no progresso da epidemia

Para representar o progresso da epidemia quando não se têm os dados completos, o gráfico com a **velocidade de propagação** é uma alternativa para a compreensão da evolução/tendência do surto através dos novos casos. O que não é possível observar no gráfico temporal.

## Gráfico Temporal



## Gráfico de "Velocidade"

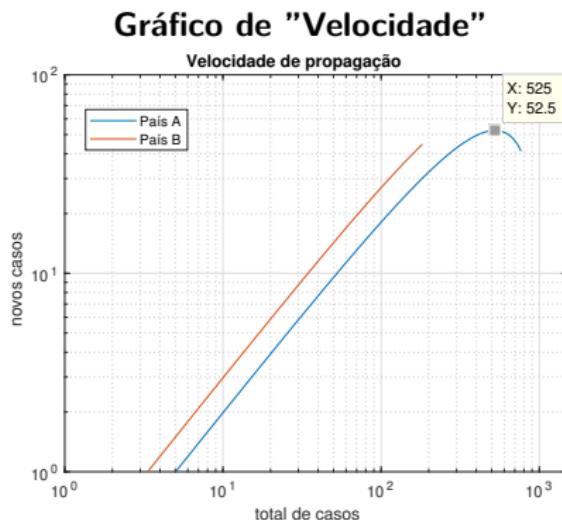


A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21<sup>a</sup> Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

# Mudança no progresso da epidemia

Para representar o progresso da epidemia quando não se têm os dados completos, o gráfico com a **velocidade de propagação** é uma alternativa para a compreensão da evolução/tendência do surto através dos novos casos.

No gráfico **velocidade de propagação** é possível observar a queda na curva do país “A”, cujo pico ocorre no 30º dia e corresponde a um total de 525 casos.



A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21ª Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

## Gráficos de monitoramento

Os tipos de gráficos gerados com o arquivo `epidemic_trends.m` para monitoramento são os seguintes:

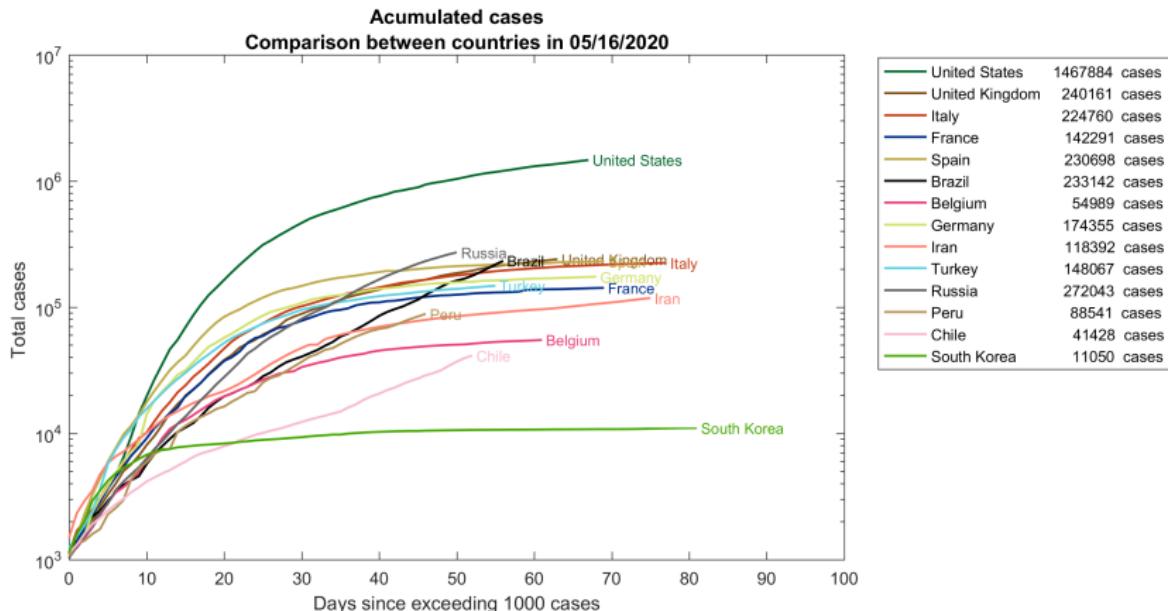
- Total de `casos/mortes` em relação ao tempo;
- Total de `casos/mortes` por 1M de habitantes em relação ao tempo;
- Novos `casos/mortes` por semana em relação ao tempo;
- Novos `casos/mortes` por semana por 1M de habitantes em relação ao tempo;
- Novos `casos/mortes` por semana `em relação ao total` de casos/mortes;

Contágio   Mortalidade   Progresso

## Exemplo: monitoramento da COVID-19

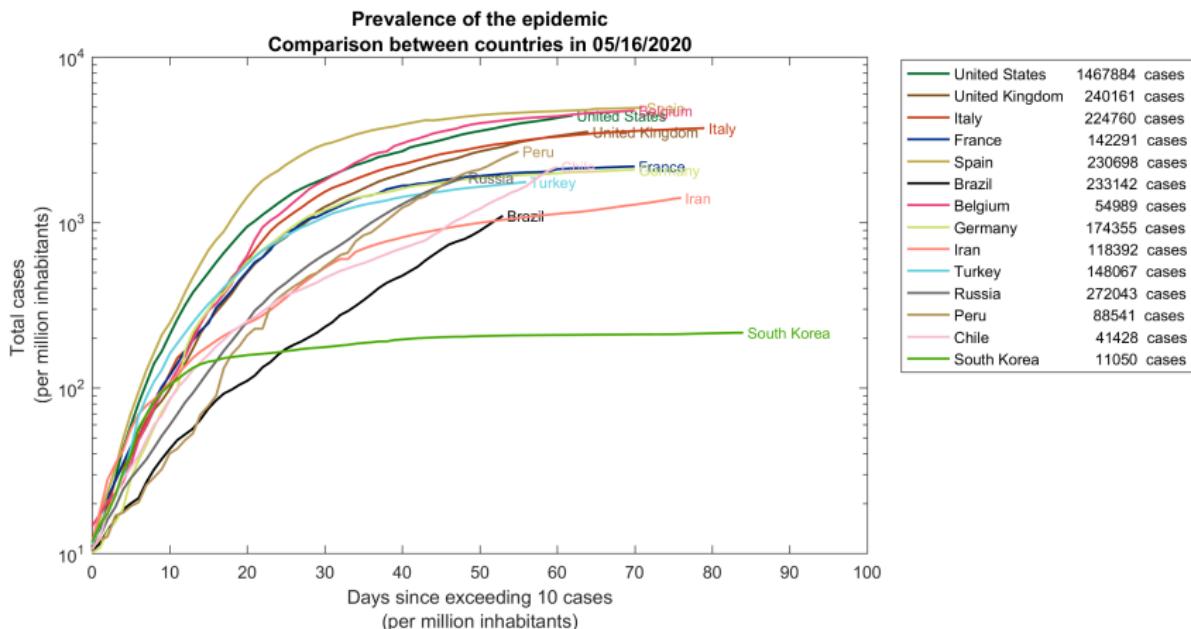
- A título de exemplificação são apresentados 6 tipos de gráficos de monitoramento. Por meio do código EPIDEMIC é possível gerar para cada um dos tipos aqui apresentados uma versão explicitando o número de casos e o número de mortes.
- Os exemplos a seguir foram gerados utilizando o módulo de tendências do código educacional EPIDEMIC para a situação da pandemia de COVID-19 em alguns países. Sendo assim, os resultados gráficos ao executar o código devem ser semelhantes aos aqui apresentados.

