

Estatística na pandemia

Alexandre Galvão Patriota

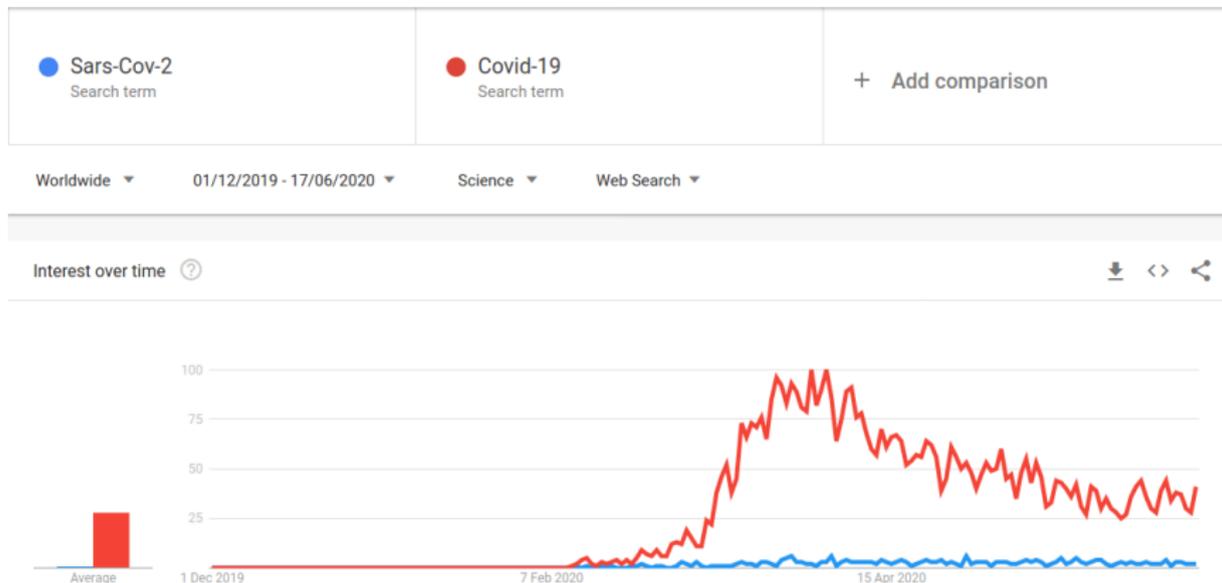
Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo

Sars-Cov-2 (Covid-19)

- Primeiro caso ocorreu em Wuhan, China (início dos sintomas em 1/12/2019).
- Primeiro artigo foi publicado em dezembro de 2019.
- Artigos sobre Sars-Cov-2 (Covid-19) 01/01/2020 a 17/06/2020:
 - 1 Pubmed 6.888 (22.201).
 - 2 MedRxiv 3.289 (4.272).
 - 3 Google Scholar 54.600 (56.100)

Sars-Cov-2 e Covid-19

Os termos começam a aparecer com frequência a partir de 7/02/2020 (google trends)



Estudos

Os estudos se dividem basicamente em

- **experimentais** (com intervenção do pesquisador).
- **observacionais** (sem intervenção do pesquisador).
- Puramente **teóricos**.

Dados de estudos **experimentais** têm menos variabilidade do que de estudos **observacionais**, pois há controle das variáveis.

Quanto mais variabilidade e menos variáveis controladas, menos confiáveis são os dados.

Dados confiáveis são **essenciais** para se obterem conclusões adequadas.

Aplicações da Estatística na pandemia

Estudos **experimentais**:

- eficácia e efeitos colaterais de vacinas e medicamentos.
- potencialidade de Testes (RT-PCR, Sorologia).¹

Estudos **observacionais** (dados disponíveis pelo governo):

- velocidade de propagação da doença, medidas mitigam o contágio.
- curvas de infectados, mortos, recuperados por meio de modelos: SIR, SEIR, Binomial Negativa, BetaBinomial, Séries temporais...

¹"the sensitivity and specificity of the real-time RT-PCR test is not 100%"(Expert Review of Molecular Diagnostics doi: 10.1080/14737159.2020.1757437)

Dados oferecidos pelo Governo

Os sites oficiais que informam casos de Covid-19 são:

- (MS) <https://covid.saude.gov.br/>
- (MS) <https://www.saude.gov.br/boletins-epidemiologicos>
- (Micro dados: Sistema de Informação de Vigilância Epidemiológica da Gripe) <http://plataforma.saude.gov.br/coronavirus/dados-abertos/>
- (Micro dados corrigidos MS) <https://opendatasus.saude.gov.br>
- (Cartórios) <https://transparencia.registrocivil.org.br/especial-covid>

Datas importantes em São Paulo:

- Decreto limitando movimentação em 16/03.
- Quarentena em 24/03.
- Uso obrigatório de máscaras 07/05.
- Rodízio ampliado 11/05 - 17/05.

Algumas características dos dados disponíveis

- 1 O procedimento de testagem gera uma “fila” de mortos.
- 2 Há atrasos na divulgação de óbitos.
- 3 Nos fins de semana e feriados, secretarias trabalham menos.
- 4 Fontes de dados divergentes.
- 5 Dados de óbitos segundo a data da divulgação tem mais variabilidade do que óbitos segundo a data de óbito.
- 6 O passado dos óbitos segundo a data de óbito é variável.

Na prática, o analista de dados escolhe uma fonte de dados para fazer suas análises.

Mensagens contraditórias das autoridades

A informação dispersa e sujeita à grande variabilidade tem enviado mensagens contraditórias para as nossas autoridades:

