

1 Painel de Subnotificação

Durante a pandemia do novo coronavírus, o Brasil enfrenta o problema da falta de testes. A incapacidade de testar um número alto de habitantes nos impede de inferir adequadamente sobre a real situação da epidemia no país. Desta forma, buscamos uma forma de incorporar essa falta de testes na análise dos dados. Por esse motivo, analisar a subnotificação existente nos dados é uma forma de trazer informação para o público quanto para indicar a confiabilidade nos dados divulgados.

O *COVID19 Analytics* agora oferece o painel de análise de subnotificação. Através da combinação de duas bases de dados, conseguimos estimar a quantidade de óbitos não divulgada causadas pelo novo coronavírus. Com base nessa estimativa, calculamos a taxa de subnotificação para cada estado e capital do Brasil.

Para o cálculo do indicador de subnotificação, extraímos dados de óbitos totais por município de 2 fontes diferentes. Os dados referentes aos anos de 2016 a 2018 têm como fonte a base de dados do DATASUS. Estes são disponíveis publicamente e são retirados do [FTP do DATASUS](#). Para os dados de 2019 a 2020, utilizamos a fonte de dados do [Portal da Transparência](#), mantido pela Associação Nacional dos Registradores de Pessoas Naturais (ARPEN Brasil). Estes dados foram obtidos através do seguinte repositório no GitHub: <https://github.com/capyvara/brazil-civil-registry-data>. Este repositório contém os dados que são coletados do Portal da Transparência por *web scrapping*. Os dados são atualizados diariamente e também estão disponíveis de forma pública.

Explicamos abaixo o tratamento dado a cada um destes dados assim como a construção de cada um dos gráficos disponíveis no nosso aplicativo.

1.1 Fontes

1.1.1 DATASUS

A base de dados do Sistema de Informação de Mortalidade (SIM) do DATASUS é a base mais completa existente no Brasil para análise de Mortalidade. Devido à demora com que os dados levam para ser coletados, temos apenas as informações para 2018. Os dados do DATASUS são retirados do FTP e estão contidos na pasta do Sistema de Informação de Mortalidade. Mais informações sobre a tabela de dados pode ser obtida na [documentação dos dados](#). Estes dados estão disponíveis de 1996 a 2018.

Foram considerados os casos de óbitos por causas não-externas (coluna *CIRCOBITO=NA*) e não fetais (coluna *TIPOBITO=1*). Para a classificação dos óbitos por município, foi considerada a cidade de ocorrência do óbito (coluna *CODMUNOCUR*) ao invés da cidade de naturalidade da vítima (coluna *CODMUNNATU*). Foi contado o número de óbitos por dia usando estas características. Para os dados estaduais, agregamos o número de óbitos segundo os dois primeiros números do código do IBGE. Como os dados são comparados entre diferentes anos, fizemos um

ajuste para igualar anos bissextos a anos regulares. Desta forma, os dados de 29 de Fevereiro (quando existentes) foram somados ao do dia anterior.

1.1.2 ARPEN

A Associação Nacional dos Registradores de Pessoas Naturais organiza os dados de cartórios brasileiros e publica no Portal da Transparência. A partir desses dados, conseguimos ter uma visão detalhada do óbito baseado no registro feito junto ao cartório. Com base em análise de texto, pode-se classificar o tipo de morte. A coleta destes dados é muito mais fácil do que a do DATASUS. Desta forma, temos atualização diária do número de óbitos com esta fonte e dados disponíveis já para o ano de 2020. Contudo, esta base não é tão ampla quanto a do DATASUS, abrangendo um número menor de cidades e apresentando normalmente um número menor de óbitos para o mesmo município que a base mais ampla. Além disso, existe uma diferença entre a data efetiva do óbito e o lançamento deste no portal. Segundo escrito no próprio site “A família tem até 24h após o falecimento para registrar o óbito em Cartório que, por sua vez, tem até cinco dias para efetuar o registro de óbito, e depois até oito dias para enviar o ato feito à Central Nacional de Informações do Registro Civil (CRC Nacional), que atualiza esta plataforma.”. Desta forma, a análise destes dados para períodos mais recentes pode ser complicada devido ao fato dos dados não estarem completos. Por este motivo, representamos a região referente aos últimos 14 dias com uma região sombreada, informando que - para aqueles dias – os dados devem ser olhados com cautela.