# Número total de casos em relação ao tempo



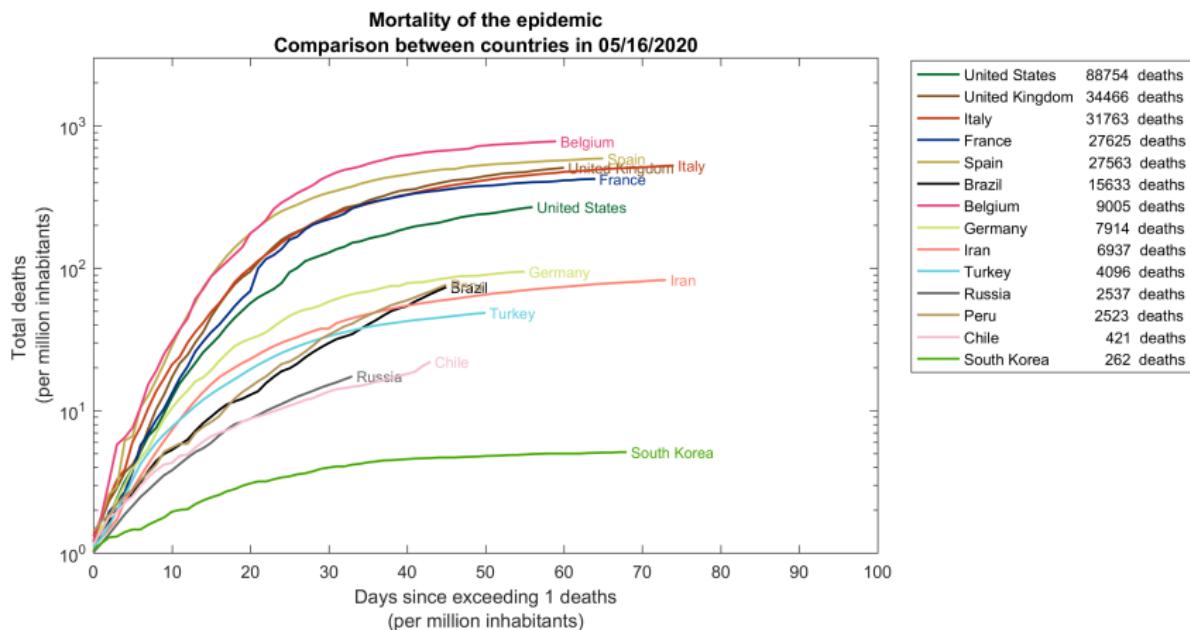
O gráfico analisa a evolução de contágio da epidemia, apresentando o total de casos em função do tempo. A mesma análise pode ser aplicada ao número de mortes (evolução da mortalidade).

# Número total de casos em relação ao tempo por 1M hab.



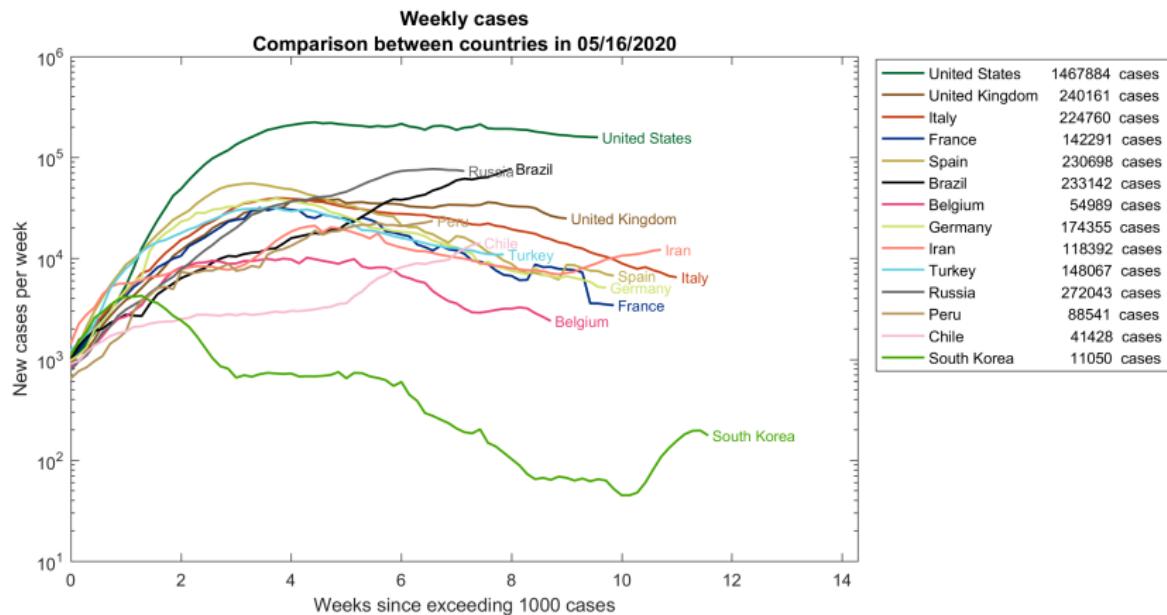
O gráfico permite analisar a prevalência da epidemia através da proporção de casos nos países listados na legenda, para isso emprega-se o total de casos por milhão de habitantes dos países avaliados.

# Número total de mortes em relação ao tempo por 1M hab.



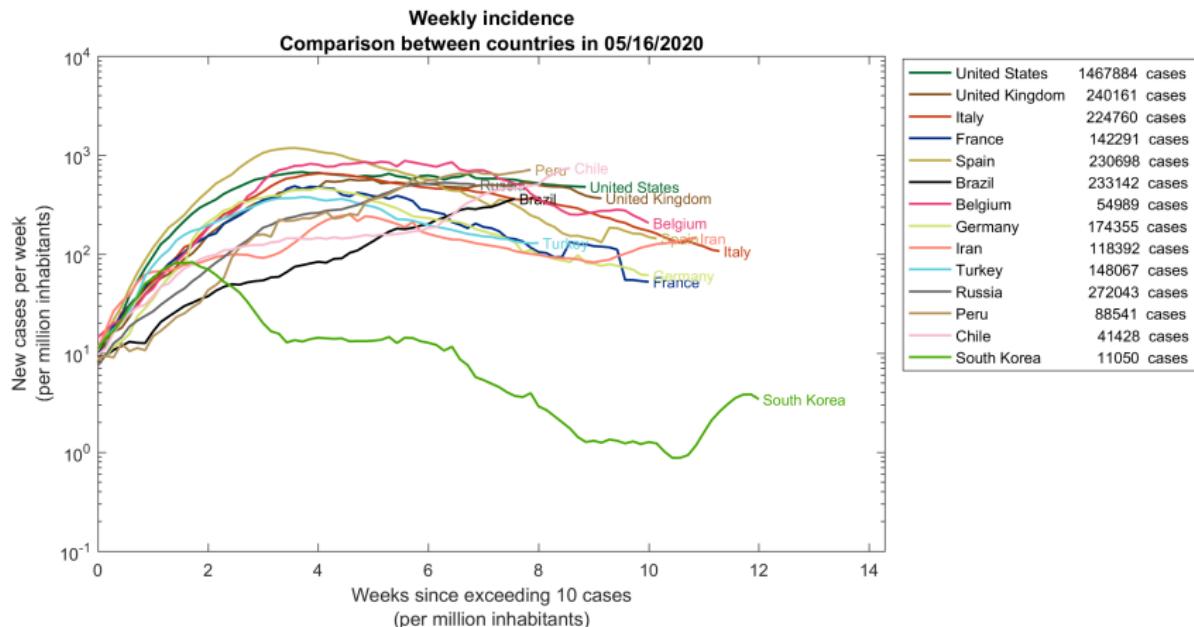
Assim como no gráfico anterior, a comparação da mortalidade entre os países emprega o total de mortes por milhão de habitantes do país.

# Novos casos por semana em relação ao tempo



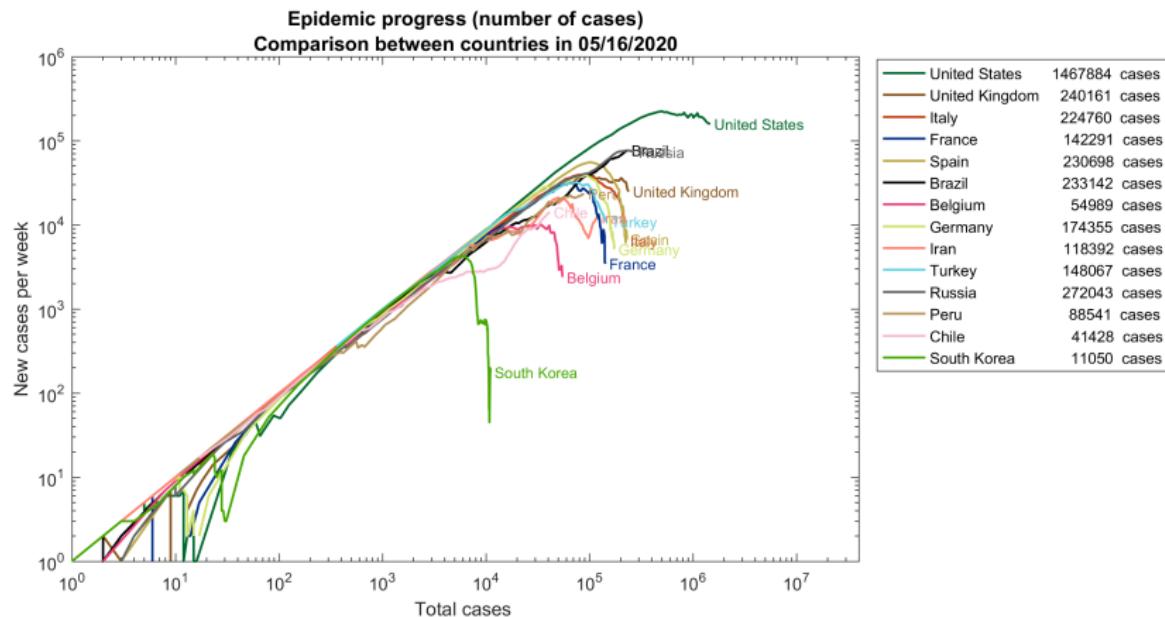
A escolha pela semana como parâmetro para a medida da evolução da epidemia deve-se ao fato de que o comportamento diário é muito oscilatório. A soma de todos os valores da semana filtra as oscilações presentes na atualização diária.

