



**UNIVERSIDADE FEDERAL DE SERGIPE
DEPARTAMENTO DE FÍSICA
GRUPO DE MECÂNICA ESTATÍSTICA**

**Nota técnica AMCS 03:
A testagem da população na pandemia do Covid-19**

Edição: 20/05/2020

Elaboração: Prof. André Maurício Conceição de Souza

Índice:

1. Introdução	2
2. Amostras sem viés	3
3. Amostras com viés	7
4. Amostras para o Covid-19.....	10
5. Comentários finais	12
6. Referências	13
7. Agradecimentos	13

Contato: amcsouza@ufs.br

Aracaju, maio de 2020.

1. Introdução

O número real de infectados pelo Covid-19 é um dos pontos controversos nesta pandemia em todo o mundo. O consenso da quase totalidade no meio científico é que há uma grande subnotificação de casos. Até mesmo as mortes por Covid-19, declaradas pelos sistemas oficiais de saúde, são contestadas por diversos artigos e reportagens^[1]. O caso de Manaus é um exemplo, a média de mortes chega a quatro vezes maior que a média dos anos anteriores neste período, e os casos notificados de mortes pelo Covid -19 não chega a ser 10% desses valores^[2]. São diversos os argumentos apresentados para discordar dos números oficiais, tais como: falta ou pouca testagem, demora de resultado dos testes, omissão de resultado positivos por interesse político. Como este tem sido um assunto bastante discutido, abordamos nesta nota técnica, a partir de princípios de estatística, como se define o número de testes e a quem testar.

Aqui, analisamos a questão do número de casos reais em função do número de testes e do viés estatístico. Demonstra-se como, certamente, os números oficiais estão subdimensionados, isto é, o número de infectados pelo Covid-19 é maior que o divulgado oficialmente. Concluiremos esta nota técnica apresentando uma série de problemas que vem ocorrendo na metodologia de testagens e no cálculo do número de infectados pelo Covid-19 em todo o mundo.

Gostaríamos de enfatizar que esta nota não representa uma introdução formal ao estudo de espaço amostral, muito menos tem a pretensão de ser um curso de Estatística^[3,4,5]. Nosso objetivo é mostrar para o público geral como uma amostra é escolhida e o que se pode concluir sobre o que se passa na população. No caso da pandemia do Covid-19, como escolher as pessoas a serem testadas e assim inferir a quantidade de pessoas infectadas.

2. Amostras sem viés

O público em geral está acostumado a acompanhar durante o período eleitoral pesquisas em que se projeta o resultado de uma eleição para populações de milhares ou milhões de eleitores, a partir de uma amostragem de 500 a 2000 pessoas. Essas pesquisas balizam as estratégias dos candidatos, e, no final, na maioria das vezes, concordam bastante com o resultado das eleições. Além do valor da projeção de votos, o público mais atento observa que são apresentadas mais duas variáveis: o grau de confiança e a margem de erro. O grau (ou intervalo) de confiança indica a probabilidade de uma pesquisa ter seus resultados dentro da margem de erro apresentada. Pesquisas eleitorais no Brasil estabelecem o grau de confiança em 95%. O que quer dizer que se fizermos 100 pesquisas com o mesmo objetivo, em 95 delas os resultados estarão dentro da margem de erro. A margem de erro significa o intervalo de valores em que o instituto de pesquisa garante estar o valor apresentado para a votação de um certo candidato. Assim, diz-se que o candidato teve 30% de intenção de votos com margem de erro de 2% para mais e para menos, portanto, o candidato está com intenção de votos entre 28% e 32%.

A ciência que baliza essas pesquisas é a estatística. A escolha da amostra (usa-se a palavra amostra para o conjunto de pessoas escolhidas para fazer parte da pesquisa) é feita fundamentada pela estatística a partir da teoria matemática^[3,4,5]. Além de pesquisas eleitorais, esta teoria pode ser aplicada numa incrível quantidade de diferentes situações. Vamos citar duas completamente diferentes: i) saber a chance de um palito não acender, naturalmente, sem testar todos os palitos em uma fábrica de palitos de fósforos. ii) a partir da testagem de um conjunto de pessoas verificar a quantidade total de pessoas infectadas pelo Covid-19. Como veremos, é muito mais fácil analisar a chance de fósforos acender, ou mesmo prever o resultado de uma eleição que saber quantas pessoas estão infectadas pelo Covid-19.

