

Seleção dos casos de COVID-19 para paper de comparação com H1N1

Gestantes

18/06/2021

Sobre a base de dados e pacotes do R utilizados

A seguir são carregados os pacotes do R (<https://www.r-project.org>) utilizados para filtragem e tratamento dos dados.

```
#carregar pacotes
loadlibrary <- function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = T)
    if (!require(x, character.only = TRUE))
      stop("Package not found")
  }
}

packages <-
  c(
    "readxl",
    "janitor",
    "dplyr",
    "skimr",
    "forcats",
    "stringr",
    "lubridate",
    "readr",
    "summarytools",
    "knitr",
    "pander",
    "modelsummary",
    "magrittr",
    "ggplot2",
    "rgdal",
    "tm",
    "wordcloud",
    "kableExtra",
    "tables",
    "questionr",
    "corrplot",
    "wordcloud2",
    "dlookr",
  )
```

```

    "patchwork",
    "sjPlot",
    "readr",
    "broom",
    "gmodels",
    "epitools",
    "survival",
    "survminer",
    "nortest",
    "car",
    "tidylog",
    "gtsummary",
    "writexl"
  )
lapply(packages, loadlibrary)

## Packages ##
library(ggplot2)      # fazer gráficos
library(rgdal)         # ler arquivos de mapas
library(dplyr)         # fazer manipulação de banco de dados
library(leaflet)      # gráficos com mapas interativos.
library(readxl)
library(highcharter)
library(geobr)

```

A base de dados SIVEP-Gripe (Sistema de Informação da Vigilância Epidemiológica da Gripe) tem os registros dos casos e óbitos de SRAG (Síndrome Respiratória Aguda Grave). A notificação é compulsória para síndrome gripal (caracterizado por pelo menos dois dos seguintes sinais e sintomas: febre, mesmo que referida, calafrios, dor de garganta, dor de cabeça, tosse, coriza, distúrbios olfatórios ou de paladar) e que tem dispneia / desconforto respiratório ou pressão persistente no peito ou Saturação de O₂ menor que 95% no ar ambiente ou cor azulada dos lábios ou rosto. Indivíduos assintomáticos com confirmação laboratorial por biologia molecular ou exame imunológico para infecção por COVID-19 também são relatados.

Para notificações no Sivep-Gripe, os casos hospitalizados em hospitais públicos e privados e todas as mortes devido a infecções respiratórias agudas graves, independentemente da hospitalização, devem ser considerados.

A vigilância da SRAG no Brasil é desenvolvida pelo Ministério da Saúde (MS), por meio da Secretaria de Vigilância em Saúde (SVS), desde a pandemia de Influenza A (H1N1) em 2009. Mais informações em <https://coronavirus.saude.gov.br/definicao-de-caso-e-notificacao>.

O período analisado compreende de dados epidemiológicos de 2020, com banco de dados obtido em 19/05/2021 no site <https://opendatasus.saude.gov.br/dataset/bd-srag-2020>, e de 2021, com banco de dados obtido em 19/05/2021 no site <https://opendatasus.saude.gov.br/dataset/bd-srag-2021>. Os dados de 2020 e de 2021 são carregados e combinados abaixo:

```

##### carregando as bases de dados #####
#2021
dados_2021 <- read_delim(
  "INFLUD21-17-05-2021.csv",
  ";",
  escape_double = FALSE,
  locale = locale(encoding = "ISO-8859-2"),
  trim_ws = TRUE
)

```

```

#2020
dados_2020 <- read_delim(
  "INFLUD-17-05-2021.csv",
  ";",
  escape_double = FALSE,
  locale = locale(encoding = "ISO-8859-2"),
  trim_ws = TRUE
)

sem <- 19

memory.limit(999999)

#### Concatenar dados 2020 e 2021 #####
dados1 <- rbind(dados_2020, dados_2021)

#Criar variavel de ano do caso
dados1 <- dados1 %>%
  dplyr::mutate(
    dt_sint = as.Date(DT_SIN_PRI, format = "%d/%m/%Y"),
    ano = lubridate::year(dt_sint),
    mes = lubridate::month(dt_sint)
  )