# Novos casos por semana no tempo por 1M hab.



Este gráfico é semelhante ao anterior, porém utiliza o número de novos casos por milhão de habitantes, possibilitando comparar o crescimento do número de mortes em relação à população (incidência).

# Novos casos por semana em relação ao total



O gráfico avalia o progresso da epidemia comparando a quantidade de novos casos semanais com o total de casos. Se a epidemia estiver em fase exponencial, a quantidade de novos casos aumenta proporcionalmente ao total, o que se reflete numa reta com inclinação positiva. Quando a epidemia entra em esgotamento nota-se uma queda.

## O código e os dados

Para o monitoramento dos números de casos e mortes de uma epidemia são necessárias fontes realistas e confiáveis de dados. Neste sentido, para compreender o funcionamento do código, os dados da pandemia de COVID-19 foram utilizados nas simulações.

Para execução do código `epidemic_trends.m` é preciso um arquivo com os dados da epidemia com a mesma disposição das informações do arquivo “owid-covid-data.csv”, o qual traz informações sobre a pandemia de COVID-19 em alguns países, e que está disponível para baixar em <https://ourworldindata.org/coronavirus-source-data>.

Além disso, esse arquivo deve ser salvo no mesmo diretório do código.

# Importando os dados

O código começa com a leitura do arquivo de dados e transformação destes para o formato matricial.

```
% Reading the data file
% (After downloading, save this file in the same directory as the code)
%
A = fileread('owid-covid-data.csv');
B = strsplit(A, '\n');

for i = 1:length(B{1,:})
    all_data(i,:) = strsplit(B{1,i}, ',', 'CollapseDelimiters', false);
end

data = [str2double(all_data(:, 4)) ...
    str2double(all_data(:, 5)) ...
    str2double(all_data(:, 6)) ...
    str2double(all_data(:, 7)) ...
    str2double(all_data(:, 8)) ...
    str2double(all_data(:, 9)) ...
    str2double(all_data(:,10)) ...
    str2double(all_data(:,11))];
```

---

## Configurando os locais analisados

Um laço (*loop*) é usado para produzir as curvas dos países monitorados. Eles são plotados na ordem definida pelo valor da variável “**init**”. Tal comando define quais países farão parte do gráfico gerado, bem como as cores. O país deve ser indicado como descrito no arquivo de dados.

```
% Running the loop command in the 14 countries studied
%
for init = 1:1:14

    % Cleaning up reused variables inside the loop
    clearvars -except plot_type init all_data data name paises tot_deaths tot_cases

    % Order by countries that have more death
    if (init == 6); country='Brazil' ; color=[ 0, 0, 0]/255; end
    if (init == 14); country='South Korea' ; color=[ 69,169, 0]/255; end
    if (init == 10); country='Turkey' ; color=[ 96,209,224]/255; end
    if (init == 12); country='Peru' ; color=[181,147, 87]/255; end
    if (init == 9); country='Iran' ; color=[255,130,113]/255; end
    if (init == 8); country='Germany' ; color=[209,227,105]/255; end
    if (init == 13); country='Chile' ; color=[248,187,208]/255; end
    if (init == 1); country='United States' ; color=[ 0,104, 44]/255; end
    if (init == 4); country='France' ; color=[ 0, 45,135]/255; end
    if (init == 2); country='United Kingdom' ; color=[135, 85, 30]/255; end
    if (init == 3); country='Italy' ; color=[203, 63, 23]/255; end
    if (init == 5); country='Spain' ; color=[191,171, 72]/255; end
    if (init == 7); country='Belgium' ; color=[236, 64,122]/255; end
    if (init == 11); country='Russia' ; color=[102,102,102]/255; end
```

## Definindo os dados a serem coletados

Os dados desejados, contidos no arquivo de dados, são então localizados por meio do comando “**find**”.

```
% Collecting country data
location = data(find(strcmp([all_data(:,2]), country)),1:8);

% Defining the matrix with dates and the final day
all_data_date = all_data(:,3);
dates         = all_data_date(find(strcmp([all_data(:,2)], country)),:);
end_time     = max(datenum(dates))-1;
```

## Vetorizando os dados de interesse

Os dados são separados em vetores contendo os casos totais, mortes totais, novos casos diários e novas mortes diárias. De modo que, esses vetores tenham comprimento igual ao número de dias avaliados e correspondam aos dados do país escolhido.

```
% Separating the data
tot_cases = location(:,1);
new_cases = location(:,2);
tot_deaths = location(:,3);
new_deaths = location(:,4);
```

A análise por milhão de habitantes também é apresentada, pois países mais populosos terão números absolutos maiores do que países menos populosos (provavelmente). Assim, a relação por milhão “**pm**” minimiza a influência do valor populacional nas comparações gráficas.

```
% Data per million
tot_cases_pm = location(:,5);
new_cases_pm = location(:,6);
tot_deaths_pm = location(:,7);
new_deaths_pm = location(:,8);
```

## Novos casos e mortes por semana

Os números de novos casos e novas mortes podem sofrer grandes variações de um dia para o outro. Neste contexto, o código a seguir é utilizado para definir que os novos casos ou as novas mortes sejam avaliadas por semana.

```
% Consolidating new cases and deaths per week
for i = 7:1:max(max(size(dates)))
    new_cases7(i,1) = sum(new_cases(i-6:i,1));
    new_deaths7(i,1) = sum(new_deaths(i-6:i,1));
    new_cases7_pm(i,1) = sum(new_cases_pm(i-6:i,1));
    new_deaths7_pm(i,1) = sum(new_deaths_pm(i-6:i,1));
end
```

## Definindo o eixo do tempo (dias)

Por fim, são gerados dados desde que se ultrapassa “X\_deaths” mortes e “X\_cases” casos. O usuário pode definir esses valores iniciais, porém é indicado que sejam valores já alcançados em todos os países analisados.

```
% Creating vectors from day zero from X deaths ('_deaths')
% or from X cases ('_cases').
% User sets zero day for cases or deaths
X_deaths_pm = 1;
X_cases_pm  = 10;
X_deaths    = 100;
X_cases     = 1000;
```

## Definindo o eixo vertical (dados)

Os comandos para a visualização de cada informação são apresentados a seguir. Sendo eles: número de mortes, número de casos, número de casos por milhão e número de mortes por milhão.

```
% Relative to total deaths
n = 0;
for i = 1:1:max(max(size(dates)))
    if (tot_deaths(i,1)    >= X_deaths)
        n = n + 1;
        tot_deaths_X (n,1) = tot_deaths (i,1);
        new_deaths_X (n,1) = new_deaths (i,1);
        new_deaths7_X(n,1) = new_deaths7(i,1);
    end
end

% Relative to total cases
n = 0;
for i = 1:1:max(max(size(dates)))
    if (tot_cases(i,1)    >= X_cases)
        n = n + 1;
        tot_cases_X (n,1) = tot_cases (i,1);
        new_cases_X (n,1) = new_cases (i,1);
        new_cases7_X(n,1) = new_cases7(i,1);
    end
end
```