A seguir vamos apresentar um pequeno estudo de como são feitas essas escolhas.

Primeiramente devemos ter certeza da representação da amostra. Isto é, os palitos escolhidos devem representar as mesmas características de todos os palitos fabricados. As pessoas escolhidas para a pesquisa eleitoral devem representar toda população do local. A forma mais simples de escolha é a que chamamos de aleatória. Para este tipo de escolha temos uma teoria matemática sólida de como escolher os indivíduos e o tamanho da amostra. Vejam o caso dos

palitos. Imaginemos que temos um milhão de palitos e vamos pegar uma amostra de mil palitos para testar. Podemos escolher os mil palitos de várias formas, os mil primeiros ou os últimos, ou em outra sequência qualquer, entretanto, considerando que não temos nenhuma informação sobre o que pode ter se passado com a máquina na sequência de produção, nossa melhor opção é escolher aleatoriamente os mil palitos. Dizemos nesse caso que é uma amostra sem viés. Imagine que na pesquisa 900 palitos funcionaram normalmente e os outros 100 palitos não acenderam. O que podemos dizer sobre o milhão de palitos? Podemos fazer uma regra de três e facilmente observar que de um milhão 900.000 acenderão e 100.000 não. Correto? Em parte. Primeiro não devemos afirmar que acenderão 900.000 palitos, e sim, afirmar que em média acenderão 900.000 palitos. Em seguida, temos que completar com o grau de confiança e com a margem de erro. Para este caso, da teoria estatística, encontramos para um grau de confiança de 95% que a margem de erro é 3,1%.

Em geral, para uma amostra sem viés, se temos T testes positivos, o número médio de casos na população (P_{med}) será igual ao produto de T vezes o tamanho da amostra dividido pelo tamanho da população, isto é, $P_{med} = T \times \text{tamanho da amostra} / \text{tamanho da população}$. A fórmula para a margem de erro é mais complicada e não apresentaremos nesta nota. Entretanto, podemos simplificá-la e dizer que a margem de erro (E%) em termos percentuais é proporcional ao inverso do tamanho da amostra. Isto é, quanto maior a amostra menor a margem de erro. Em primeira aproximação, para um grau de confiança de 95%, $E\% = 98 / \sqrt{\text{tamanho da amostra}}$. Usando as fórmulas temos $P_{med} = 900 \times 1.000.000 / 1.000 = 900.000$ e $E\% = 98 / \sqrt{1000} = 3,1\%$. Isto é, o número de palitos que irão acender estará entre 869.000 a 931.000 mil palitos de fósforo. Para ilustrar o que foi dito até o momento, fizemos uma simulação no computador de 1 milhão de palitos de fósforos. Assumimos que 900 mil acendem e 100 mil não. Em seguida, pegamos aleatoriamente 1.000 palitos e testamos se acendem. Na Figura 01 apresentamos um gráfico com o número de palitos que acendem na amostra de 1.000 palitos repetindo o teste 100 vezes (vantagem de fazer simulação no computador, pois na realidade perderíamos todos os palitos). Obtivemos uma média $T=901$. Na figura 02, utilizamos esses resultados e aplicamos as fórmulas de P_{med} e E% para calcular o número de palitos que acendem quando temos os 1 milhão de palitos. Indicamos em linha tracejada as margens de erro e em linha cheia a média teórica. Naturalmente a simulação está de acordo com a teoria. É interessante observar as flutuações nos resultados simulados. A teoria nos dá um número, enquanto que a simulação no computador, que deve refletir a realidade, os números flutuam em torno da teoria.