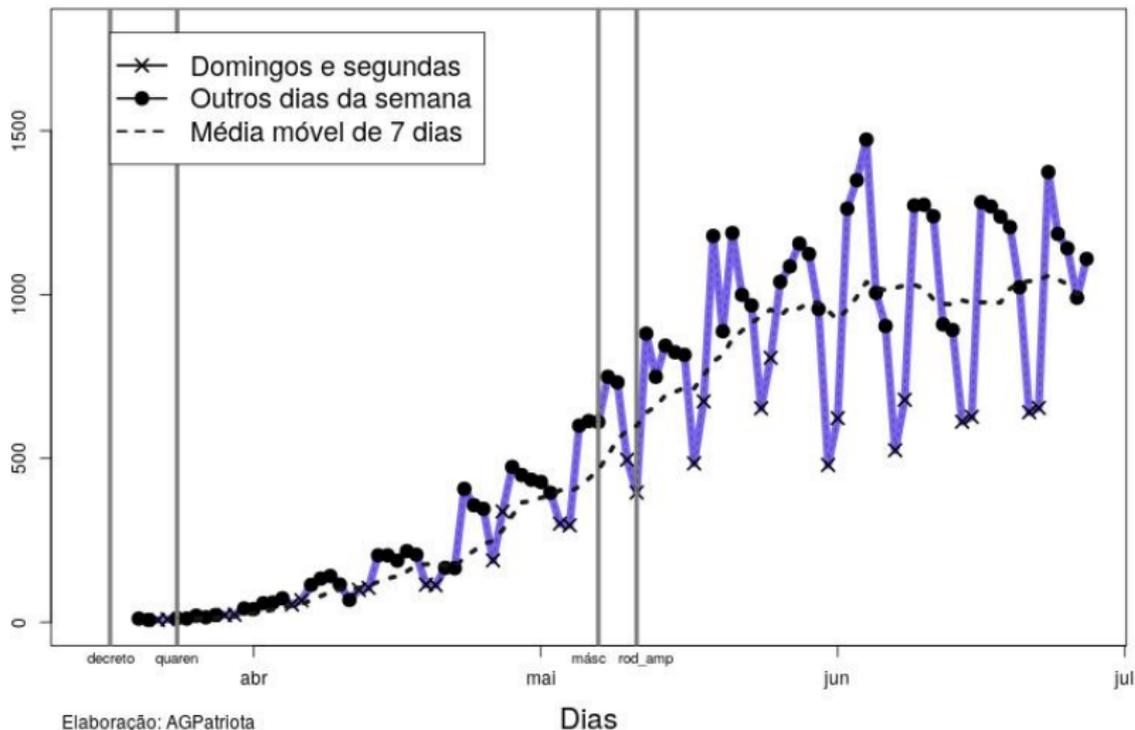
- A OMS **não recomendava** o uso de máscaras pela população.
- A OMS **recomenda** o uso de máscara desde abril.
- A OMS **proibe** pesquisa com hidroxicloroquina (25/05).
- A OMS **libera** pesquisa com hidroxicloroquina (03/06).
- A OMS **proibe** pesquisa com hidroxicloroquina (17/06).

Muitos artigos científicos publicados às pressas (alguns contém falhas metodológicas) sobre o tema fomentam várias posições "não científicas".

Alguns gráficos

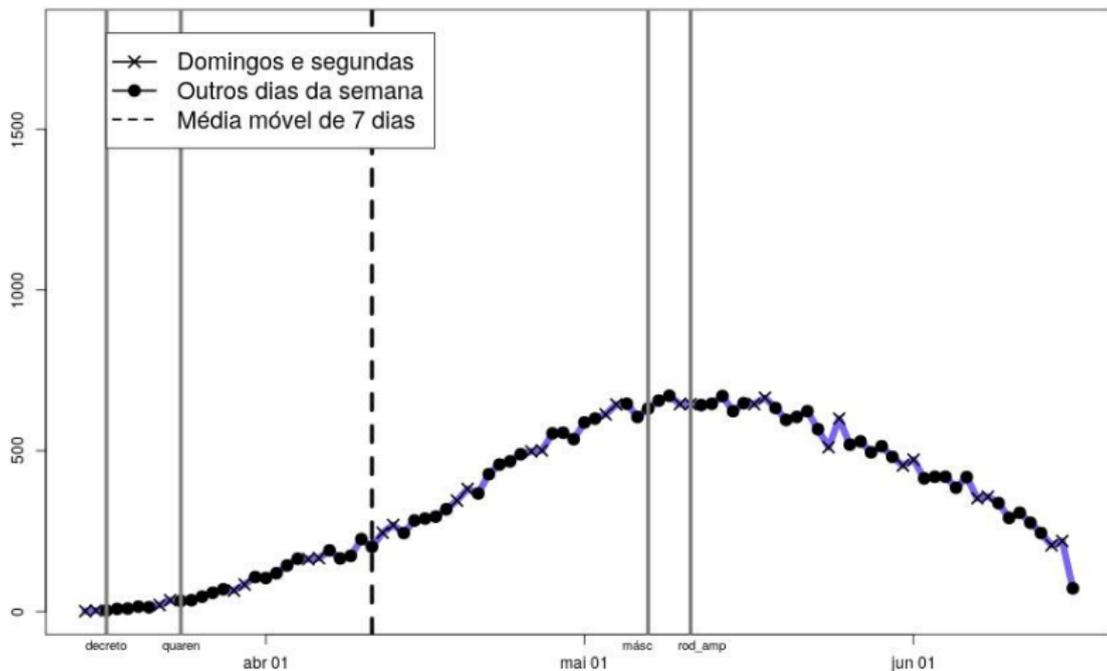
Óbitos na data de divulgação (Brasil)

Óbitos diários segundo data da divulgação (Brasil)



Óbitos na data de óbito (Brasil)

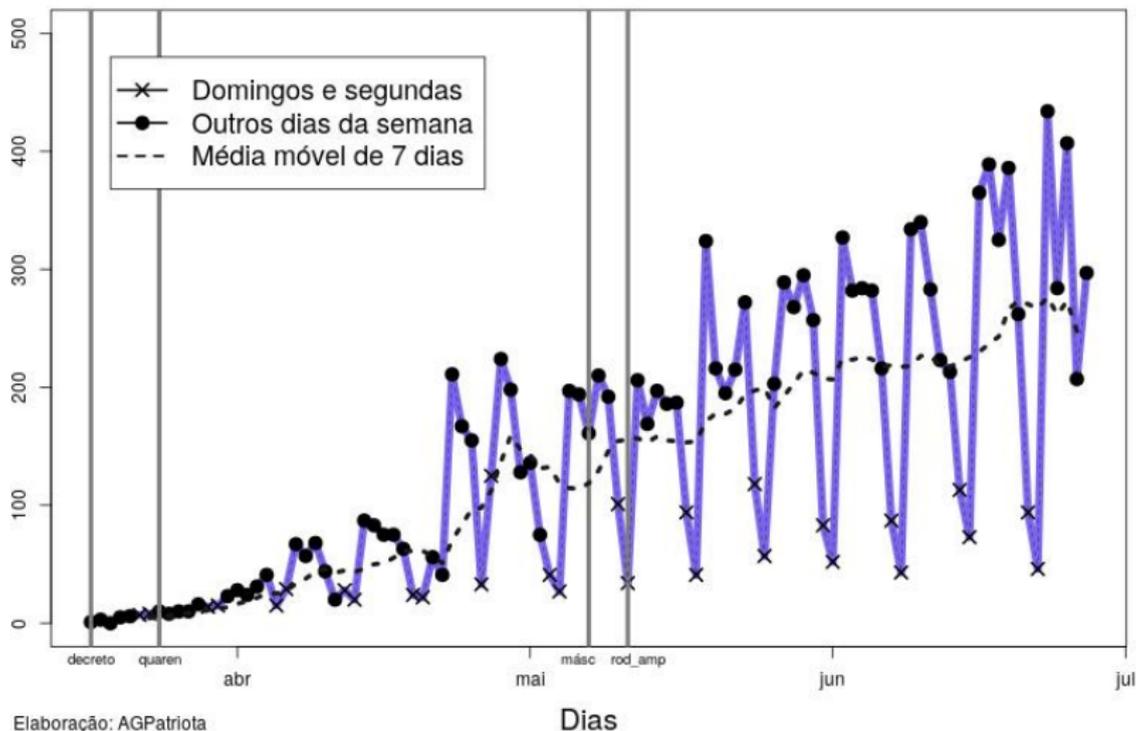
Óbitos por COVID-19 segundo a data do óbito (Brasil)