# Definindo o eixo vertical (dados)

```
% Related deaths per million
n = 0;
for i = 1:1:max(max(size(dates)))
    if (tot_cases_pm(i,1)  >= X_cases_pm)
        n = n + 1;
        tot_cases_pmX (n,1) = tot_cases_pm (i,1);
        new_cases_pmX (n,1) = new_cases_pm (i,1);
        new_cases7_pmX(n,1) = new_cases7_pm(i,1);
    end
end

% Related cases per million
n = 0;
for i = 1:1:max(max(size(dates)))
    if (tot_deaths_pm(i,1)  >= X_deaths_pm)
        n = n + 1;
        tot_deaths_pmX (n,1) = tot_deaths_pm (i,1);
        new_deaths_pmX (n,1) = new_deaths_pm (i,1);
        new_deaths7_pmX(n,1) = new_deaths7_pm(i,1);
    end
end
```

# Configurações básicas dos gráficos

Para a plotagem devem ser incluídas algumas informações iniciais como:

- Definição das fontes utilizadas no título, nos eixos e demais textos contidos nos gráficos;
- Definição da quantidade de dias que serão expressos no eixo x nos gráficos temporais.

```
% Basics graphics settings
```

```
% Fonts
font_title    = 10.5;
font_labels   = 10;
font_default  = 9;
font_location = 8;
```

```
% X-axis size
day_axis = 100;
```

## Alinhamento da legenda

Para melhorar a apresentação da relação de países e os seus respectivos números de casos ou mortes na legenda, são utilizados espaçamento após o nome dos países através dos comandos a seguir:

```
% Aligning the legend
if strcmp(country, 'Brazil')      ); country_leg='Brazil'      ' ; end
if strcmp(country, 'South Korea') ); country_leg='South Korea' ' ; end
if strcmp(country, 'Turkey')       ); country_leg='Turkey'       ' ; end
if strcmp(country, 'Peru')        ); country_leg='Peru'         ' ; end
if strcmp(country, 'Iran')        ); country_leg='Iran'         ' ; end
if strcmp(country, 'Germany')     ); country_leg='Germany'     ' ; end
if strcmp(country, 'Chile')       ); country_leg='Chile'        ' ; end
if strcmp(country, 'United States'); country_leg='United States' ' ; end
if strcmp(country, 'France')      ); country_leg='France'       ' ; end
if strcmp(country, 'United Kingdom'); country_leg='United Kingdom' ' ; end
if strcmp(country, 'Italy')        ); country_leg='Italy'         ' ; end
if strcmp(country, 'Spain')       ); country_leg='Spain'        ' ; end
if strcmp(country, 'Belgium')     ); country_leg='Belgium'      ' ; end
if strcmp(country, 'Russia')      ); country_leg='Russia'       ' ; end
```

## Alinhamento da legenda

Em seguida, é realizado o alinhamento a direita dos números de casos ou mortes, melhorando visualmente a apresentação dos dados.

```
deaths_leg = [ ' ', num2str(max(tot_deaths))];  
if size(num2str(max(tot_deaths)))<5; deaths_leg=[ ' ', ' ', ' ', ' ', ' ', num2str(max(tot_deaths))]; end  
if size(num2str(max(tot_deaths)))<4; deaths_leg=[ ' ', ' ', ' ', ' ', num2str(max(tot_deaths))]; end  
if size(num2str(max(tot_deaths)))<3; deaths_leg=[ ' ', ' ', ' ', num2str(max(tot_deaths))]; end  
if size(num2str(max(tot_deaths)))<2; deaths_leg=[ ' ', ' ', num2str(max(tot_deaths))]; end  
  
cases_leg = [ ' ', num2str(max(tot_cases))];  
if size(num2str(max(tot_cases)))<7; cases_leg=[ ' ', ' ', ' ', ' ', ' ', ' ', ' ', num2str(max(tot_cases))]; end  
if size(num2str(max(tot_cases)))<6; cases_leg=[ ' ', ' ', ' ', ' ', ' ', ' ', num2str(max(tot_cases))]; end  
if size(num2str(max(tot_cases)))<5; cases_leg=[ ' ', ' ', ' ', ' ', num2str(max(tot_cases))]; end  
if size(num2str(max(tot_cases)))<4; cases_leg=[ ' ', ' ', ' ', num2str(max(tot_cases))]; end
```

# Plotagem dos gráficos

Para a plotagem dos gráficos existem dois formatos que podem ser utilizados. O primeiro exemplo é referente às **análises temporais**. A seguir é apresentado um exemplo utilizando o total de mortes pelo tempo.

```
% Plot total deaths by time (zero day defined by deaths)
figure (1);
n    = max(max(size(tot_deaths_X)));
days = 0:1:n-1;
fig = semilogy(days,tot_deaths_X, ...
                'DisplayName',[country_leg,' ',deaths_leg,' ','deaths'],...
                'color',color, ...
                'LineWidth',1.25);
hold on;
text (n-1,tot_deaths_X(n,1),[' ',country_leg],...
      'FontSize',font_location, ...
      'color',color, ...
      'Clipping','on');
```

As demais plotagens temporais são obtidas trocando o termo **tot\_deaths** por **tot\_cases** para casos e o **tot\_deaths\_X** por:

- **new\_deaths7\_X** (novas mortes) ou **new\_deaths7\_pmX** (por milhão);
- **tot\_cases\_X** (total de casos) ou **tot\_cases\_pmX** (por milhão);
- **new\_cases7\_X** (novos casos) ou **new\_cases7\_pmX** (por milhão).

## Plotagem dos gráficos

O segundo exemplo é referente às **análises atemporais**, as quais relacionam os dados novos com os dados totais. A seguir é apresentado um exemplo das novas mortes pelo total.

```
% Plot new deaths by time (zero day defined by deaths)
figure (5);
n    = max(max(size(new_deaths_X)));
days = 0:1:n-1;
fig = semilogy(days/7,new_deaths7_X, ...
                'DisplayName',[country_leg , ' ', deaths_leg , ' ', 'deaths'] , ...
                'color',color , ...
                'LineWidth',1.25);
hold on;
text ((n-1)/7,new_deaths7_X(n,1),[ ' ',country_leg ] , ...
      'FontSize',font_location , ...
      'color',color , ...
      'Clipping','on');
```

As demais plotagens atemporais são obtidas trocando o termo **deaths** por **cases**.

# Configurações estéticas dos gráficos

Por fim, são realizadas configurações em relação ao tamanho da figura, título do gráfico e dos eixos e a posição da legenda. A seguir são apresentadas as configurações para dois exemplos de plotagens mostrados anteriormente.

```
% Aesthetic graphics settings
%
figure (1);
set(gca,'FontSize',font_default)
set(gcf,'units','normalized','OuterPosition',[0 0 0.7 0.7])
title ({'Acumulated deaths',...
    ['Comparison between countries in ',datestr(end_time,23)]},...
    'FontSize',font_title);
xlabel (['Days since exceeding ',num2str(X_deaths),' deaths'] ,...
    'FontSize',font_labels);
ylabel ('Total deaths ','FontSize',font_labels);
legend ('location','northeastoutside');
y_init = 100;
max_y = 500000;
axis([0 day_axis y_init max_y]);
```

# Configurações estéticas dos gráficos

O comando “**axis**” é utilizado para definir o intervalo dos eixos. Essa configuração pode ser alterada conforme o interesse do usuário.

```
figure (5);
set(gca,'FontSize',font_default)
set(gcf,'units','normalized','OuterPosition',[0 0 0.7 0.7])
title ({'Weekly deaths',...
    ['Comparison between countries in ',datestr(end_time,23)]},...
    'FontSize',font_title);
xlabel ({{'Weeks since exceeding ',num2str(X_deaths),' deaths'}},...
    'FontSize',font_labels);
ylabel ({{'New deaths per week'}}, 'FontSize', font_labels);
legend ('location', 'northeastoutside');
y_init = 1;
max_y = 100000;
axis([0 day_axis/7 y_init max_y]);
```

# Salvando os gráficos

São dez tipos de gráficos plotados por este código. Sendo eles:

Tipo de gráfico	Nome da figura
1) Mortes acumuladas - valor absoluto (total de mortes por tempo)	deaths-total-abs_AAAA-MM-DD
2) Casos acumulados - valor absoluto (total de casos por tempo)	cases-total-abs_AAAA-MM-DD
3) Progresso de mortes - valor absoluto (novas mortes por total de mortes)	deaths-progress-abs_AAAA-MM-DD
4) Progresso de casos - valor absoluto (novos casos por total de casos)	cases-progress-abs_AAAA-MM-DD
5) Mortes semanais - valor absoluto (novas mortes por tempo)	deaths-weekly-abs_AAAA-MM-DD
6) Casos semanais - valor absoluto (novo casos por tempo)	cases-weekly-abs_AAAA-MM-DD
7) Mortalidade - por milhão habitantes (total de mortes por milhão)	mortality-pm_AAAA-MM-DD
8) Prevalência - por milhão habitantes (total de casos por milhão no tempo)	prevalence-pm_AAAA-MM-DD
9) Mortes semanais - por milhão habitantes (novas mortes semanais por milhão)	deaths-weekly-pm_AAAA-MM-DD
10) Incidência - por milhão habitantes (novos casos semanais por milhão)	incidence-weekly-pm_AAAA-MM-DD

# Salvando os gráficos

Caso seja necessário salvar as plotagens, basta utilizar o comando “*print*”, definindo o nome do arquivo e o formato no qual o mesmo será salvo. A seguir são apresentados os comandos utilizados para salvar os exemplos anteriores.

