

Análise exploratória sobre variantes

Elias, Lucas e Ornella

2022-07-02

Tratamento dos dados

De início iremos importar a base de dados e criar a variável de **variante**. A criação dessa variável se dá sabendo que as variantes gama, delta e omicron tiveram seus inícios em 01/02/2021, 01-08-2021 e 01-01-2022, respectivamente.

```
# Carregando a base de dados ----
dados <- readRDS("dados6.rds")

# Somente COVID ----
dados <- dados %>%
  filter(CLASSI_FIN == "5")

#criando variável chamada variante(original,gama,delta,omicron)
in_gama <- as.Date("01-02-2021", format = "%d-%m-%Y")
in_delta <- as.Date("01-08-2021", format = "%d-%m-%Y")
in_omicron <- as.Date("01-01-2022", format = "%d-%m-%Y")

dados <- dados %>%
  mutate(
    variante = case_when(
      dt_sint < in_gama ~ "original",
      dt_sint >= in_gama &
        dt_sint < in_delta ~ "gama",
      dt_sint >= in_delta &
        dt_sint < in_omicron ~ "delta",
      dt_sint >= in_omicron ~ "omicron"
    )
  ) %>%
  mutate(month_year = paste(
    formatC(
      month(dt_sint),
      width = 2,
      format = "d",
      flag = "0"
    ),
    year(dt_sint),
    sep = "/"
  )) %>%
  mutate(mes = month(dt_sint))
```

```

dados <- dados %>%
  mutate(vacina_cov = ifelse(variante == "original", "não", vacina_cov))

dados$variante <- factor(dados$variante,
  levels = c("original", "gama", "delta", "omicron"))

dados <- dados %>%
  mutate(dt_1dose = as.Date(DOSE_1_COV, format = "%d/%m/%Y")) %>%
  mutate(dt_2dose = as.Date(DOSE_2_COV, format = "%d/%m/%Y")) %>%
  mutate(
    doses = case_when(
      vacina_cov == "sim" & is.na(dt_1dose)
      &
      is.na(dt_2dose) ~ "pelo menos uma dose",
      !is.na(dt_2dose) ~ "duas doses",
      !is.na(dt_1dose) &
      is.na(dt_2dose) ~ "pelo menos uma dose",
      TRUE ~ "não informado"
    )
  )

# Alterações iniciais na base ----

## Reajuste de levels ----
dados$faixa_et <- factor(dados$faixa_et,
  levels = c("<20", "20-34", ">=35"))

dados$escol <- factor(dados$escol,
  levels = c("sem escol", "fund1", "fund2", "medio", "superior"))

dados$suport_ven <- factor(dados$suport_ven,
  levels = c("não", "não invasivo", "invasivo"))

## Restante de alterações ----
dados$raca_sel <- dados$raca
dados$raca_sel <-
  ifelse(is.na(dados$raca_sel), "não informado", dados$raca_sel)

dados$vacina_cov_sel <-
  ifelse(is.na(dados$vacina_cov), "nao informado", dados$vacina_cov)

dados$CLASSI_FIN <- as.factor(dados$CLASSI_FIN)

dados$DT_SIN_PRI <- dmy(dados$DT_SIN_PRI)
dados$DT_EVOLUCA <- dmy(dados$DT_EVOLUCA)

dados <- dados %>%
  mutate(variante2 = as.numeric(
    case_when(
      variante == "original" ~ "1",
      variante == "gama" ~ "2",
      variante == "delta" ~ "3",
      variante == "omicron" ~ "4",

```

```

      TRUE ~ NA_character_
    )
  ))

dados <- dados %>%
  mutate(vacina_cov2 = as.numeric(
    case_when(
      vacina_cov == "sim" ~ "1",
      vacina_cov == "não" ~ "0",
      TRUE ~ NA_character_
    )
  ))

dados$vacinacov_variante <-
  as.factor(dados$vacina_cov2 * dados$variante2)