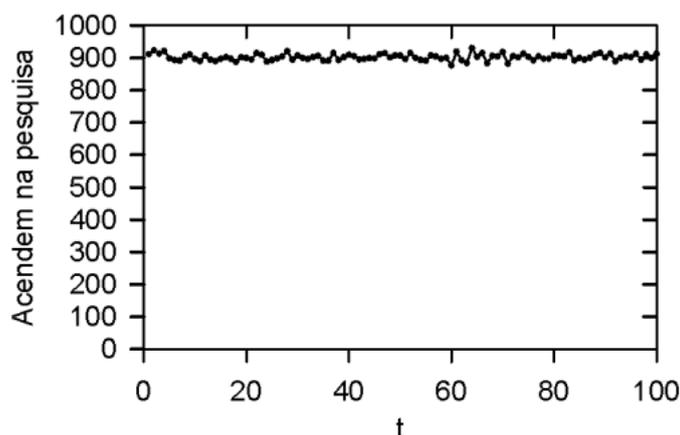


Figura 01: Número de palitos de fósforos que acendem numa amostra de 1000 palitos escolhidos aleatoriamente numa população de 1 milhão de palitos. A simulação é repetida 100 vezes.

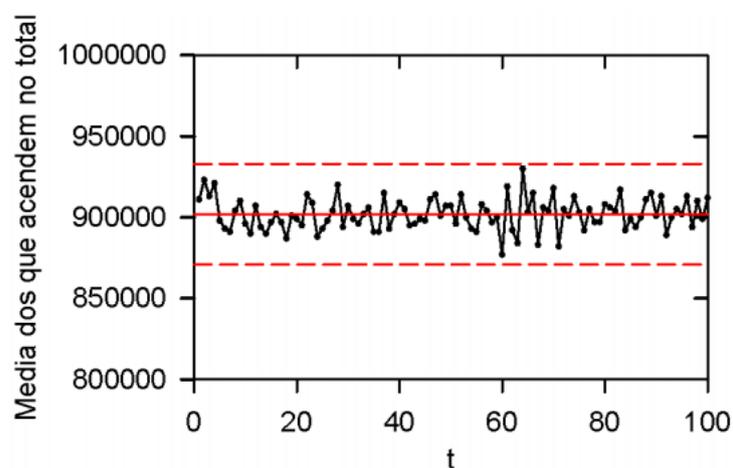


Figura 02: Número médio de palitos de fósforos que acendem numa população de 1 milhão de palitos, a partir dos dados da simulação da figura 1. A simulação é repetida 100 vezes. A linha vermelha cheia indica o valor teórico da média e as linhas vermelhas tracejadas indicam a margem de erro teórica calculada pela fórmula apresentada no texto desta nota.

Para o caso de uma pesquisa eleitoral a escolha da amostra é parecida, também se usa a escolha sem viés, entretanto, é necessário estratificar a amostra. Ao contrário dos palitos de fósforo, em que todos têm as mesmas características, numa população de pessoas numa cidade, caso se faça a escolha completamente aleatória não estaremos levando em conta o sexo, a escolaridade, a renda, o local de moradia. Sabe-se que cada uma dessas características leva a um perfil de escolha de candidato diferente. Assim, antes de fazer a escolha aleatória, separa-se a amostra em proporções de pessoas com estes extratos, e dentro destes escolhe-se aleatoriamente. Com isso, podemos usar as fórmulas apresentadas no caso dos palitos, pois estamos tratando de escolha sem viés.

Como exemplo, imaginemos que fizemos uma pesquisa estratificada com 400 pessoas em Sergipe e um candidato obteve 100 indicações. Assim, assumindo que votam 1.600.000 pessoas em Sergipe, pelas fórmulas, o candidato terá uma média de $P_{med} = 100 \times 1.600.000 / 400 = 400.000$ votos para uma margem de erro de $E\% = 98/20 = 4,9\%$ votos. Portanto, o candidato deve receber dos eleitores entre 321.600 e 478.400 votos.

Salientamos que, do ponto de vista científico, não se pode simplesmente apresentar um resultado de uma pesquisa amostral indicando apenas a média obtida. Para apresentá-la corretamente, além da média, tem-se que indicar o grau de confiança e a margem de erro da pesquisa.