Os dados da ARPEN são retirados por *web scrapping* do Portal da Transparência e são guardados no GitHub: <https://github.com/capyvara/brazil-civil-registry-data>. Agradecemos a Marcelo Oliveira (infovid) pelo gerenciamento e atualização deste repositório. Nesta base de dados, os óbitos são classificados como causados por:

- COVID-19
- Síndrome respiratória aguda grave (SRAG)
- Pneumonia
- Insuficiência respiratória
- Septicemia (sepse/choque séptico)
- Indeterminadas (causas mortes ligadas a doenças respiratórias, mas não conclusivas)
- Demais óbitos (todos os outros tipos de óbitos que não estão listados a cima)

Ainda construímos 2 colunas adicionais:

- Óbitos totais (soma de todos os itens)
- Óbitos por causas respiratórias (COVID+SRAG+Pneumonia+Septicemia+Insuficiência Respiratória)

Para os dias em que não obtivemos nenhum registro de morte na base de dados, foi considerado o número de 0 mortes. Como os dados são comparados entre diferentes anos, fizemos um ajuste para igualar anos bissextos a anos regulares. Desta forma, os dados de 29 de Fevereiro (quando existentes) foram somados ao do dia anterior.

1.2 Indicadores

No aplicativo de análise de subnotificação do *COVID19 Analytics* você acompanha as séries de óbitos totais, excesso de mortes e indicador de subnotificação. Explicamos a fundo a construção de cada uma dessas séries. Estes dados estão disponíveis para os 27 estados e as 27 capitais do Brasil nas 3 frequências: diário, semanal e acumulado no ano. Para os dados em formato diário, foi construída uma média móvel de 7 dias para suavização.

1.2.1 Número de Óbitos

O gráfico de número de óbitos contém os dados brutos encontrados em cada uma das bases de dados.

1.2.2 Excesso de Mortes

O gráfico de excesso de mortes é usado como um comparativo para os dados de COVID-19. Este indicador é calculado como sendo a diferença no número de óbitos em 2020 da sua média histórica no mesmo período, isso é, tomando como exemplo os dados diários:

$$Excesso_{(01/abr)} = \acute{O}bitos_{(01/abr)}^{2020} - M\acute{e}dia_Hist\acute{o}rica_{(01/abr)}$$

Para o cálculo do excesso de mortes, usamos 3 tipos diferentes de Média Histórica. A primeira é calculada usando apenas os dados disponíveis na base do DATASUS, indo de 2016 a 2018:

$$M\acute{e}dia_Hist\acute{o}rica_{(01/abr)}^{2016-2018} = \frac{1}{3} \left(\acute{O}bitos_{(01/abr)}^{2016} + \acute{O}bitos_{(01/abr)}^{2017} + \acute{O}bitos_{(01/abr)}^{2018} \right)$$

O segundo cálculo da média histórica é feito utilizando apenas o ano de 2019, que corresponde à base de dados do ARPEN (a mesma que é usado para 2020).

$$M\acute{e}dia_Hist\acute{o}rica_{(01/abr)}^{2019} = \acute{O}bitos_{(01/abr)}^{2019}$$

A última base do cálculo da média histórica utiliza apenas as mortes definidas como causadas por problemas respiratórios. Esta também só está disponível para o ano de 2019 e é calculada de forma equivalente.

$$M\acute{e}dia_Hist\acute{o}rica_{(01/abr)}^{CR} = \acute{O}bitos_por_Causas_Respirat\acute{o}rias_{(01/abr)}^{2019}$$

Anterior ao período de COVID-19, deveríamos ter um alinhamento dos dados de 2020 com alguma média histórica. Desta forma, qualquer discrepância que exista entre os óbitos ocorridos no ano atual e a sua base histórica pode ser associada ao COVID-19, ou seja:

$$Excesso_{(01/abr)} \approx \acute{O}bitos_COVID_Divulgados_{(01/abr)} + \acute{O}bitos_COVID_N\tilde{a}oDivulgados_{(01/abr)}$$

1.2.3 Indicador de Subnotificação

O indicador de subnotificação é calculado usando as duas bases anteriores. Como é esperado que o número de mortes em 2020 siga aproximadamente a sua média histórica, temos que o excesso de mortes pode ser interpretado pelas mortes totais causadas pelo novo coronavírus. Desta forma, comparando as mortes totais (consideradas pelo excesso de mortes) com o total de mortes registrados como efetivamente causadas pelo COVID-19, temos a taxa de subnotificação de casos. Quanto maior o indicador de subnotificação, maior a discrepância entre os dados divulgados de COVID-19 e o excesso de mortes. A fórmula do indicador de subnotificação é dada por:

$$Subnotificac\tilde{a}o_{(01/abr)} = 1 - \frac{\acute{O}bitos_COVID_Divulgados_{(01/abr)}}{Excesso_{(01/abr)}}$$

Novamente temos 3 tipos de média histórica assim como explicado na seção de Excesso de Mortes. O Indicador é calculado para cada uma dessas médias e já está disponível no nosso aplicativo.