```

Há atualmente 2007693 observações na base de dados e são as variáveis:

```
names(dados1)
```

```

## [1] "DT_NOTIFIC" "SEM_NOT" "DT_SIN_PRI" "SEM_PRI" "SG_UF_NOT"
## [6] "ID_REGIONA" "CO_REGIONA" "ID_MUNICIP" "CO_MUN_NOT" "ID_UNIDADE"
## [11] "CO_UNI_NOT" "CS_SEXO" "DT_NASC" "NU_IDADE_N" "TP_IDADE"
## [16] "COD_IDADE" "CS_GESTANT" "CS_RACA" "CS_ETINIA" "CS_ESCOL_N"
## [21] "ID_PAIS" "CO_PAIS" "SG_UF" "ID_RG_RESI" "CO_RG_RESI"
## [26] "ID_MN_RESI" "CO_MUN_RES" "CS_ZONA" "SURTO_SG" "NOSOCOMIAL"
## [31] "AVE_SUINO" "FEBRE" "TOSSE" "GARGANTA" "DISPNEIA"
## [36] "DESC_RESP" "SATURACAO" "DIARREIA" "VOMITO" "OUTRO_SIN"
## [41] "OUTRO_DES" "PUERPERA" "FATOR_RISC" "CARDIOPATI" "HEMATOLOGI"
## [46] "SIND_DOWN" "HEPATICA" "ASMA" "DIABETES" "NEUROLOGIC"
## [51] "PNEUMOPATI" "IMUNODEPRE" "RENAL" "OBESIDADE" "OBES_IMC"
## [56] "OUT_MORBI" "MORB_DESC" "VACINA" "DT_UT_DOSE" "MAE_VAC"
## [61] "DT_VAC_MAE" "M_AMAMENTA" "DT_DOSEUNI" "DT_1_DOSE" "DT_2_DOSE"
## [66] "ANTIVIRAL" "TP_ANTIVIR" "OUT_ANTIV" "DT_ANTIVIR" "HOSPITAL"
## [71] "DT_INTERNA" "SG_UF_INTE" "ID_RG_INTE" "CO_RG_INTE" "ID_MN_INTE"
## [76] "CO_MU_INTE" "UTI" "DT_ENTUTI" "DT_SAIDUTI" "SUPOORT_VEN"
## [81] "RAIOX_RES" "RAIOX_OUT" "DT_RAIOX" "AMOSTRA" "DT_COLETA"
## [86] "TP_AMOSTRA" "OUT_AMOST" "PCR_RESUL" "DT_PCR" "POS_PCRFLU"
## [91] "TP_FLU_PCR" "PCR_FLUASU" "FLUASU_OUT" "PCR_FLUBLI" "FLUBLI_OUT"
## [96] "POS_PCROUT" "PCR_VSR" "PCR_PARA1" "PCR_PARA2" "PCR_PARA3"
## [101] "PCR_PARA4" "PCR_ADENO" "PCR_METAP" "PCR_BOCA" "PCR_RINO"
## [106] "PCR_OUTRO" "DS_PCR_OUT" "CLASSI_FIN" "CLASSI_OUT" "CRITERIO"
## [111] "EVOLUCAO" "DT_EVOLUCA" "DT_ENCERRA" "DT_DIGITA" "HISTO_VGM"
## [116] "PAIS_VGM" "CO_PS_VGM" "LO_PS_VGM" "DT_VGM" "DT_RT_VGM"
## [121] "PCR_SARS2" "PAC_COCBO" "PAC_DSCBO" "OUT_ANIM" "DOR_ABD"