3. Amostras com viés

Em princípio podemos escolher amostras sem viés para todos os problemas, com o cuidado apenas de escolher extratos que caracterizem os diferentes perfis do que se quer analisar. É o caso da pesquisa eleitoral que separamos a amostra de pessoas em sexo, idade, renda e local de moradia. Entretanto, observa-se que há um enorme problema quando se quer analisar uma situação em que temos na população um número muito pequeno de casos. Imagine, na eleição que tratamos na seção anterior, se houvesse um candidato que somente uma pessoa tivesse indicado que votaria nele. Pelas fórmulas, o candidato terá uma média de $P_{med} = 1 \times 1.600.000 / 400 = 4.000$ votos (0,29% dos votos) para uma margem de erro de $E\% = 4,9\%$, isto é, 78.400 votos. Assim, o candidato receberia dos eleitores entre 0 e 82.400 votos. Isto é, a margem de erro é cerca de 20 vezes a média dos votos do candidato!

Chegamos ao primeiro grande problema dos testes para o Covid-19. Como o número de infectados ainda representa uma pequena parcela da população, a margem de erro dos resultados é do tamanho ou maior que o resultado! Imagine que de 400 testes, apenas uma pessoa seja positiva. Se usarmos as fórmulas, como no caso acima, o que podemos afirmar é que entre 1 e 82.400 pessoas estão infectadas e que em média 4.000 estão infectadas. De apenas um caso positivo, devemos ter entre 1 e 82.400 infectados. Este resultado é muito forte e determina porque não devemos usar amostras sem viés nesta fase de pandemia.

Nestes casos o que fazer? As amostras devem ser escolhidas com viés. Isto significa que não se deve fazer escolha aleatória das pessoas. A escolha das pessoas agora depende de um protocolo. E nestes casos não temos uma teoria estatística geral. Para cada protocolo de plano amostral se deve preparar uma estatística que irá inferir os dados na população.

Diante do que foi exposto até o momento, fica claro que o número de casos positivos nos testes para detectar o Covid-19 não representa em absoluto o número de pessoas infectadas com o vírus! Vimos que se as amostras fossem sem viés, apenas um positivo em 400 testes, poderia representar milhares de pessoas infectadas em Sergipe. Somente se o protocolo de testes garantisse que todos os infectados estão dentro da amostra é que teríamos uma concordância entre positivos e infectados, isto é uma situação quase que improvável. Como escolher um protocolo que garanta que estamos testando todos os possíveis infectados? Somente se tivéssemos uma amostra do tamanho da população. Certamente isto não acontece, ao contrário,

sabemos que no Brasil o número de testes tem sido relativamente muito pequeno.

Antes de analisar os protocolos reais para o Covid-19, para fins didáticos, como na seção anterior, apresentamos agora uma simulação computacional de uma enfermidade. A enfermidade acomete parte da população e não muda no tempo, imaginemos que demora muito de curar, mas que não afeta gravemente a saúde da população. Dos doentes, 50% são sintomáticos e 50% assintomáticos. Vamos imaginar uma população de 2.290.00 habitantes (similar a população de Sergipe) e que 1000 pessoas estão infectadas.

Passemos agora para a escolha da amostra, isto é, como faremos os testes na população. Vimos que fazer amostras sem viés não funciona, pois 1000 infectados representam apenas 0,04% da população. Se fizermos testes com 1000 pessoas, a margem de erro seria de 71.000 pessoas. Assumimos, que no nosso protocolo os pacientes sintomáticos têm 100% de chance de serem escolhidos nos testes e os assintomáticos, 50%. As pessoas sem a enfermidade são testadas aleatoriamente. A Figura 03 mostra a evolução de casos positivos, caso sejam feitos 1000 testes por dia. É fácil ver que o número de infectados flutua em torno de 750 pessoas. Pelo nosso protocolo, este resultado é simples de ver. Como são 1.000 infectados, são 500 sintomáticos e 500 assintomáticos. Como fazemos um mil testes e o protocolo indica que escolhemos 100% de infectados (500) e 50% de assintomáticos (250), o que dá uma média de 750 testes positivos. Portanto, dos 1000 infectados, somente 750 são testados positivos diariamente e assim, como esperado, o número de infectados é maior que de testes positivos. Neste caso, com o uso deste protocolo, 1,33 vezes maior que o de testes positivos.