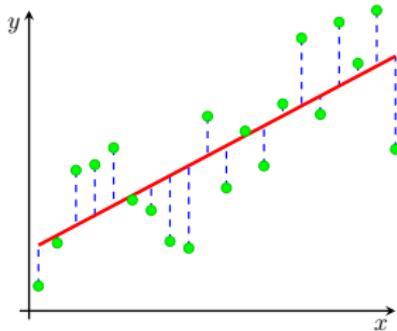
```
% Saving the charts
%
print(figure( 1 ),[ 'deaths-total-abs_' , datestr(end_time,29) , '.png' ], '-dpng', '-r300');
print(figure( 5 ),[ 'deaths-weekly-abs_' , datestr(end_time,29) , '.png' ], '-dpng', '-r300');
```

O nome dos gráficos é apresentado conforme o código original e no padrão definido anteriormente neste documento.

# Previsões

# Regressor

- Previsões são feitas com o auxílio de regressores estatísticos;
- O processo de estimativa usa o **Método dos Mínimos Quadrados**.



## Objetivo:

Obter uma curva de ajuste (—) a um certo conjunto de dados (•), visando minimizar a soma dos quadrados das diferenças (---) entre os valores estimados e os dados observados.

## Metodologia

Na elaboração do módulo de previsões levou-se em conta os seguintes pontos:

- Considerar os últimos **5 dias** da amostra de dados, pois estes refletem a **tendência mais recente**;
- Inserir os últimos 5 dias na **escala logarítmica** (base 10), pois caso o crescimento seja exponencial, isto refletirá em uma reta;
- Realizar uma **regressão linear**, pelo **Método dos Mínimos Quadrados**. Considerar a banda de confiança de 95%;
- Plotar os valores previstos dentro do envelope de confiabilidade estimado.

# Método dos Mínimos Quadrados

A regressão linear utilizando o **Método dos Mínimos Quadrados** é feita obtendo os estimadores  $\beta_1$  e  $\beta_2$ , considerando uma equação do tipo  $Y_i = \beta_1 + \beta_2 X_i$ .

$$\beta_2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

$$\beta_1 = \bar{Y}_n - \beta_2 \bar{X}_n$$

Onde:

- $\bar{X}_n$  e  $\bar{Y}_n$  são as médias amostrais.



L. Wasserman. **All of Statistics: A Concise Course in Statistical Inference**. Springer, New York, NY, 2004. DOI: <https://doi.org/10.1007/978-0-387-21736-9>

## Banda de Confiabilidade de 95%

A banda de confiabilidade de 95% é obtida através das variâncias da regressão e da predição, dada pela equação abaixo:

$$\xi_n^2 = \sigma^2 \left( \frac{\sum_{i=1} n (X_i - X_*)^2}{n \sum_i (X_i - \bar{X})^2} + 1 \right)$$

$$Y_* \pm 1.96 \xi_n$$

Considerando:

- $X_*$  e  $Y_*$  são os valores estimados;
- $\sigma^2 = (\frac{1}{n-2}) \sum_{i=1} n \epsilon_i^2$ ;
- $\epsilon_i = Y_i - (\beta_1 + \beta_2 X_i)$ .

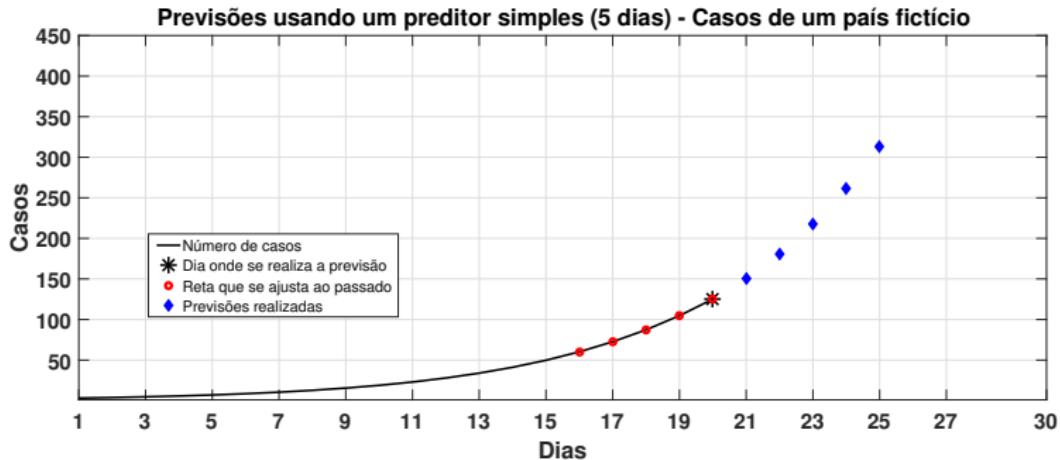


L. Wasserman. All of Statistics: A Concise Course in Statistical Inference. Springer, New York, NY, 2004. DOI: <https://doi.org/10.1007/978-0-387-21736-9>.

## Observações

- A progressão das estimativas pode-se alterar com o decorrer do tempo, por isso utilizar dados muito pregressos pode não gerar uma amostra confiável para a previsão;
- Realizar previsões para muitos dias à frente também pode não vislumbrar bem a realidade, por isso a escolha pela previsão para um horizonte de 5 dias.

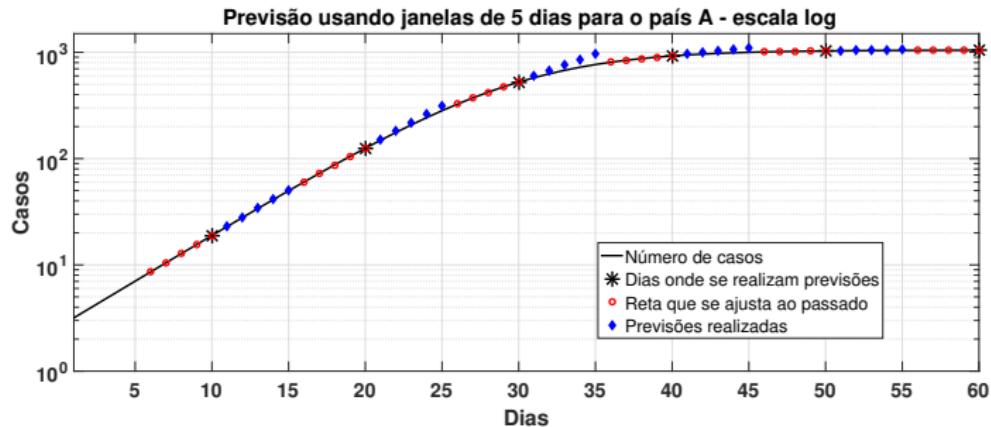
# Previsões de casos ou óbitos no curto prazo



- 1 Encontrar a curva que melhor se aproxima ao logaritmo da sequência do número de casos até o 5º dia anterior à data em que a previsão é realizada (dia 20). Os círculos vermelhos “○” são pontos da curva;
- 2 Extrapolar essa curva a partir da data de realização da previsão tentando prever os 5 pontos futuros da curva de casos, marcados no gráfico com ‘◊’.