Fonte: Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

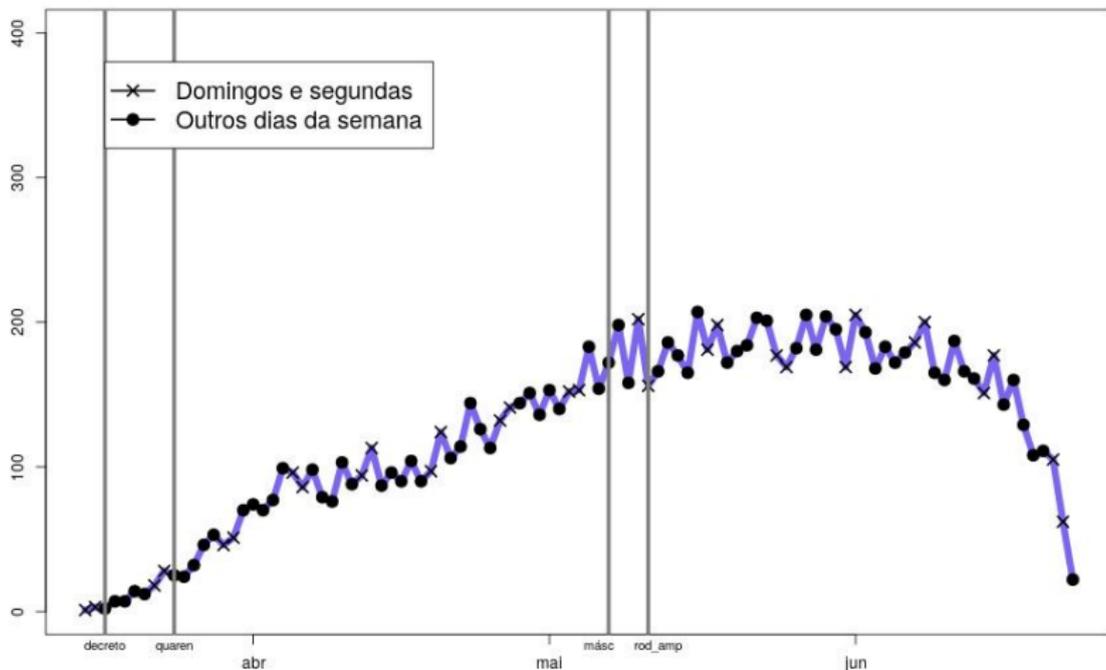
Óbitos na data de divulgação (Estado de SP)

Óbitos diários segundo data da divulgação (Estado de São Paulo)



Óbitos na data de óbito (SP)

Óbitos por COVID-19 segundo a data do óbito (SP)



Fonte: Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

Variabilidade

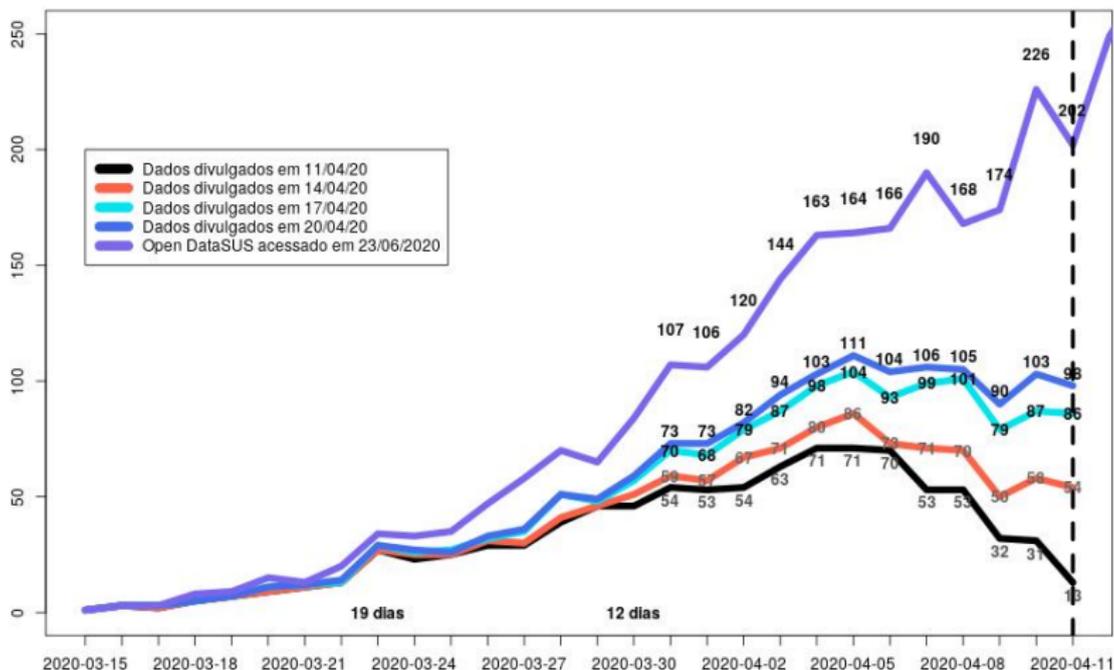
Observamos que para os dados obtidos segundo a data de óbito:

- há menos variabilidade.
- não parece haver efeito de fins de semana na contagem.
- podem ser utilizados para comparar intervenções, pois o marcador tempo estará alinhado.