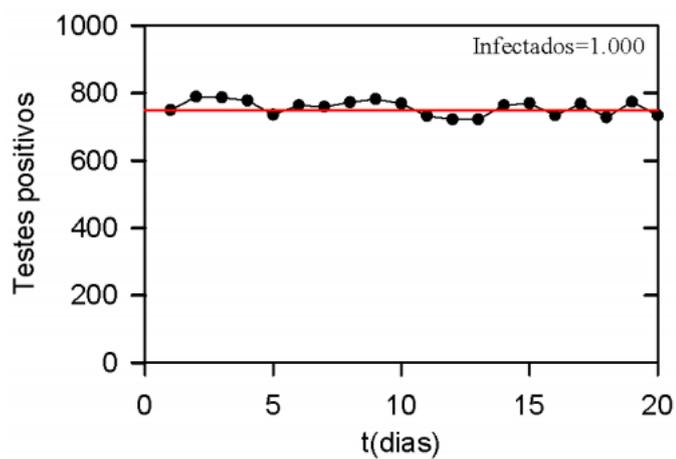


Figura 03: Número de testes positivos para uma enfermidade com 1000 infectados, sendo 500 sintomáticos e 500 assintomáticos. São feitos 1000 testes por dia. A média de testes positivos quando se faz 20 dias de testes é de 750 ao invés de 1000 infectados.

4. Amostras para o Covid-19

Vimos nas seções anteriores que a escolha das pessoas que compõem uma amostra é de fundamental importância para se saber o que de fato acontece na população. Para a pandemia do Covid-19, tem que ser feito um protocolo que atende dois princípios: i) deve ser consistente com a teoria estatística, de modo que possamos saber o que se passa com a população geral; ii) deve ser priorizada a escolha de pessoas com a saúde debilitada, e que necessitam caracterizar a presença do vírus como questão de vida e morte. Para ambos princípios, a amostra com viés é a resposta. E qual protocolo de viés está sendo usado? Podemos afirmar que, em praticamente todos os lugares do mundo, o princípio de consistência com a teoria estatística está sendo desconsiderado. O princípio de escolha de pessoas com saúde debilitada é preponderante. A falta de testes é um fator crucial para essa escolha. Naturalmente, se não se pode fazer testes em massa, a prioridade não será a teoria estatística e sim a saúde. Contudo, deve ficar claro que esta limitação quebra completamente a possibilidade de inferir o real número de infectados, a partir de uma teoria científica sólida e estabelecida. O que se vê, portanto, são diversas linhas de análises, em que os autores assumem hipóteses em função dos resultados de casos positivos, mas são resultados que ficam obscuros uma vez que não há um protocolo estatístico bem definido para as amostras.

Diversos sites apresentam dados sobre a pandemia do Covid-19^[6,7]. Além de outras variáveis, mostram os casos positivos por país, estado e municípios. Não vemos em nenhum dos lugares um resultado formal do número de infectados, do ponto de vista científico, em que se apresente o protocolo, o grau de confiança e margem de erro. Pior, muitas vezes, o que observamos são dados coletados em diferentes instituições e diferentes protocolos e somados numa mesma tabela. Na maioria das vezes, isto vai gerar erros estatísticos grosseiros.

Em uma análise mais detalhada dos boletins diários de praticamente todos os lugares do mundo, vê-se uma variação imensa de número de testes por dia, mudanças de protocolo de testagem, mudanças de metodologia na forma de definir quando o paciente está com o vírus. E mesmo cientes de toda essa falta de regra, os dados são colocados juntos numa mesma tabela de evolução. A teoria estatística passa ao longe.