```

```
## [126] "FADIGA"      "PERD_OLFT"  "PERD_PALA"  "TOMO_RES"   "TOMO_OUT"
## [131] "DT_TOMO"     "TP_TES_AN"  "DT_RES_AN"  "RES_AN"     "POS_AN_FLU"
## [136] "TP_FLU_AN"   "POS_AN_OUT" "AN_SARS2"    "AN_VSR"     "AN_PARA1"
## [141] "AN_PARA2"    "AN_PARA3"   "AN_ADENO"    "AN_OUTRO"    "DS_AN_OUT"
## [146] "TP_AM_SOR"   "SOR_OUT"    "DT_CO_SOR"   "TP_SOR"     "OUT_SOR"
## [151] "DT_RES"      "RES_IGG"    "RES_IGM"     "RES_IGA"     "dt_sint"
## [156] "ano"         "mes"
```

Para ver o dicionário das variáveis, acesse: <https://opendatasus.saude.gov.br/dataset/ae90fa8f-3e94-467e-a33f-94adbb66edf8/resource/8f571374-c555-4ec0-8e44-00b1e8b11c25/download/dicionario-de-dados-srag-hospitalizado-27.07.2020-final.pdf>

Filtragem e tratamento dos dados para projeto

A primeira filtragem consiste em selecionar as semanas epidemiológicas de sintomas da análise em 2020 e em 2021.

```
#seleção das semanas epidemiológicas de 2020 e de 2021
dados2 <- dados1 %>%
  filter((ano==2020 & SEM_PRI >=8) | ano ==2021)

dados2 <- dados2 %>%
  mutate(ano = ifelse(ano ==2021 & SEM_PRI ==53, 2020, ano)) %>% filter(ano==2020 | (ano ==2021 & SEM_PRI >=8))

with(dados2, ctable(SEM_PRI, ano))
```

```
## Cross-Tabulation, Row Proportions
## SEM_PRI * ano
## Data Frame: dados2
##
## -----
##      ano                2020                2021                Total
## SEM_PRI
##      1      0 ( 0.0%)    35407 (100.0%)    35407 (100.0%)
##      2      0 ( 0.0%)    34176 (100.0%)    34176 (100.0%)
##      3      0 ( 0.0%)    31415 (100.0%)    31415 (100.0%)
##      4      0 ( 0.0%)    29396 (100.0%)    29396 (100.0%)
##      5      0 ( 0.0%)    35312 (100.0%)    35312 (100.0%)
##      6      0 ( 0.0%)    38422 (100.0%)    38422 (100.0%)
##      7      0 ( 0.0%)    48352 (100.0%)    48352 (100.0%)
##      8      925 ( 1.8%)    51278 ( 98.2%)    52203 (100.0%)
##      9     1165 ( 1.6%)    71029 ( 98.4%)    72194 (100.0%)
##     10     1984 ( 2.7%)    71533 ( 97.3%)    73517 (100.0%)
##     11     5139 ( 6.8%)    70292 ( 93.2%)    75431 (100.0%)
##     12    12825 (18.7%)    55824 ( 81.3%)    68649 (100.0%)
##     13    14978 (22.3%)    52277 ( 77.7%)    67255 (100.0%)
##     14    16298 (25.1%)    48646 ( 74.9%)    64944 (100.0%)
##     15    19581 (32.3%)    41096 ( 67.7%)    60677 (100.0%)
##     16    24877 (39.2%)    38575 ( 60.8%)    63452 (100.0%)
##     17    30814 (46.6%)    35298 ( 53.4%)    66112 (100.0%)
##     18    34864 (62.3%)    21087 ( 37.7%)    55951 (100.0%)
```