Devemos utilizá-los para fazer as inferências estatísticas?
(passado distante x passado próximo)

Óbitos segundo data do óbito (Brasil)

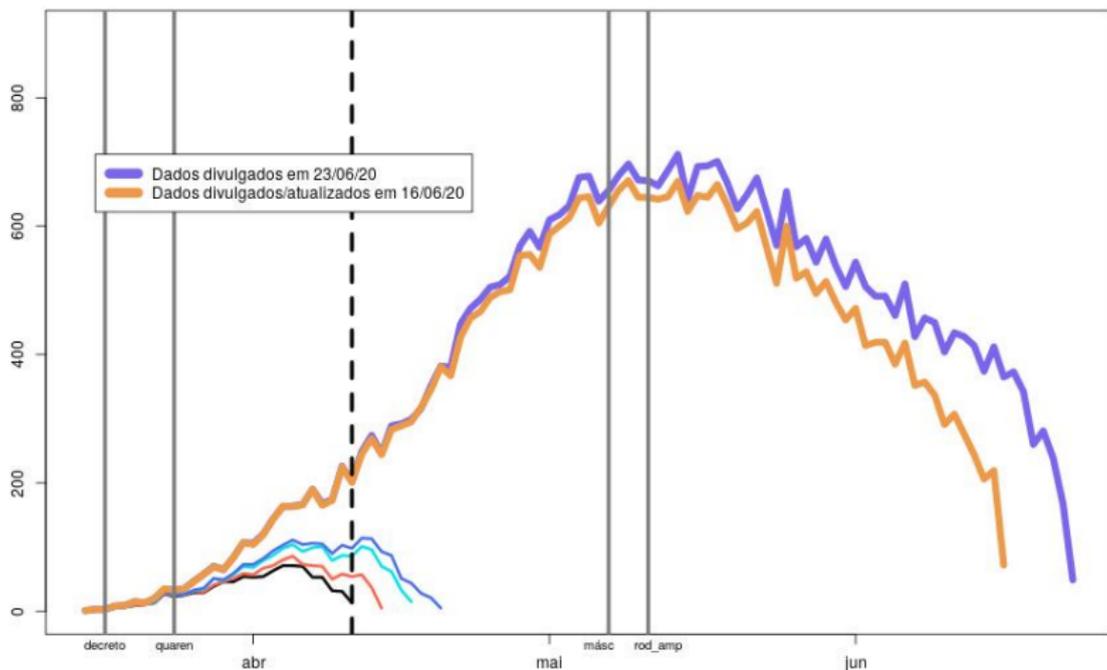
Óbitos por COVID-19 segundo data do óbito (Brasil: 15/03 a 11/04)



Fonte: Boletins epidemiológicos. Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

Óbitos segundo data do óbito (Brasil)

Óbitos por COVID-19 segundo a data do óbito (Brasil)



Fonte: Ministério da Saúde e OpenDataSUS

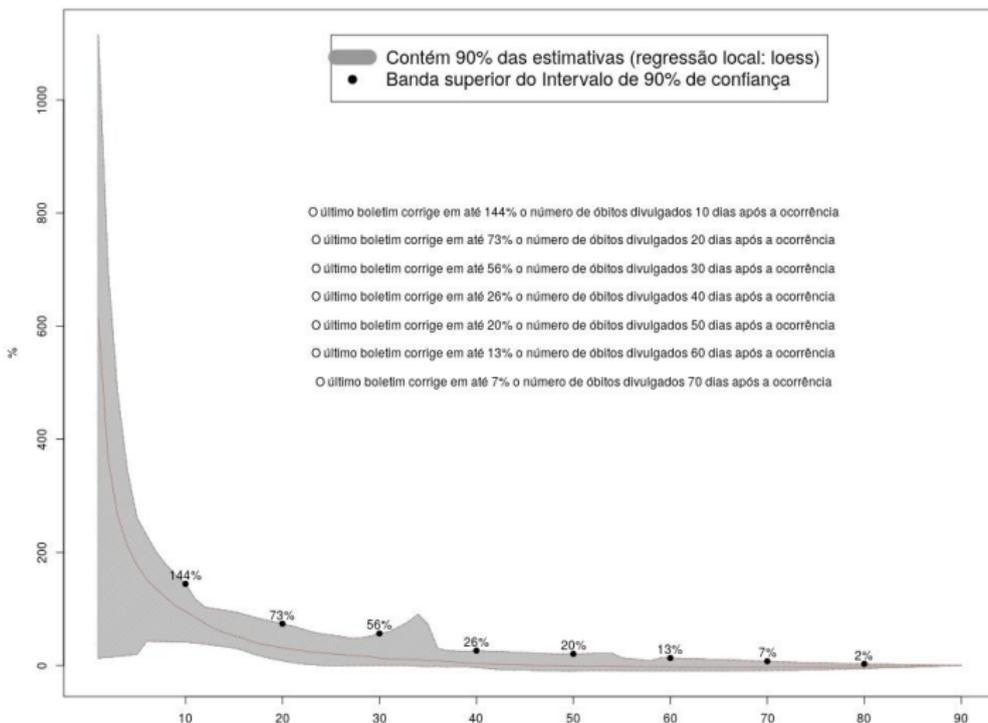
Elaboração: AGPatriota

<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>



Óbitos segundo data do óbito (Brasil)

Correção em relação ao boletim mais recente (23/06/2020)



Fonte: Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

Dias decorridos desde a data do óbito

Atrasos

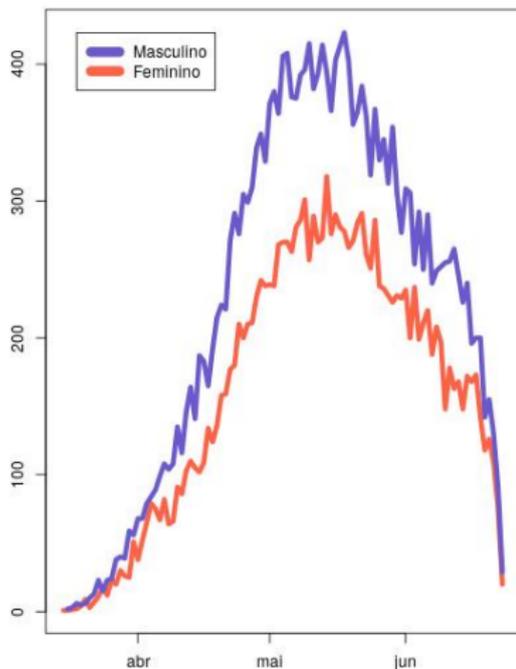
- Há atrasos na liberação dos resultados.
- O passado é atualizado.
- O atraso o atraso se mantém constante com o tempo?

Deve-se ter cuidado ao utilizar esses dados.