## Exemplo: previsões de casos ou óbitos no curto prazo

- Gráfico em escala logarítmica das previsões de casos num país fictício A;



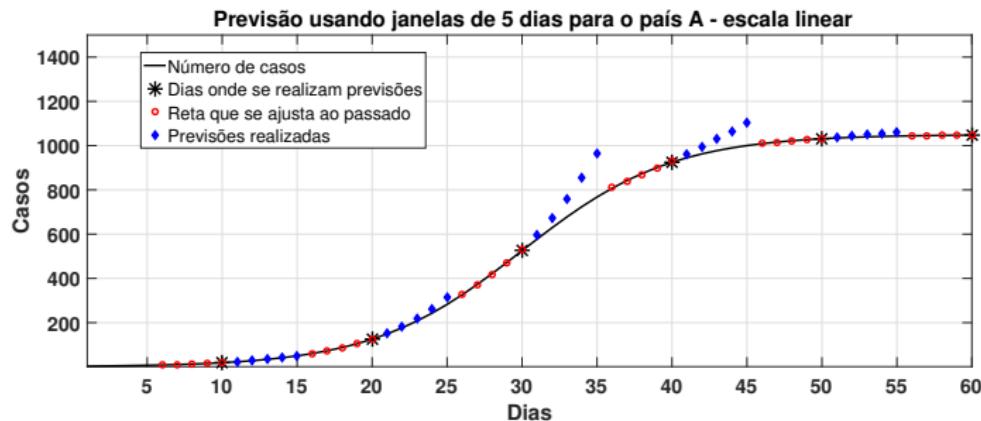
- Neste gráfico temos a curva original de casos num período de 60 dias, com 5 previsões realizadas “\*”, possibilitando uma comparação previsto e realizado;
- A medida em que a curva se afasta do comportamento exponencial, a diferença entre “ $\diamond$ ” e a reta “—”, ou seja, os **erros da previsão aumentam**.



A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21<sup>a</sup> Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

# Exemplo: previsões de casos ou óbitos no curto prazo

- Gráfico em escala linear das previsões de casos num país fictício A;



- Usando a mesma ferramenta de previsão, mas agora na escala linear, as diferenças entre as previsões e a curva original são visivelmente maiores entre o 30º e 45º dia;
- Isso ocorre porque a previsão é realizada usando uma reta na escala logarítmica e nessa região a curva original deixa de ter o comportamento exponencial.



A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21ª Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>

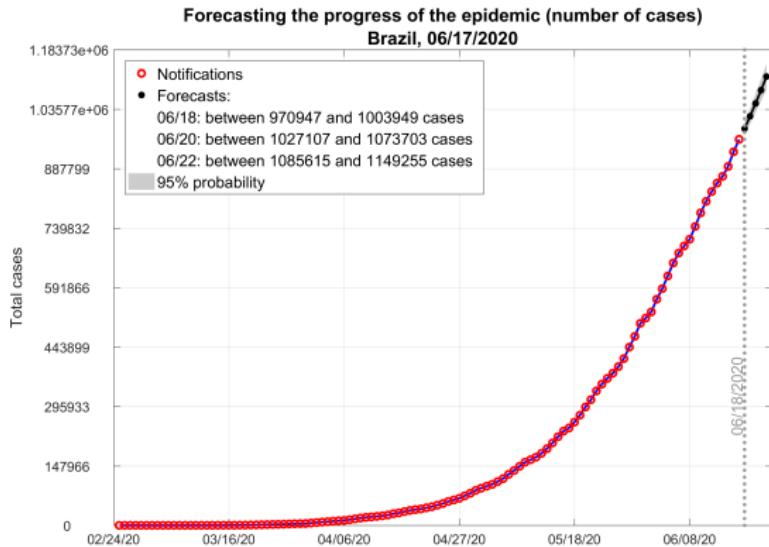
## Exemplo: COVID-19

- Utilização da situação-problema da COVID-19 no Brasil;
- Dois tipos de gráficos: **previsão de casos** e **previsão de mortes**;
- O local analisado pode ser modificado.

## Exemplo COVID-19

Previsão do nº total de casos da COVID-19 no Brasil (em função do tempo)

- Resultado gráfico no EPIDEMIC:

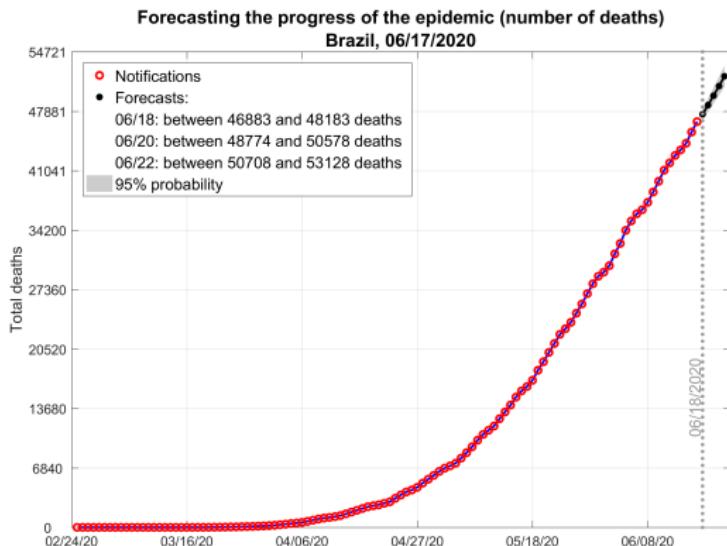


- No gráfico, tem-se a banda de confiança, em cinza, de 95%.

# Exemplo COVID-19

Previsão do nº total de mortes da COVID-19 no Brasil (em função do tempo)

- Resultado gráfico no EPIDEMIC:



- No gráfico, tem-se a banda de confiança, em cinza, de 95%.

## O código e os dados

Assim como no módulo tendências, para as previsões de casos e mortes numa epidemia são necessárias fontes confiáveis de dados. Para compreender o funcionamento do código, foram utilizados os dados da pandemia de COVID-19.

Para execução do código `epidemic_forecasts.m` é necessário um arquivo com os dados da epidemia com a mesma disposição das informações do arquivo “`cases-brazil-states.csv`”, o qual traz informações sobre a pandemia de COVID-19 nos estados federativos do Brasil, e que está disponível para baixar em <https://github.com/wcota/covid19br>. Além disso, esse arquivo deve ser salvo no mesmo diretório do código.

O algoritmo está dividido em duas seções. A primeira com os comandos para importação e organização dos dados referentes ao local analisado, e a segunda é referente à plotagem dos gráficos de previsões de casos e mortes, respectivamente.

## Importando dados

O código inicia com a data da última atualização de dados. Em seguida realiza-se a leitura do arquivo de dados e os dados numéricos são transferidos para uma matriz.

```
%% Organizing death and case data
%
% Insert the last data update date, in the format MM/DD/AAAA
%
up_date = '06/17/2020';
%
%
% Reading the data file
%(After downloading, save this file in the same directory as the code)
%
A = fileread('cases-brazil-states.csv');
B = strsplits(A, '\n');

for i = 1:length(B(1,:))
    all_data(i,:) = strsplits(B{1,i}, ',', 'CollapseDelimiters', false);
end
%
```

## Definindo o local de análise

Com os dados importados, é necessário definir o local, que neste exemplo é o Brasil, e que será a soma dos dados de seus estados definidos por suas siglas. Os comandos “**find**” e “**for**” são utilizados para encontrar e definir o tamanho dos dados baseado no estado com maior número de dados.

```
% Defining country as the sum of data for all its federal states
%
country = 'Brazil';
states = {'RO' 'AC' 'AM' 'RR' 'PA' 'AP' 'TO' 'MA' 'PI' ...
          'CE' 'RN' 'PB' 'PE' 'AL' 'SE' 'BA' 'MG' 'ES' ...
          'RJ' 'SP' 'PR' 'SC' 'RS' 'MS' 'MT' 'GO' 'DF'};
%
%
% Checking the state data size
%
for i = 1:length(states)
    location = all_data(find(strcmp(all_data(:,3),states(1,i))),:);
    maxcases(1,i) = length(cellfun(@str2double,(location(:,8))));
    maxdeaths(1,i) = length(cellfun(@str2double,(location(:,6))));
end
%
```