```

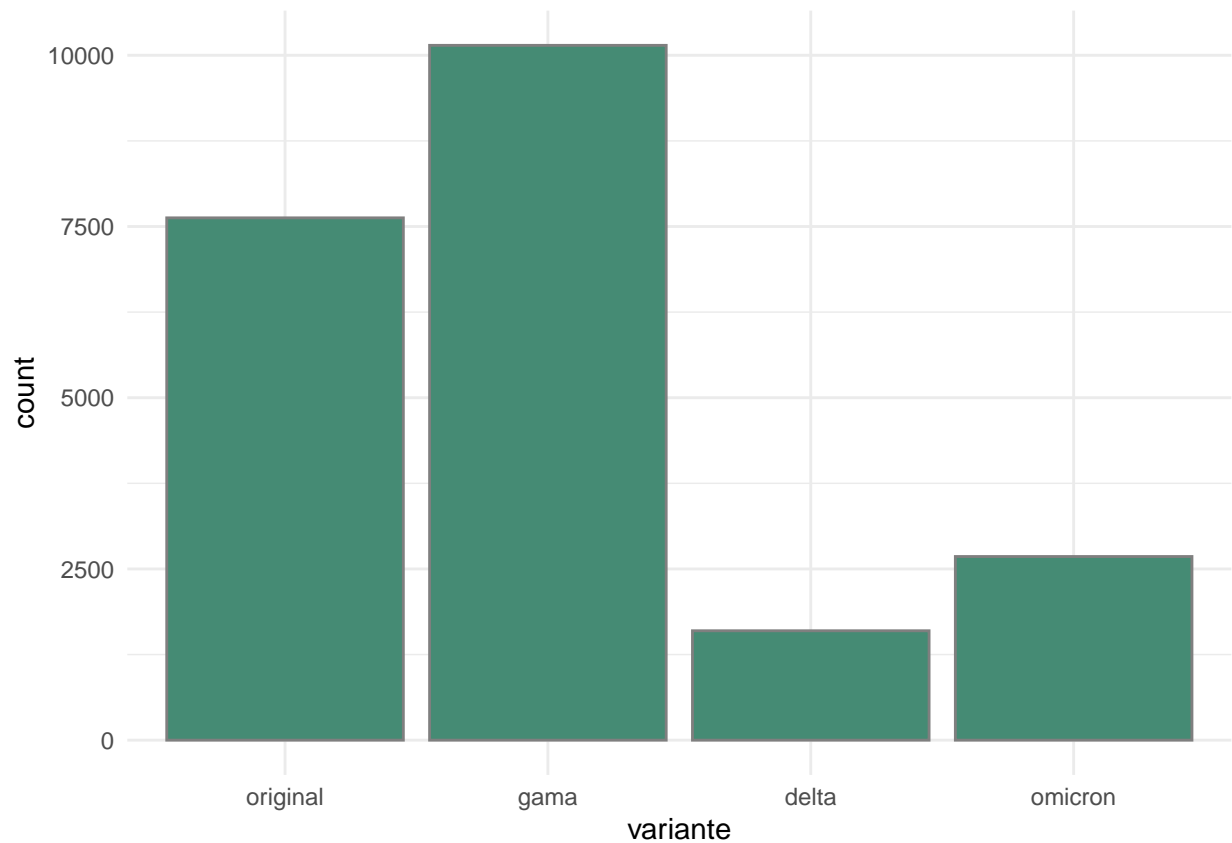
Gráficos de frequência

Casos para cada uma das variantes

```

dados %>%
  ggplot(aes(variante)) +
  geom_bar(fill = "aquamarine4", color = "gray50") +
  theme(legend.position = "none")

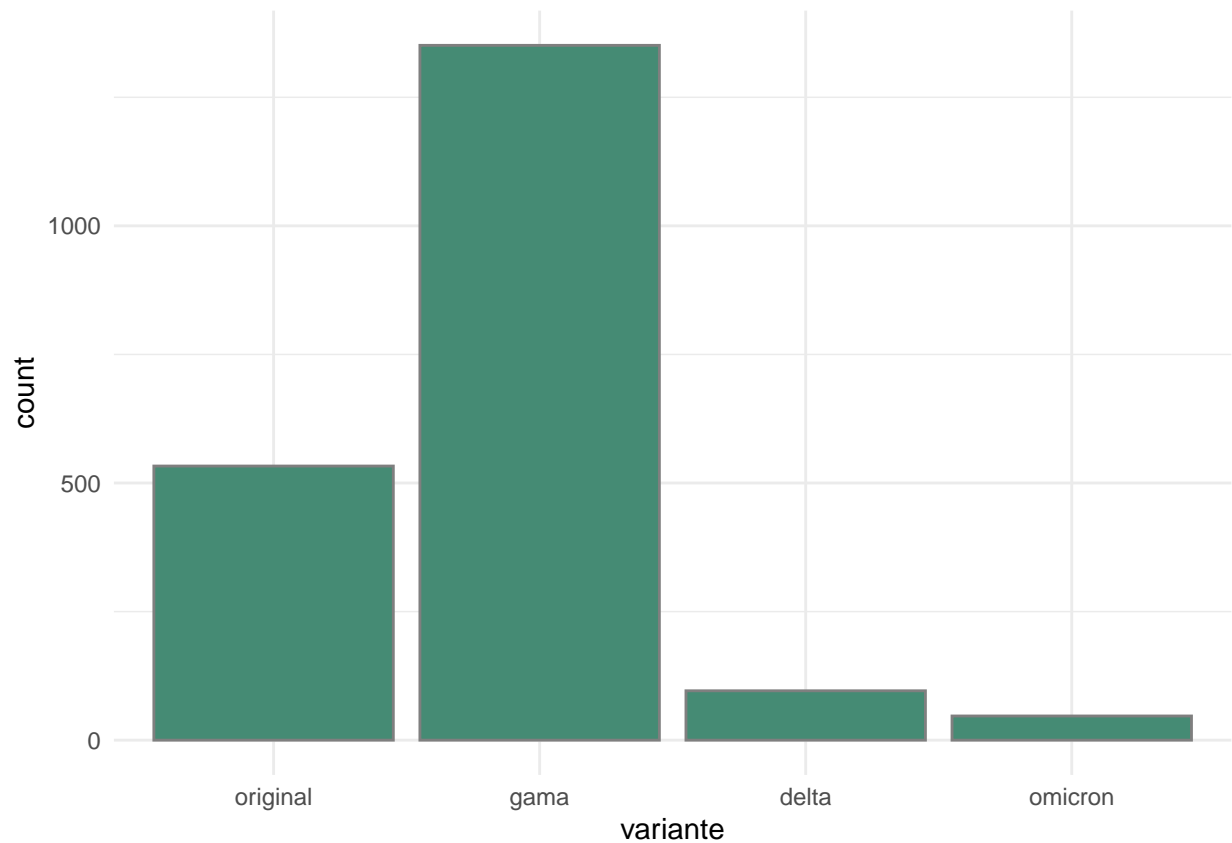
```



Podemos ver que, para a população de gestantes se puérperas, a variante gama foi a que teve maior número de casos. No tempo podemos tirar algumas informações mais concretas sobre essa questão. Para isso iremos gerar o gráfico a seguir.

Óbitos para cada uma das variantes

```
dados %>%
  filter(evolucao == "obito") %>%
  ggplot(aes(variante)) +
  geom_bar(fill = "aquamarine4", color = "gray50") +
  theme(legend.position = "none")
```