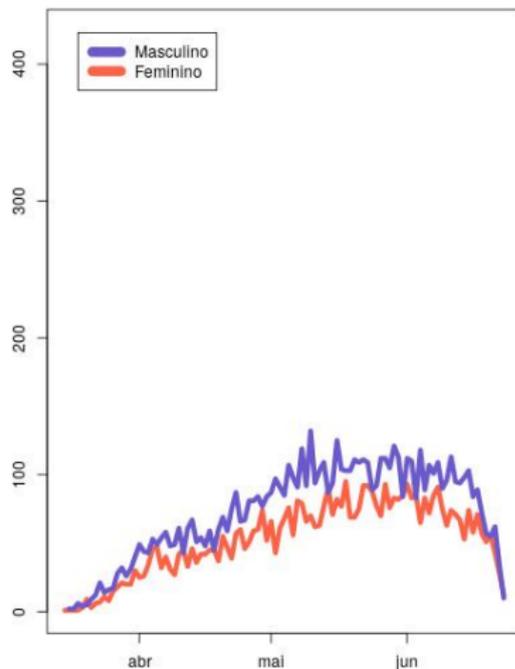
Estatísticas por Sexo, Idade e alfabetização

Óbitos segundo data do óbito e sexo (Brasil e SP)

Óbitos por COVID-19 segundo a data do óbito e sexo
Brasil



Óbitos por COVID-19 segundo a data do óbito e sexo
Estado de São Paulo

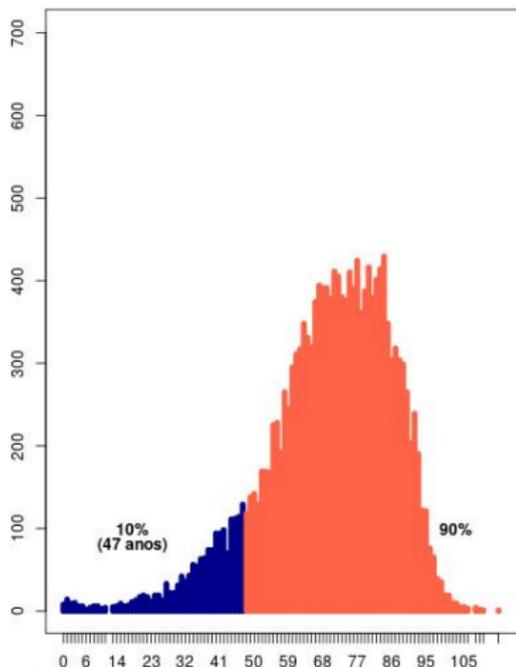


Fonte: Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

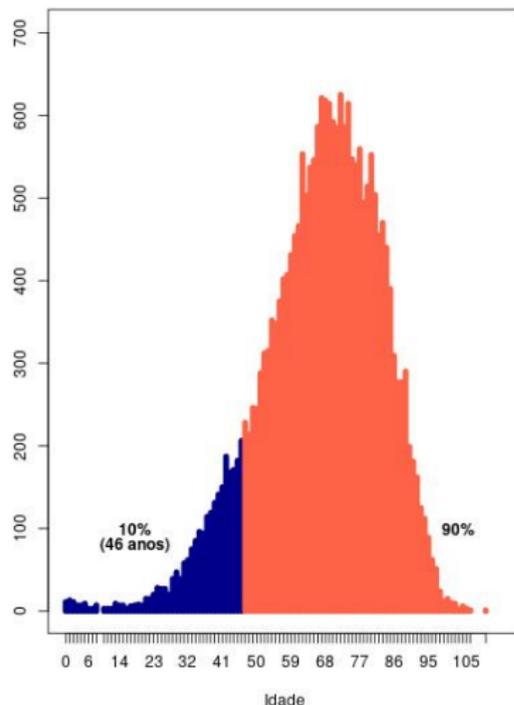
<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Óbitos segundo data do óbito, idade e sexo (Brasil)

Óbitos por COVID-19 segundo a idade (Brasil)
Sexo Feminino



Óbitos por COVID-19 segundo a idade (Brasil)
Sexo Masculino

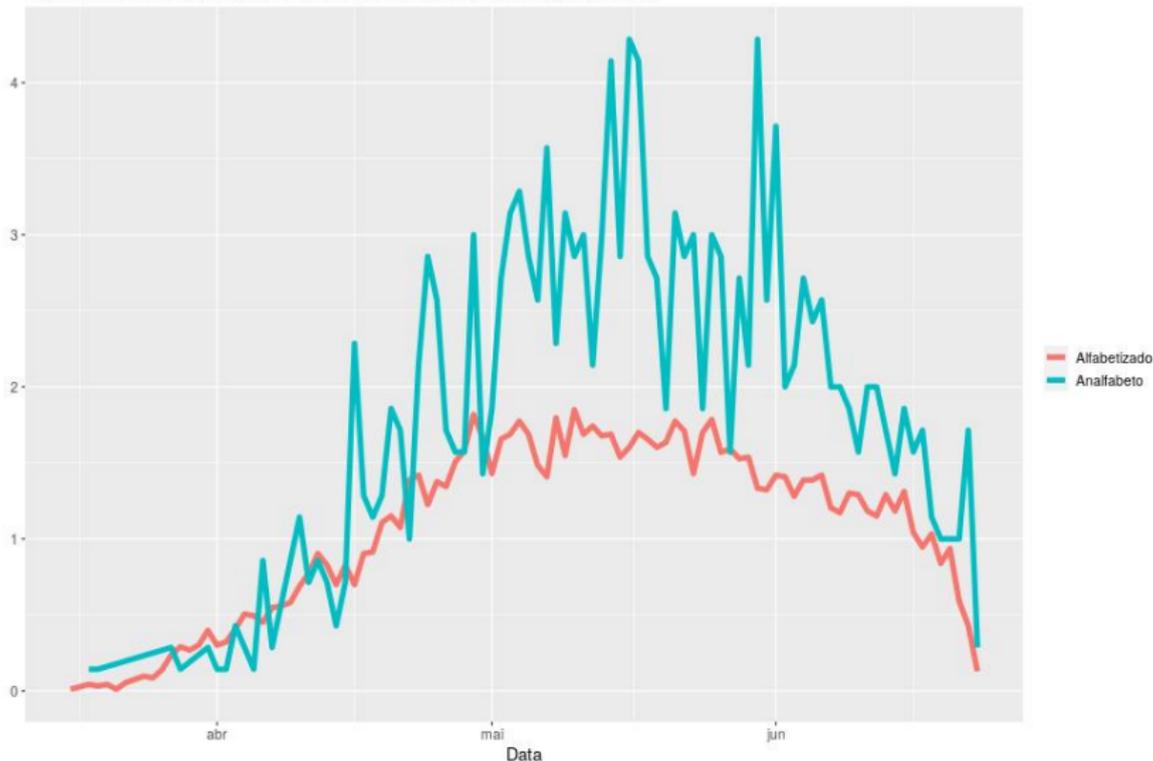


Fonte: Ministério da Saúde e OpenDataSUS
Elaboração: AGPatriota

<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Óbitos na data do óbito e alfabetização (Brasil)

Óbitos (por mil hab) segundo a data de óbito e alfabetização (Brasil).