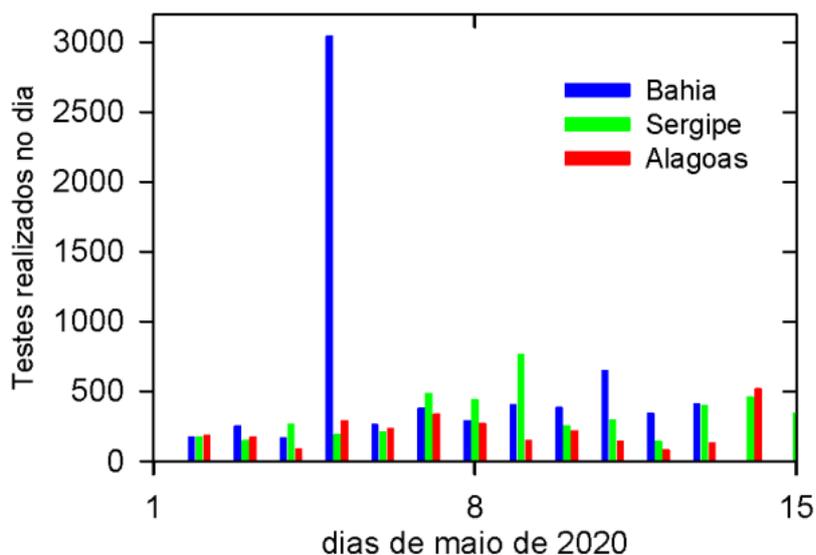


Figura 04: Número de testes diários realizados para detectar pessoas com Covid-19 na Bahia, Sergipe e Alagoas durante os primeiros 15 dias do mês de maio

A figura 04 mostra o número de testes para detectar a presença do vírus do Covid-19 na Bahia^[8], Sergipe^[9] e Alagoas^[10]. No nosso estado, somos mais de 2 milhões de pessoas e uma média de 304 teste são feitas por dia. Nos três estados vemos uma grande flutuação diária dos testes. Neste ponto da evolução da pandemia a quantidade de testes está muito abaixo do aceitável. Sabemos que é um problema que vem ocorrendo em todo Brasil, fruto da falta de uma política unificada para enfrentamento da pandemia, e da falta de autossuficiência de tecnologia nacional, em que nos deixa vulneravelmente dependente de importação de testes e de diversos equipamentos necessários.

Pesquisadores vêm enfrentando o problema da falta de precisão dos dados criando métodos de compensação para o problema da subnotificação de casos. Aqui em Sergipe, a profa. Márcia Montalvão e o prof. Jugurta Montalvão aplicam uma compensação baseada em número de mortes, considerando a letalidade e a latência entre contaminação e morte pela Covid-19^[11]. Para os óbitos não há necessidade de uso de amostras estatísticas, contudo, bem observado pelos autores, mesmo nos óbitos temos visto subnotificações, a exemplo da cidade de Manaus^[2].

4. Comentários finais

Apresentamos aqui a terceira nota técnica sobre a pandemia do Covid-19^[12,13]. Fizemos uma breve introdução da teoria estatística em que se baseia a escolha de uma amostra, isto é, a escolha de uma pequena parte da população, que deve representar as características da população total. Vimos que essa escolha tem duas grandes classes: sem e com viés. Sem viés, a amostra é escolhida aleatoriamente, podendo ser feitos extratos para melhor separar diferentes perfis da população. Para amostras sem viés, para casos em que queremos analisar situações em que a quantidade de indivíduos com a enfermidade é muito pequena, mostramos que a margem de erro é muito maior que o valor que obtemos. Esse é o caso do estágio atual da pandemia do Covid-19, por exemplo, em Sergipe. Para pesquisas sem viés, devemos esperar um pouco mais. Vimos que uma possível pesquisa sem viés em Sergipe com mil pessoas têm uma margem de erro de 3.1%, isto é, 71 mil pessoas. Portanto, não sugerimos que sejam realizadas em estágios da pandemia com menos de 200 mil pessoas infectadas. Pensando em outros lugares, isto significa que não sugerimos amostras sem viés feitas com mil testes enquanto o Covid-19 não atinja 10% da população.