## Coletando os dados de casos e mortes

Para concluir a primeira seção de organização dos dados, são utilizados novamente os comandos “**for**” e “**find**”, porém agora para coletar os dados de casos e mortes.

```
% Collecting data on total cases and deaths for each state
%
cases_states = zeros(length(states),max(maxcases) );
deaths_states = zeros(length(states),max(maxdeaths));

for i = 1:length(states)
    location = all_data(find(strcmp(all_data(:,3),states(1,i))),:);

    cases_states(i,:) = [ zeros(1,max(maxcases)-maxcases(1,i)) ,...
        nonzeros(cellfun(@str2double,location(:,8))) ];

    deaths_states(i,:) = [ zeros(1,max(maxdeaths)-maxdeaths(1,i)) ,...
        cellfun(@str2double,location(:,6)) ];
end
```

## Consolidando os dados de casos e mortes

Em seguida, são consolidados os dados do país como a soma dos estados utilizando os mesmos comandos.

```
% Consolidating data on total cases and deaths for the country
%
cases_country = zeros(1,length(cases_states (1,:)));
deaths_country = zeros(1,length(deaths_states(1,:)));

for i = 1:length(states)
    cases_country = cases_country + cases_states (i,:);
    deaths_country = deaths_country + deaths_states(i,:);
end

cases = cases_country ;
deaths = deaths_country;
%
```

# Loop para cálculo da regressão e da plotagem

- Nessa seção do código, serão realizadas a regressão linear e as configurações para a plotagem dos dados e das previsões.
- Como os comandos para plotagem da previsão são os mesmos para casos e mortes, um laço de comandos realiza-a na ordem definida pela variável “**init**”.

```
%% Plotting the forecast of cases/deaths by time
%
for init = 1:1:2

    % Cleaning up reused variables inside the loop
    clearvars -except up_date init cases country deaths all_data ...
               cases_country deaths_country cases_states ...
               deaths_states location maxcases maxdeaths states

    % Data loop for plotting cases and then deaths
    if (init == 1); data = cases ; info = 'cases' ; end
    if (init == 2); data = deaths; info = 'deaths' ; end
```

## Ajuste dos dias

O primeiro passo é realizar o ajuste de como serão lidas as datas. Logo, a data é transformada de uma **"string"** para um dado numérico através do comando **"str2num"**. Em seguida é gerado o vetor dos dias, um por um para o número de casos e mortes.

```
% Defaults: Adjust days and dates
%
date = [up_date(1,1:2), '—', up_date(1,4:5), '—', up_date(1,7:10)];
days = length (data);
day = str2num(up_date(1,1:2));
%
```

## Configurando as datas em números

É realizado o ajuste das datas em “números” através do comando “**datenum**” para os dias de previsão, que são o próximo, o terceiro e o quinto dia seguinte à data do último valor.

```
% Setting string for days ahead (Transforming the date to strings)
% This make sure that, given the day, the program gives the days ahead
% without end of month problems like "30/05, 32/05, 34/05, ..."
%
Datestring = date;
formatIn  = 'mm-dd-yyyy';

ND = datenum(Datestring,formatIn)+1;
ND3 = ND+2;
ND5 = ND+4;
%
```

# Configurando as datas em "string"

Em seguida, é realizado o ajuste das datas em “string” para inclusão na legenda e no eixo X dos gráficos de previsão.

```
% Setting to include in the legend
%
str1 = datestr(ND , 'mm-dd-yyyy');
str3 = datestr(ND3, 'mm-dd-yyyy');
str5 = datestr(ND5, 'mm-dd-yyyy');

LINE_DATE = [ str1(1,1:2) , '/' , str1(1,4:5) , '/' , str1(1,7:10) ];
d1 = [ str1(1,1:2) , '/' , str1(1,4:5) ];
d3 = [ str3(1,1:2) , '/' , str3(1,4:5) ];
d5 = [ str5(1,1:2) , '/' , str5(1,4:5) ];
%
```

```
% Setting to include in the X-axis
%
final = datenum(Datestring , formatIn);
first = final - days;
str = datestr(first : 21 : final , 'mm/dd/yy');
%
```

## Vetorizando os dias

Em seguida, é gerado o vetor dos dias, um por um para o número de casos e mortes.

```
% Vector of days, one by one for the number of cases/deaths.  
%  
days = sparse(1,length(data));  
  
for i = 1:length(data)  
    days(1,i) = i;  
end  
%
```

# Regressão linear

Para realizar a regressão linear, toma-se os últimos 5 dias dos dados como amostra e, em seguida, aplica-se a regressão utilizando o comando “**polyfit**”, gerando os coeficientes da reta que será utilizada para calcular a previsão dos 5 dias.

```
% Take the last 5 days of the data as a sample
% Note: "5s" refers to the last 5 days taken as a sample
%
data5s = sparse(1,5);
days5s = sparse(1,5);

for i = 0:4
    data5s(1,i+1) = data(1,length(data)-4+i);
    days5s(1,i+1) = length(data)-4+i;
end
%
%
% Linear regression of log points vs days of the sample (last 5 days)
%
[coeffs5s,S_5s] = polyfit(days5s,log10(data5s),1);
%
```

## Previsão de 5 dias

Aplicando a regressão linear, os valores preditos para os 5 dias são gerados utilizando o comando “**polyval**” com os coeficientes da regressão e com os dias de previsão.

```
% 5-day prediction
% Note: "5f" refers to the 5-day forecast
%
data5f = sparse(1,5);
days5f = sparse(1,5);

for i = 1:5
    days5f(1,i) = length(data) + i;
end

data5f = polyval(coeffs5s, days5f);
%
```

## Banda de confiabilidade

Para finalizar, obtém-se a banda de confiabilidade de 95%. Para isso, são calculadas as variâncias da predição e da correção.

```
% Confidence band calculation
%
% Method 1
%
% Fitted line
datafit = coeffs5s(1,2) + coeffs5s(1,1)*days5s;
% Variance estimation
N = length(days5s);
sigma2 = sum((log10(data5s) - datafit).^2)/(N-1);
% Denominator
den = N*sum((days5s-mean(days5s)).^2);

for i = 1:N;
    % Numerator
    num = sum((days5s-days5f(1,i)).^2);
    % Variance square correction
    ksi2(1,i) = sigma2*((num/den)+1);
end
% Variance correction
ksi = sqrt(ksi2);

% Confidence band
upperCB = data5f + 1.96*ksi;
lowerCB = data5f - 1.96*ksi;
```

# Banda de confiabilidade

Outro método para calcular, é pelo comando “**polyconf**”.

```
% Method 2 (This method uses a direct command)
%
% To run on MATLAB disable the commands below
pkg install -forge optim
pkg load optim

[deaths5f,dy] = polyconf(coeffs5s ,days5f ,S_5s);
upperCB_data5f = data5f + dy;
lowerCB_data5f = data5f - dy;
%
```

# Configurações do gráfico

A partir de agora, são apresentadas as configurações dos gráficos. Iniciando pela plotagem dos casos/mortes notificados e pela definição dos títulos.

```
% Chart configuration
%
figure

% Plot command notifications points
set(gca , 'YScale','linear')
plot(days,data , 'b' , 'LineWidth' ,2);
hold on
c = scatter(days,data , 'r' , 'LineWidth' ,2);

% Titles configuration
set(gca , 'FontSize',12)
title({{['Forecasting the progress of the epidemic '...
    '(number of ',info , ')']},...
    [country , ' ', ' ',up_date]} ,...
    'FontSize',16);
ylabel({{['Total ',info]}}, 'FontSize',14)
hold on
```

# Plotagem da previsão e banda de confiabilidade

Em seguida, plota-se os casos/mortes para os 5 dias de previsão e a banda de confiabilidade.

```
% Confidence band
CB = fill([days5f fliplr(days5f)] ,...
[10.^lowerCB fliplr(10.^upperCB)] ,...
[0.8 0.8 0.8] ,...
'LineStyle','none');

% Forecasts points
plot(days5f,10.^(data5f),'k','LineWidth',2);
hold on
b = scatter(days5f,10.^(data5f),'k','filled');
hold on
```