<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Efeitos

Os gráficos mostram que

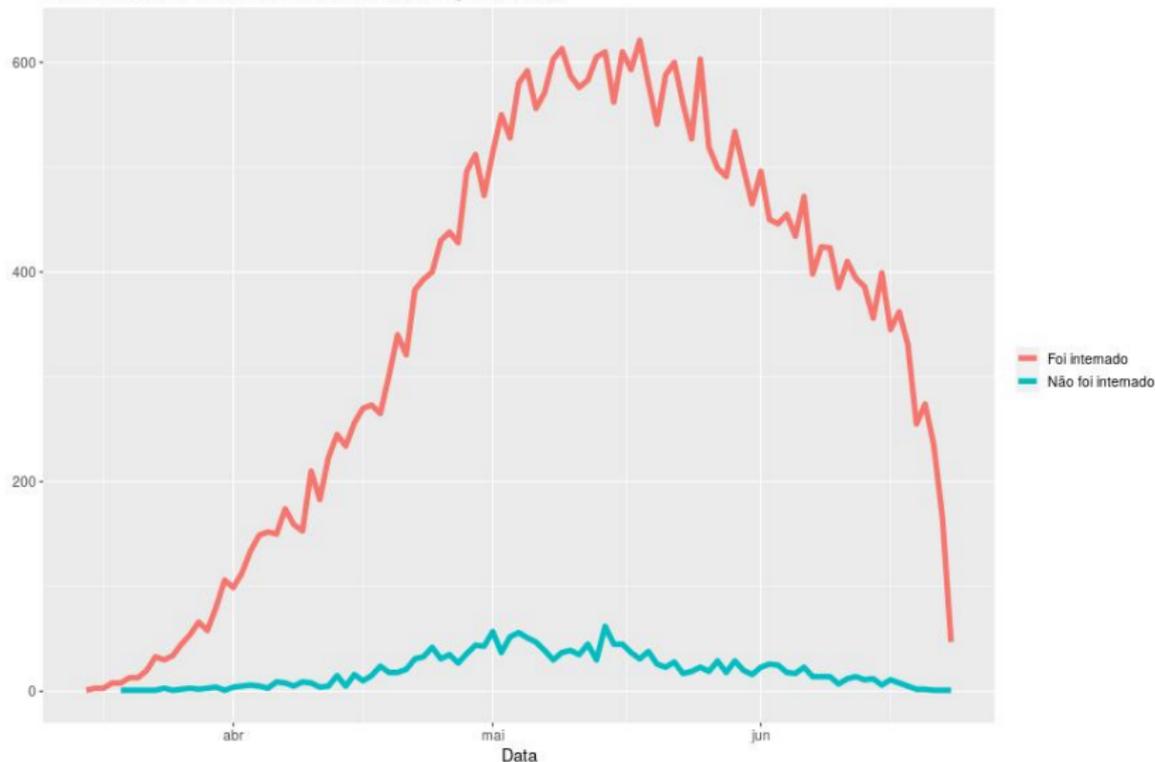
- há efeito da idade.
- possivelmente há efeito do sexo (pois homens têm mais comorbidades?).
- possível efeito da escolaridade (que também pode estar relacionado com a renda familiar e número de moradores na mesma casa).

O problema em se fazer testes estatísticos é que os dados tem problemas sérios.

Internações

Óbitos na data do óbito e hospitalização (Brasil)

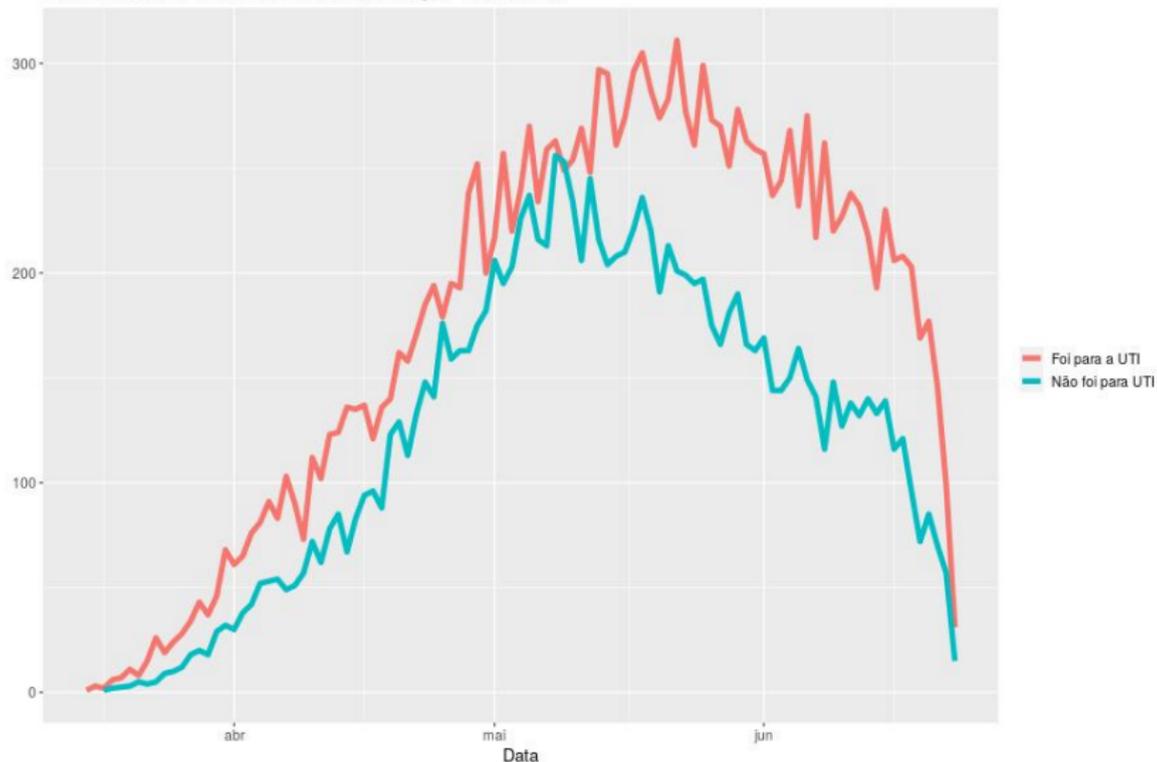
Óbitos segundo a data de óbito e hospitalização (Brasil).