##	19	34576 (87.3%)	5016 (12.7%)	39592 (100.0%)
##	20	37150 (100.0%)	0 (0.0%)	37150 (100.0%)
##	21	33800 (100.0%)	0 (0.0%)	33800 (100.0%)
##	22	31235 (100.0%)	0 (0.0%)	31235 (100.0%)
##	23	35607 (100.0%)	0 (0.0%)	35607 (100.0%)
##	24	34139 (100.0%)	0 (0.0%)	34139 (100.0%)
##	25	36700 (100.0%)	0 (0.0%)	36700 (100.0%)
##	26	32968 (100.0%)	0 (0.0%)	32968 (100.0%)
##	27	37431 (100.0%)	0 (0.0%)	37431 (100.0%)
##	28	37026 (100.0%)	0 (0.0%)	37026 (100.0%)
##	29	34453 (100.0%)	0 (0.0%)	34453 (100.0%)
##	30	33718 (100.0%)	0 (0.0%)	33718 (100.0%)
##	31	32203 (100.0%)	0 (0.0%)	32203 (100.0%)
##	32	30043 (100.0%)	0 (0.0%)	30043 (100.0%)
##	33	31091 (100.0%)	0 (0.0%)	31091 (100.0%)
##	34	28286 (100.0%)	0 (0.0%)	28286 (100.0%)
##	35	26357 (100.0%)	0 (0.0%)	26357 (100.0%)
##	36	26499 (100.0%)	0 (0.0%)	26499 (100.0%)
##	37	24059 (100.0%)	0 (0.0%)	24059 (100.0%)
##	38	22244 (100.0%)	0 (0.0%)	22244 (100.0%)
##	39	21578 (100.0%)	0 (0.0%)	21578 (100.0%)
##	40	22454 (100.0%)	0 (0.0%)	22454 (100.0%)
##	41	21032 (100.0%)	0 (0.0%)	21032 (100.0%)
##	42	19028 (100.0%)	0 (0.0%)	19028 (100.0%)
##	43	19470 (100.0%)	0 (0.0%)	19470 (100.0%)
##	44	18727 (100.0%)	0 (0.0%)	18727 (100.0%)
##	45	23281 (100.0%)	0 (0.0%)	23281 (100.0%)
##	46	25803 (100.0%)	0 (0.0%)	25803 (100.0%)
##	47	29261 (100.0%)	0 (0.0%)	29261 (100.0%)
##	48	29161 (100.0%)	0 (0.0%)	29161 (100.0%)
##	49	33057 (100.0%)	0 (0.0%)	33057 (100.0%)
##	50	30608 (100.0%)	0 (0.0%)	30608 (100.0%)
##	51	28490 (100.0%)	0 (0.0%)	28490 (100.0%)
##	52	30470 (100.0%)	0 (0.0%)	30470 (100.0%)
##	53	34280 (100.0%)	0 (0.0%)	34280 (100.0%)
##	Total	1189735 (59.4%)	814431 (40.6%)	2004166 (100.0%)
##	-----	-----	-----	-----

Há 2004166 observações na base de dados.

Vamos seleccionar os casos de março de 2020 até março de 2021.

```
#filtrando só os casos de março de 2020 até março de 2021
dados3 <- filter(dados2,
                 (mes >=3 & ano == 2020) | (mes <=3 & ano == 2021))
```

```
with(dados3, ctable(SEM_PRI, ano))
```

```
## Cross-Tabulation, Row Proportions
```

```
## SEM_PRI * ano
```

```
## Data Frame: dados3
```

```
##
```

```
## -----
## ano 2020 2021 Total
```