## Configuração da legenda

Configura-se a legenda, inserindo o intervalo de casos e mortes previstos para os dias seguintes.

```
% Legend configuration
aaa = scatter(1,1,1,[1 1 1]);
bbb = scatter(1,1,1,[1 1 1]);
ccc = scatter(1,1,1,[1 1 1]);
lgd = legend([c b aaa bbb ccc CB] ,...
    'Notifications' ,...
    'Forecasts: ' ,...
    [d1 ,': between ' ,num2str(round(10.^lowerCB(1,1))) ,...
        ' and ' ,...
        num2str(round(10.^upperCB(1,1))) ,...
        ' ',info] ,...
    [d3 ,': between ' ,num2str(round(10.^lowerCB(1,3))) ,...
        ' and ' ,...
        num2str(round(10.^upperCB(1,3))) ,...
        ' ',info] ,...
    [d5 ,': between ' ,num2str(round(10.^lowerCB(1,5))) ,...
        ' and ' ,...
        num2str(round(10.^upperCB(1,5))) ,...
        ' ',info] ,...
    '95% probability' ,...
    'Location','northwest','Orientation','vertical');
set(lgd , 'FontSize',14);
```

# Separando notificações e previsões

Uma linha pontilhada separando as notificações e as previsões é utilizada para evidenciar quais são os dados reais dos previstos.

```
% Dotted line between notifications and forecasts
scat = scatter(ones(1,70)*days5f(1,1) ,...
    0:...
    (10^(upperCB(1,5))*1.03) / 69:...
    10^(upperCB(1,5))*1.03 ,...
    10,[0.6 0.6 0.6], 'filled');

txxt = text(days5f(1,1)-1.5 ,...
    0.2*10^(data5f(1,5)) ,...
    LINE_DATE ,...
    'FontSize' ,14 ,...
    'Color',[0.6 0.6 0.6]);
set(txxt , 'Rotation',90);
```

# Configuração da escala

Para finalizar as configurações do gráfico, ajusta-se as escalas.

```
% Scale configuration
grid

ylim([0 10^(upperCB(1,5))*1.03])
set(gca,'ytick',...
    floor(0:(10^(upperCB(1,5))*1.03)/8:10^(upperCB(1,5))*1.03),...
    'yticklabel',...
    floor(0:(10^(upperCB(1,5))*1.03)/8:10^(upperCB(1,5))*1.03));

set(gca,'xtick',0:21:length(days),'xticklabel',str);
set(gcf,'Position',[100, 100, 1000, 700]);
```

# Salvando o gráfico

Por fim, utiliza-se o comando “**print**” para salvar o gráfico no mesmo diretório no qual o código foi executado.

```
% Saving the chart
%
print(figure(1) ,['forecasts-cases','.png'], '-dpng', '-r300');
%
print(figure(2) ,['forecasts-deaths','.png'], '-dpng', '-r300');
```

# Agradecimentos

## Suporte financeiro



Fundaçao Carlos Chagas Filho de Amparo  
à Pesquisa do Estado do Rio de Janeiro



Conselho Nacional de Desenvolvimento  
Científico e Tecnológico



## Suporte institucional



UFRJ



INSTITUTO  
FEDERAL  
Mato Grosso  
do Sul



# Referências

## Referências: livros

-  F. Brauer, P. van den Driessche and J. Wu. **Mathematical Epidemiology**, Springer-Verlag, Berlin, 2008. DOI: <https://doi.org/10.1007/978-3-540-78911-6>
-  F. Brauer and C. Castillo-Chávez. **Mathematical Models in Population Biology and Epidemiology**, Springer, New York, 2012. DOI: <https://doi.org/10.1007/978-1-4614-1686-9>
-  L. Wasserman. **All of Statistics: A Concise Course in Statistical Inference**, Springer, New York, NY, 2004. DOI: <https://doi.org/10.1007/978-0-387-21736-9>
-  M. Martcheva. **An Introduction to Mathematical Epidemiology**, Springer US, New York, 2015. DOI: <https://doi.org/10.1007/978-1-4899-7612-3>
-  P. C. Hansen, V. Pereyra and G. Scherer. **Least Squares Data Fitting with Applications**. Baltimore: Johns Hopkins University Press., 2012. DOI: [10.1353/book.21076](https://doi.org/10.1353/book.21076)
-  R. M. Anderson and R. M. May. **Infectious Diseases of Humans: Dynamics and control**, Oxford University Press, Oxford, 1992. DOI: <https://doi.org/10.1002/hep.1840150131>

## Referências: artigos e relatórios

-  A. Cunha Jr et al. Relatório 01 Progresso da COVID-19 no Brasil e no Estado do Rio de Janeiro 21ª Semana Epidemiológica do Calendário 2020 (17/05/2020) até (23/05/2020). COVID-19: Observatório Fluminense, (24/05/2020). DOI: <https://doi.org/10.12957/eduerj.covid19rj.relatorio1>
-  D. Volpatto, A. C. Resende, L. Dos Anjos, J. Silva, C. Dias, R. Almeida and S. Malta. **Assessing social distancing release strategies in Brazil and Rio de Janeiro state**, Scientific Electronic Library Online, 2020. DOI: <https://doi.org/10.1590/SciELOPreprints.595>
-  D. Volpatto, A. C. Resende, L. Dos Anjos, J. Silva, C. Dias, R. Almeida and S. Malta. **Spreading of COVID-19 in Brazil: Impacts and uncertainties in social distancing strategies**, medrxiv, 2020. DOI: <https://www.medrxiv.org/content/10.1101/2020.05.30.20117283v1>
-  E. Dantas, M. Tosin and A. Cunha Jr. **Calibration of a SEIR–SEI epidemic model to describe the Zika virus outbreak in Brazil**, Applied Mathematics and Computation, vol.338, p.249-259, 2018. DOI: <https://doi.org/10.1016/j.amc.2018.06.024>
-  M. Begon, M. Bennett, R. G. Bowers, et al. **A clarification of transmission terms in host-microparasite models: numbers, densities and areas**, Epidemiol. Infect., 129, 2002. DOI: <https://doi.org/10.1017/S0950268802007148>

## Referências: artigos e relatórios

-  P. Diaz, P. Constantine, K. Kalmbach, E. Jones and S. Pankavichm. **A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation**, Applied Mathematics and Computation, 324(1): 141-155, 2018.  
DOI: <https://doi.org/10.1016/j.amc.2017.11.039>
-  W. Cota. **Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level**. Scientific Electronic Library Online, 2020.  
DOI: <https://doi.org/10.1590/SciELOPreprints.362>
-  W. Lyra, J. Nascimento, J. Belkhiria, L. de Almeida, P. P. Chrispim and I. Andrade. **COVID-19 pandemics modeling with SEIR(+CAQH), social distancing, and age stratification. The effect of vertical confinement and release in Brazil.**, medRxiv, 2020.  
DOI: <https://doi.org/10.1101/2020.04.09.20060053>
-  Y. Li, X. Liu, and L. Wang. **Modelling the transmission dynamics and control of Mumps in Mainland China**, Int J Environ Res Public Health, 15(1): 33., 2018.  
DOI: <https://doi.org/10.3390/ijerph15010033>

## Referências: sites

- 🌐 Centers for Disease Control and Prevention. **Smallpox.** <[shorturl.at/hm146](http://shorturl.at/hm146)>. Acesso: 31/05/2020.
- 🌐 COVID-19: Observatório Fluminense (COVID19RJ). <<https://www.covid19rj.org/>>. Acesso: 06/07/2020.
- 🌐 Covid Trends. <<https://aatishb.com/covidtrends/>>. Acesso: 06/07/2020.
- 🌐 Grupo de Métodos Analíticos em Vigilância Epidemiológica. <[shorturl.at/ILTW3](http://shorturl.at/ILTW3)>. Acesso: 06/07/2020.
- 🌐 Ministério da Saúde. **Caxumba: o que é, causas, sintomas, tratamento, diagnóstico e prevenção.** <<https://saude.gov.br/saude-de-a-z/caxumba>>. Acesso: 31/05/2020.

## Referências: sites

-  Nature. **Five ways to ensure that models serve society: a manifesto.** <[shorturl.at/uFN67](https://shorturl.at/uFN67)>. Acesso: 04/07/2020.
-  Observatório Covid-19 BR. <<https://covid19br.github.io/>>. Acesso: 06/07/2020.
-  Our World in Data. **Coronavirus Source Data.** <[shorturl.at/hsCOZ](https://shorturl.at/hsCOZ)>. Acesso: 13/07/2020.
-  Towards data science. **Transparency, Reproducibility, and Validity of COVID-19 Projection Models.** <[shorturl.at/luK49](https://shorturl.at/luK49)>. Acesso: 30/06/2020.
-  Wikipédia. **Método dos Mínimos Quadrados.** <[shorturl.at/isASX](https://shorturl.at/isASX)>. Acesso: 04/07/2020.