<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Óbitos na data do óbito e internação UTI (Brasil)

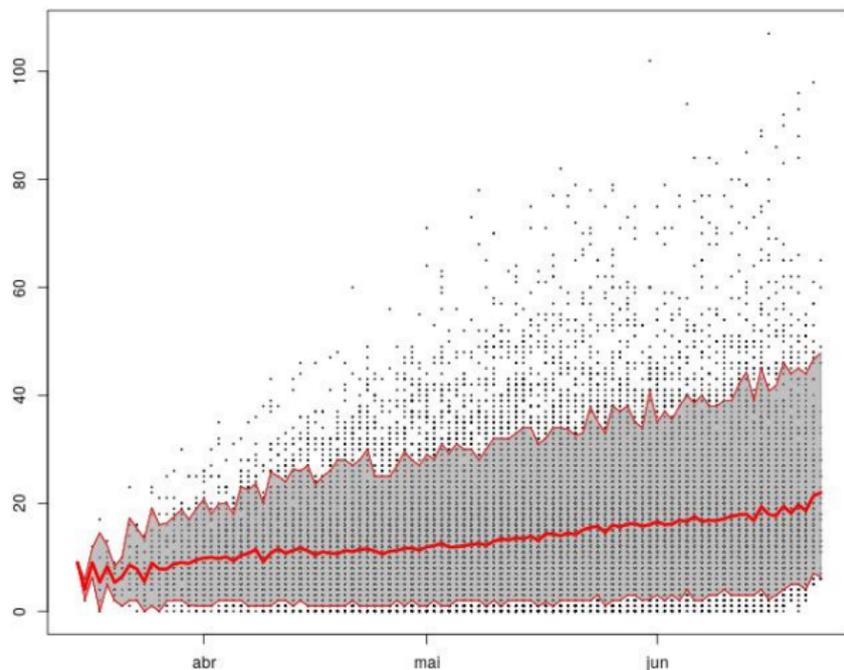
Óbitos segundo a data de óbito e Internação UTI (Brasil).



<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Tempo óbito desde primeiros sintomas (Brasil)

Tempo decorrido dos primeiros sintomas ao óbito



Fonte: Ministério da Saúde e OpenDataSUS

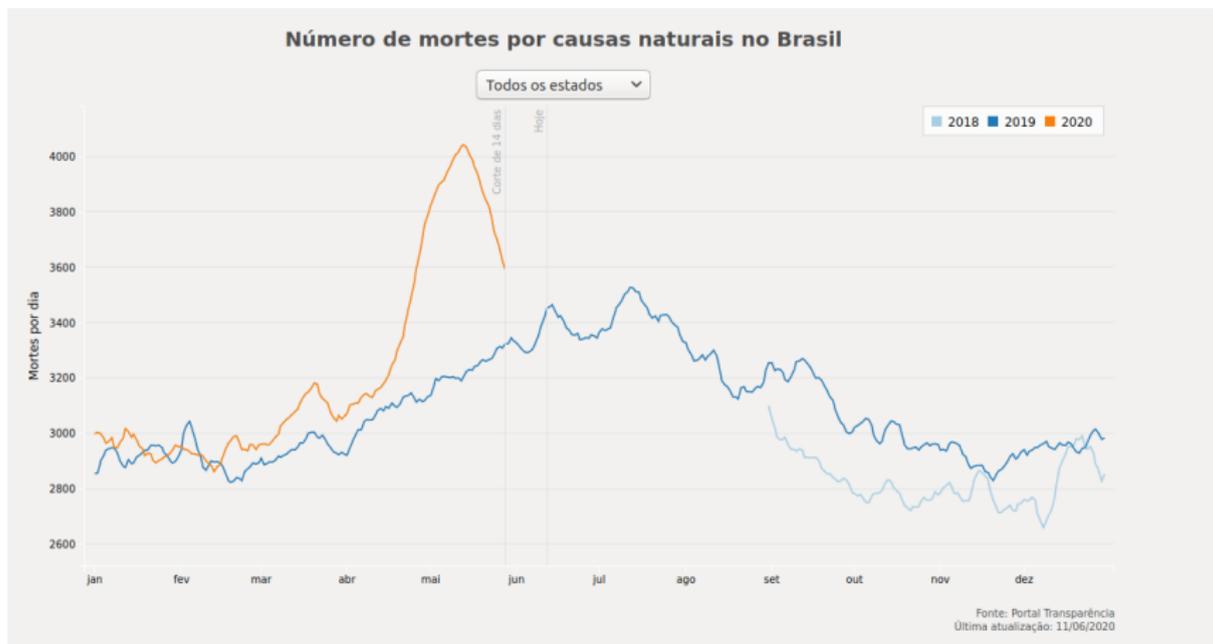
Data de óbito

Elaboração: AGPatriota

<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

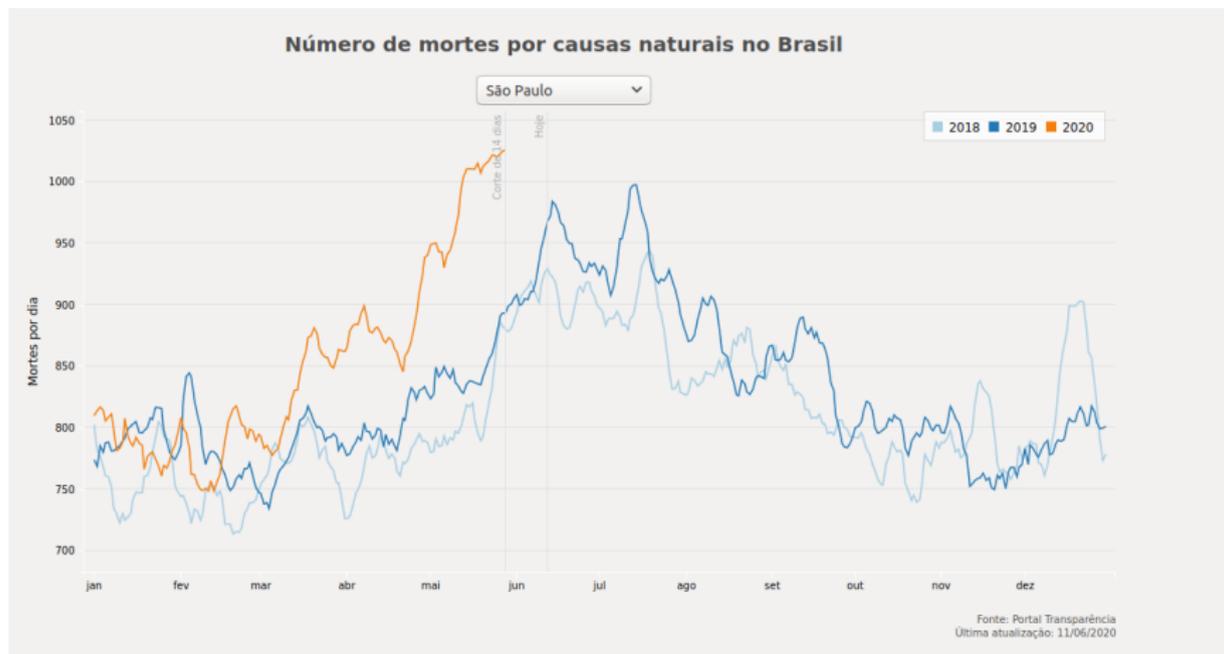
Mortes naturais (2020 x 2019)

Mortes Naturais (Brasil)



<https://mortalidade.com.br/>

Mortes Naturais (SP)



<https://mortalidade.com.br/>

Estatística

Estatística

A estatística é uma linguagem formal para lidar com a incerteza.