##	SEM_PRI			
##	1	0 (0.0%)	35407 (100.0%)	35407 (100.0%)
##	2	0 (0.0%)	34176 (100.0%)	34176 (100.0%)
##	3	0 (0.0%)	31415 (100.0%)	31415 (100.0%)
##	4	0 (0.0%)	29396 (100.0%)	29396 (100.0%)
##	5	0 (0.0%)	35312 (100.0%)	35312 (100.0%)
##	6	0 (0.0%)	38422 (100.0%)	38422 (100.0%)
##	7	0 (0.0%)	48352 (100.0%)	48352 (100.0%)
##	8	0 (0.0%)	51278 (100.0%)	51278 (100.0%)
##	9	0 (0.0%)	71029 (100.0%)	71029 (100.0%)
##	10	1984 (2.7%)	71533 (97.3%)	73517 (100.0%)
##	11	5139 (6.8%)	70292 (93.2%)	75431 (100.0%)
##	12	12825 (18.7%)	55824 (81.3%)	68649 (100.0%)
##	13	14978 (35.2%)	27559 (64.8%)	42537 (100.0%)
##	14	16298 (100.0%)	0 (0.0%)	16298 (100.0%)
##	15	19581 (100.0%)	0 (0.0%)	19581 (100.0%)
##	16	24877 (100.0%)	0 (0.0%)	24877 (100.0%)
##	17	30814 (100.0%)	0 (0.0%)	30814 (100.0%)
##	18	34864 (100.0%)	0 (0.0%)	34864 (100.0%)
##	19	34576 (100.0%)	0 (0.0%)	34576 (100.0%)
##	20	37150 (100.0%)	0 (0.0%)	37150 (100.0%)
##	21	33800 (100.0%)	0 (0.0%)	33800 (100.0%)
##	22	31235 (100.0%)	0 (0.0%)	31235 (100.0%)
##	23	35607 (100.0%)	0 (0.0%)	35607 (100.0%)
##	24	34139 (100.0%)	0 (0.0%)	34139 (100.0%)
##	25	36700 (100.0%)	0 (0.0%)	36700 (100.0%)
##	26	32968 (100.0%)	0 (0.0%)	32968 (100.0%)
##	27	37431 (100.0%)	0 (0.0%)	37431 (100.0%)
##	28	37026 (100.0%)	0 (0.0%)	37026 (100.0%)
##	29	34453 (100.0%)	0 (0.0%)	34453 (100.0%)
##	30	33718 (100.0%)	0 (0.0%)	33718 (100.0%)
##	31	32203 (100.0%)	0 (0.0%)	32203 (100.0%)
##	32	30043 (100.0%)	0 (0.0%)	30043 (100.0%)
##	33	31091 (100.0%)	0 (0.0%)	31091 (100.0%)
##	34	28286 (100.0%)	0 (0.0%)	28286 (100.0%)
##	35	26357 (100.0%)	0 (0.0%)	26357 (100.0%)
##	36	26499 (100.0%)	0 (0.0%)	26499 (100.0%)
##	37	24059 (100.0%)	0 (0.0%)	24059 (100.0%)
##	38	22244 (100.0%)	0 (0.0%)	22244 (100.0%)
##	39	21578 (100.0%)	0 (0.0%)	21578 (100.0%)
##	40	22454 (100.0%)	0 (0.0%)	22454 (100.0%)
##	41	21032 (100.0%)	0 (0.0%)	21032 (100.0%)
##	42	19028 (100.0%)	0 (0.0%)	19028 (100.0%)
##	43	19470 (100.0%)	0 (0.0%)	19470 (100.0%)
##	44	18727 (100.0%)	0 (0.0%)	18727 (100.0%)
##	45	23281 (100.0%)	0 (0.0%)	23281 (100.0%)
##	46	25803 (100.0%)	0 (0.0%)	25803 (100.0%)
##	47	29261 (100.0%)	0 (0.0%)	29261 (100.0%)
##	48	29161 (100.0%)	0 (0.0%)	29161 (100.0%)
##	49	33057 (100.0%)	0 (0.0%)	33057 (100.0%)
##	50	30608 (100.0%)	0 (0.0%)	30608 (100.0%)
##	51	28490 (100.0%)	0 (0.0%)	28490 (100.0%)
##	52	30470 (100.0%)	0 (0.0%)	30470 (100.0%)
##	53	21659 (100.0%)	0 (0.0%)	21659 (100.0%)

```
##      Total      1175024 ( 66.2%)   599995 ( 33.8%)   1775019 (100.0%)
## -----
```

Há 1775019 observações na base de dados.

A próxima seleção de pessoas do sexo feminino:

```
#filtrando F
dados4 <- filter(dados3, CS_SEXO == "F")
```

Há 813023 observações na base de dados.

O próximo passo é filtrar só as mulheres entre 10 e 49 anos.

```
# filtro de idade
dados5 <- filter(dados4, NU_IDADE_N > 9 & NU_IDADE_N < 50)
```

Há 216608 observações na base de dados.

A próxima seleção são os casos de covid indicado pela variável CLASSI_FIN.

```
with(dados5, freq(CLASSI_FIN))
```

```
##      n      % val%
## 1     372    0.2  0.2
## 2     714    0.3  0.4
## 3     507    0.2  0.3
## 4    65416   30.2 33.1
## 5   130340   60.2 66.0
## NA    19259    8.9  NA
```

```
dados6 <- dados5 %>%
  filter(CLASSI_FIN == 5)
```

Há 130340 observações na base de dados.