Vimos ainda, que, independente da classe ou protocolo de amostra, o número de casos positivos nos testes para detectar o Covid-19, não representam em absoluto o número de pessoas infectadas. Sempre o número de infectados na população será maior ou igual aos casos positivos detectados pelos testes. Quanto mais infectados e menos testes realizados, maior a possibilidade da subnotificação.

Estudos com protocolos de seleção para testagem de indivíduos devem ter claras as premissas estatísticas na sua divulgação. A apresentação de resultados somando amostras com protocolos diferentes pode gerar erros estatísticos grosseiros. Considerando a imprecisão científica de apresentação de dados por parte dos sites e boletins, devemos ter muito cuidado ao informar a população em geral o significado de cada resultado.

Para fugir dessas imprecisões, pesquisadores estão preferindo analisar dados de internações e óbitos, que são dados diretos da população total. Infelizmente, mesmos estes, temos visto subnotificações, observadas quando se faz a comparação da média de mortes em um certo lugar nos meses de março e abril deste ano com os anos anteriores.

Finalizamos alertando que, seja pela falta de rigor científico na apresentação de dados ou falha

formal na sua aplicação, a teoria estatística está sendo inadequadamente usada em toda parte do mundo quando aplicada à pandemia do Covid-19. Isso tem trazido consequências sérias quando induz aplicações de políticas públicas errôneas, bem como, incertezas para a população em geral.

5. Referências

- [1] <https://g1.globo.com/al/alagoas/noticia/2020/05/13/obitos-por-covid-19-provocam-aumento-da-medida-diaria-de-sepultamentos-em-maceio.ghtml>
- [2] <https://g1.globo.com/jornal-nacional/noticia/2020/04/29/em-manaus-a-medida-de-mortes-por-dia-quadruplicou-desde-o-inicio-da-pandemia.ghtml>
- [3] LARSON, Ron & FARBER, Betsy. *Estatística Aplicada*, Pearson Prentice Hall, São Paulo, 4ª Ed., 2010.
- [4] MOOD, A. M., GRAYBILL, F. A., and BOES, D. C. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 3d ed., 1974.
- [5] DOROFEEV, Sergey & GRANT, Peter. *Statistics for Real-Life Sample Surveys: Non-Simple Random Samples and Weighted Data*. Cambridge University Press, New York; 1ª Edição, 2006.
- [6] <https://www.worldometers.info/coronavirus/#countries>
- [7] <https://coronavirus.jhu.edu/map.html>
- [8] <http://www.saude.ba.gov.br/temasdesaude/coronavirus/notas-tecnicas-e-boletins-epidemiologicos-covid-19/>
- [9] <https://todoscontraocorona.net.br/>
- [10] <http://www.alagoascontraocoronavirus.al.gov.br/#boletim>
- [11] MONTALVÃO, Márcia V. P. & MONTALVÃO, Jugurta. *Covid-19 em Sergipe - Projeções e método para compensação de subnotificações*, preprint 2020 (<https://www.researchgate.net/publication/341369731>)
- [12] SOUZA, André M. C. *Nota técnica sobre a pandemia de Covid-19 em Sergipe*. Nota técnica AMCS01, Edição 06/05/2020. (https://www.researchgate.net/publication/341432348_Nota_tecnica_sobre_a_pandemia_de_Covid-19_em_Sergipe.)
- [13] SOUZA, André M. C. *Sobre a pandemia de Covid-19 em Sergipe*. Nota técnica AMCS 02, Edição 13/05/2020. https://www.researchgate.net/publication/341489720_Sobre_a_pandemia_de_Covid-19_em_Sergipe_nota_tecnica_02 (DOI: [0.13140/RG.2.2.30522.62402](https://doi.org/10.13140/RG.2.2.30522.62402)).

6. Agradecimentos

Aos professores André Ramos, Francisco Almeida e Raimundo Oliveira Filho pelas sugestões.

ERRATA DA NOTA TÉCNICA 2^[13]: Nas páginas 2 e 8, onde se lê “22”, leia-se “27 novos leitos de UTI”.