Karl Pearson disse “*Statistics is the grammar of science*”.

The grammar of Science:

<https://archive.org/details/grammarofscience00pearuoft>

Sem dados confiáveis, as inferências estatísticas têm pouca validade.

Escassez de dados confiáveis (ou excesso de dados problemáticos) degradam os resultados científicos e fomentam narrativas não científicas.

Modelos estatísticos

Os principais modelos estatísticos usados são:

- SIR e SEIR.
Variáveis: número de Suceptíveis, Expostos, Infectados, {Recuperados, Mortos}.
- Modelo Beta Binomial:
Variável: número de Mortos (explicado por região, UITs, leitos).
- **Modelo Binomial Negativo** (Imperial College):
Variável: número de mortos (explicado por região, UTIs, leitos).
- Outros modelos de séries temporais multivariadas que admitam covariáveis.

Potencialidades e limitações dos modelos

Os modelos acima nos ajudam a:

- explicar/descrever o que aconteceu no passado (por meio de alguma relação entre as variáveis).
- verificar se medidas mitigaram a taxa de contágio e os óbitos (*ceteris paribus*).
- projetar número de infectados, mortos e recuperados para um futuro próximo.

Entretanto,

- há subnotificações, erros de classificação e amostras enviesadas,
- falta informação sobre ocupação de UTIs, leitos respiradores,
- há muitas variáveis que afetam os resultados (idade, comorbidades pré-existentes, etc).
- o que não é observado é extremamente relevante

Eu não faço projeções de casos/óbitos

- Os dados não são confiáveis,
- os dados sobre mortalidade disponíveis até o ano de 2018 no Datasus,
- muitos fatores não conhecidos,
- a variabilidade real desconhecida.
- as curvas estimadas variam muito mesmo sob as premissas.

A confiabilidade nas inferências estatísticas depende dos dados e da complexidade das relações entre as variáveis.

Garbage in, garbage out.

Comentários:

- 1 Causalidade não pode ser (com)provada por meio de testes estatísticos.
- 2 Podemos falar sobre efeitos estatísticos (em termos de probabilidades e médias).
- 3 Inserir no modelo as principais fontes de incerteza (impossível),
- 4 Estudar o que ocorre com as inferências sob premissas falsas
- 5 Buscar modelos robustos a perturbações nas premissas,

O cientista deve deixar claro quais são as limitações dos modelos e das inferências para evitar fomentar mais desinformação²

²Imperial College paper: a palavra “assume/assumption” foi usada 27 vezes, nada sobre as limitações dos métodos foi dito.

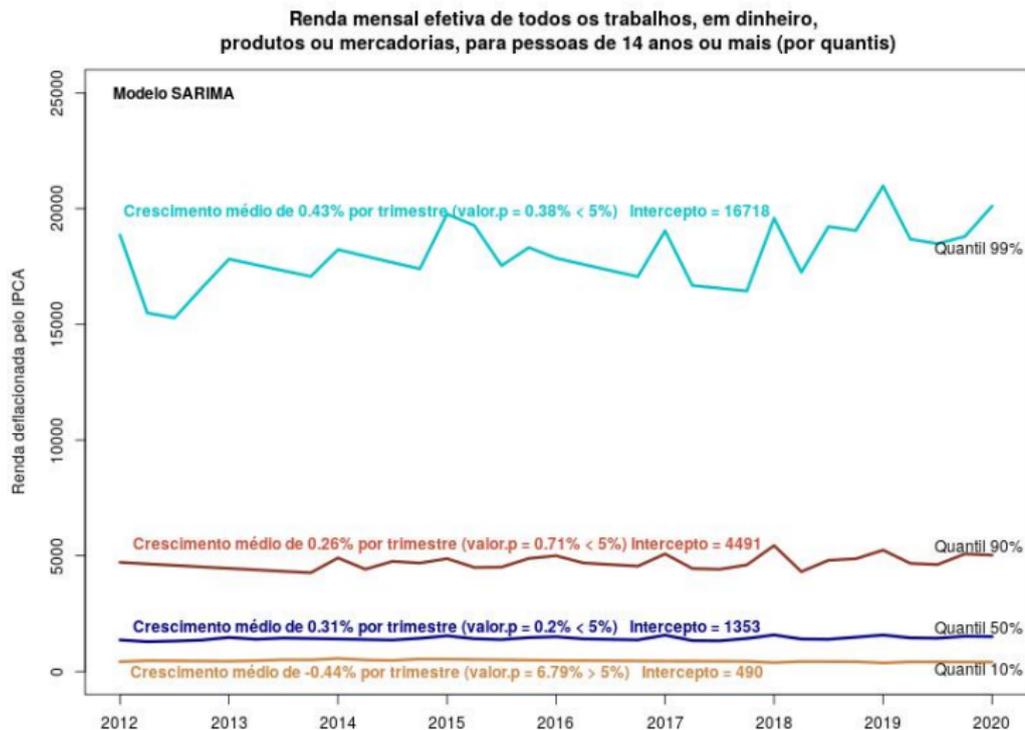
Por fim:

- 7 Fatores sócioeconômicos devem ser levados em conta.
- 8 Qual será o maior número de mortos nos próximos 5 anos: por Covid ou por desnutrição?

Em meio a tanta variabilidade e incerteza, as análises tem valor marginal. Apenas coisas que são evidentes podem são confiáveis dos modelos.

Saberemos o que de fato aconteceu em análise retrospectiva.

Renda mensal efetiva em 2019



Elaboração: agpatriota
Fonte dos dados: IBGE

Método científico e estatística

Em geral, os seguintes passos (não necessariamente nessa ordem) são adotados:

- 1 postulam-se as hipóteses científicas,
- 2 definem-se as quantidades de interesse,
- 3 executa-se o experimento para coletar os dados,
- 4 aplicam-se os métodos estatísticos,
- 5 conclue-se sobre a plausibilidade das hipóteses (*ceteris paribus*).

A conclusão em 5 precisa ser inferida para a hipótese científica (modus tollens: se $(A \implies B) \wedge (\neg B)$, então $\neg A$).