Agora vamos criar a variável se CLASSI_FIN==5 por PCR ou outro tipo de diagnóstico.

Essa variável é `pcr_test`, com as categorias: `pcr_sars2` se `PCR_SARS2 == 1`; `ds_pcr_out` se `PCR_SARS2` for NA e tem as palavras “SARS|COVID|COV|CORONA|CIVID” no `DS_PCR_OUT`; `pcr_pos_nada_outro` se PCR positivo (`CRITERIO == 1` e `PCR_RESUL == 1`) e não tem resultado positivo para nenhum outro agente etiológico.

```
#criar a variavel de pcr_test pelas variaveis de pcr
dados6 <- dados6 %>%
  mutate(pcr_test = case_when(PCR_SARS2 == 1 ~ "pcr_sars2",
                              (is.na(PCR_SARS2) &
                               stringr::str_detect(DS_PCR_OUT, "SARS|COVID|COV|CORONA|CIVID") &
                               !stringr::str_detect(DS_PCR_OUT, "63|43|229|HK|RINO|SINCI|PARE")) ~ "ds_pcr_out",
                              (
                                PCR_RESUL == 1 & #pcr positivo e caso finalizado,
                                #mas nenhum outro agente etiologico
                                CRITERIO == 1 &
```

```

is.na(PCR_SARS2) &
is.na(DS_PCR_OUT) &
(PCR_RINO != 1 |
  is.na(PCR_RINO)) &
(POS_PCRFLU != 1 | is.na(POS_PCRFLU)) &
(PCR_OUTRO != 1 | is.na(PCR_OUTRO)) &
(POS_PCROUT != 1 | is.na(POS_PCROUT)) &
(is.na(PCR_VSR)) &
(is.na(PCR_METAP)) &
(is.na(PCR_PARA1))
) ~ "pcr_pos_nada_outro",
TRUE ~ "não"
)
)

```

```
with(dados6, freq(pcr_test))
```

```

##           n      % val%
## ds_pcr_out      109  0.1  0.1
## não           37376 28.7 28.7
## pcr_pos_nada_outro 6863  5.3  5.3
## pcr_sars2       85992 66.0 66.0

```

Agora iremos filtrar os casos com `pcr_test=="pcr_sars2"` que são os casos onde podem ser covid-19 apenas por PCR :

```

#dados7<- read.csv("dados_COVID_21-05-2021.csv")
dados7 <- dados6 %>%
  filter(pcr_test == "pcr_sars2")

```

Há 85992 observações na base de dados.

Agora vamos selecionar só as pessoas gestantes ou não gestante.

```
with(dados7, freq(CS_GESTANT))
```

```

##      n      % val%
## 0      1  0.0  0.0
## 1    484  0.6  0.6
## 2   1376  1.6  1.6
## 3   3271  3.8  3.8
## 4    252  0.3  0.3
## 5  64224 74.7 74.7
## 6   3430  4.0  4.0
## 9  12954 15.1 15.1

```

```

dados7 <- dados7 %>%
  mutate(
    classi_gesta = case_when(
      CS_GESTANT == 1 ~ "1tri",
      CS_GESTANT == 2 ~ "2tri",

```



```

    CS_GESTANT == 3 ~ "3tri",
    CS_GESTANT == 4 ~ "IG_ig",
    CS_GESTANT == 5 ~ "não",
    TRUE ~ NA_character_
  )
)

#filtrando só gestante ou não gestante
dados8 <- dados7 %>%
  filter(!is.na(classi_gesta))

# Criando a variável gestante_SN
dados8 <- dados8 %>%
  mutate(gestante_SN = ifelse(CS_GESTANT == 5, "não", "sim"))

freq(dados8$gestante_SN)

##           n      % val%
## não 64224 92.3 92.3
## sim  5383  7.7  7.7

```

Há 69607 observações na base de dados.