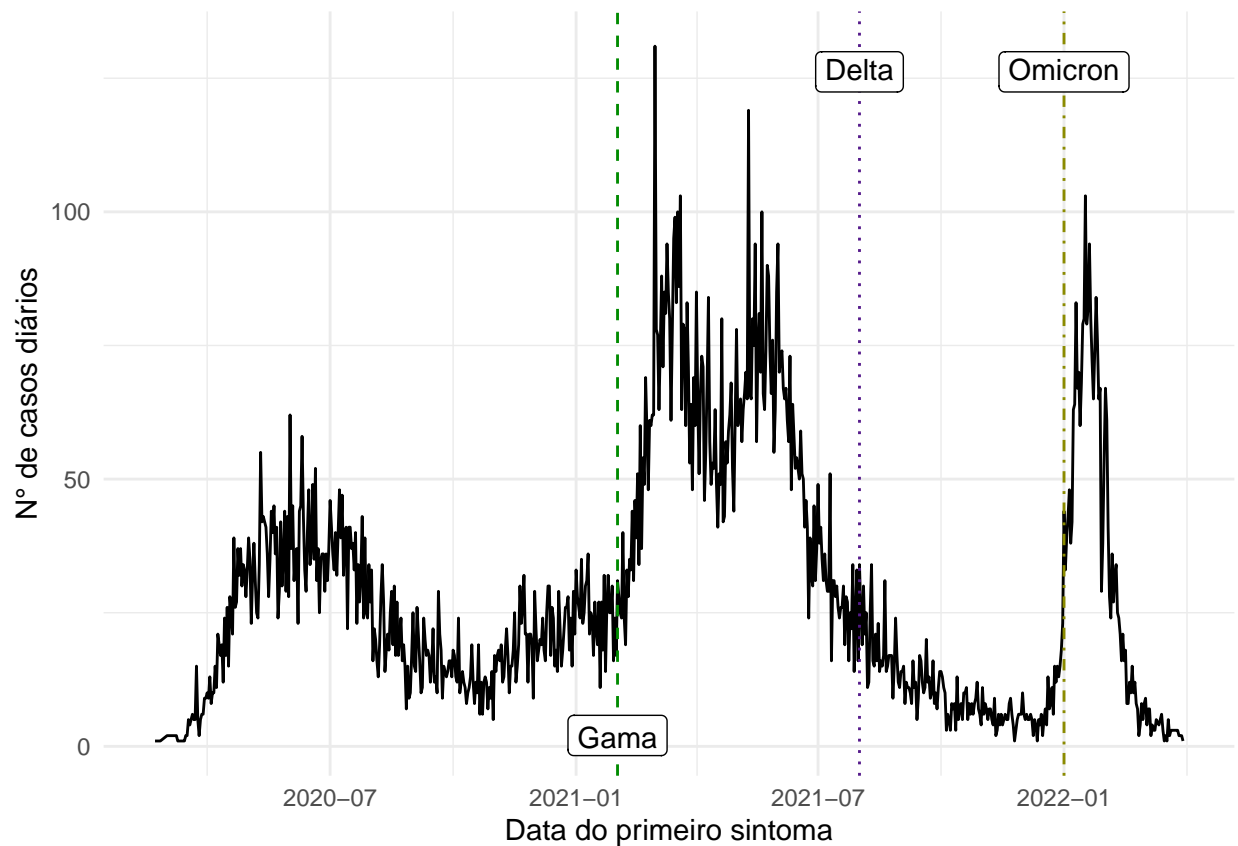
Evolução no tempo - casos

```
dados %>%
  group_by(DT_SIN_PRI) %>%
  summarise(casos = n()) %>%
  ggplot(aes(DT_SIN_PRI, casos)) +
  geom_line() +
  geom_vline(xintercept = in_gama,
             colour = "green4",
             linetype = 2) +
  annotate(
    x = in_gama,
    y = -Inf,
    label = "Gama",
    vjust = -0.5,
    geom = "label"
  ) +
  geom_vline(xintercept = in_delta,
             colour = "purple4",
             linetype = 3) +
  annotate(
    x = in_delta,
    y = +Inf,
```

```

    label = "Delta",
    vjust = 2,
    geom = "label"
  ) +
  geom_vline(xintercept = in_omicron,
             colour = "yellow4",
             linetype = 4) +
  annotate(
    x = in_omicron,
    y = +Inf,
    label = "Omicron",
    vjust = 2,
    geom = "label"
  ) +
  labs(
    x = "Data do primeiro sintoma",
    y = "N° de casos diários"
  )
)

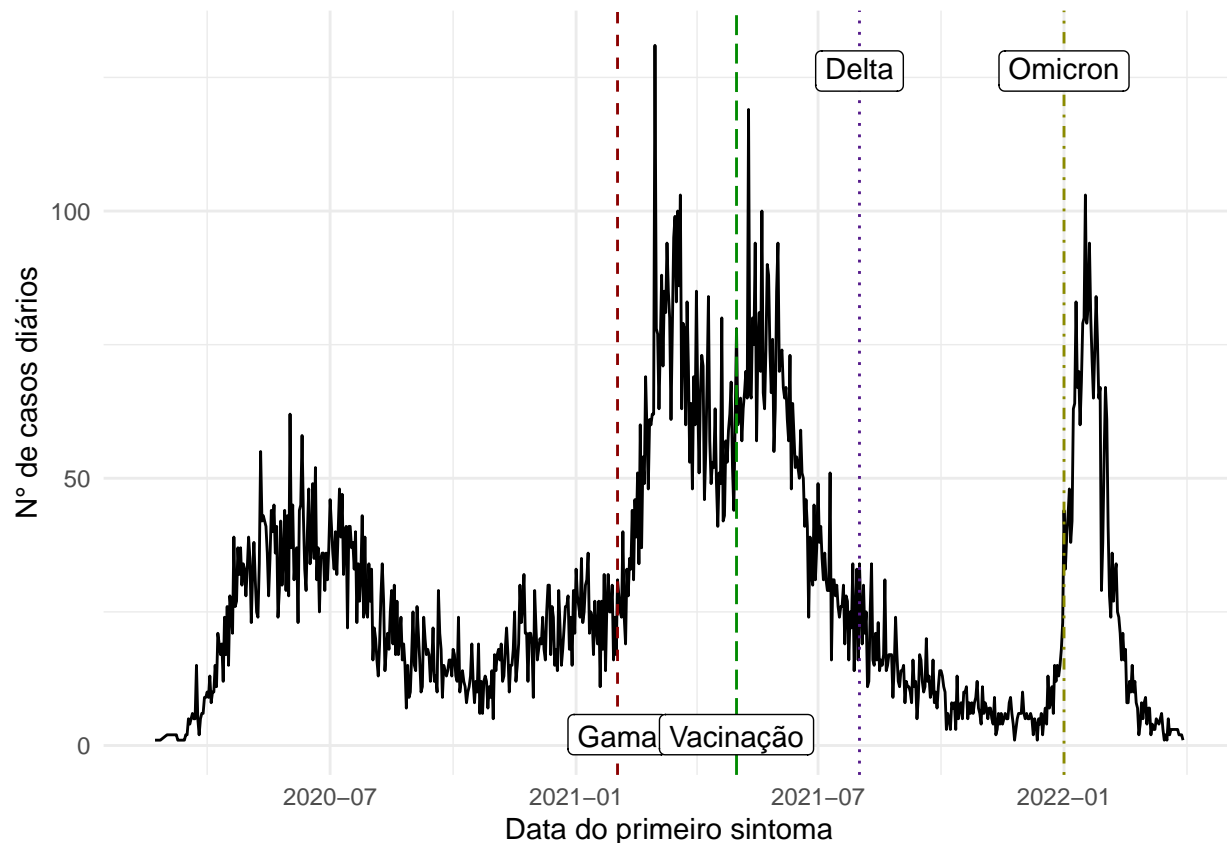
```



No gráfico acima podemos ver que a variante gama além de ter o maior pico de casos diários, teve uma constância maior em relação ao período de outras variantes como a delta que não teve um contágio alto e a omicron que teve um pico relevante de casos diários mas por um intervalo bem menor. Um dos fatores que podem ter impactado o número de casos que são notificados como SRAG é o início da vacinação. Não temos ao certo uma data em que a vacinação para gestantes e puérperas se iniciou no Brasil a nível nacional, mas

temos uma data de referência do início de maio. Sendo assim podemos agregar essa informação ao nosso gráfico anterior, que fica da seguinte forma:

```
dados %>%
  group_by(DT_SIN_PRI) %>%
  summarise(casos = n()) %>%
  ggplot(aes(DT_SIN_PRI, casos)) +
  geom_line() +
  geom_vline(xintercept = in_gama,
             colour = "red4",
             linetype = 2) +
  annotate(
    x = in_gama,
    y = -Inf,
    label = "Gama",
    vjust = -0.5,
    geom = "label"
  ) +
  geom_vline(xintercept = in_delta,
             colour = "purple4",
             linetype = 3) +
  annotate(
    x = in_delta,
    y = +Inf,
    label = "Delta",
    vjust = 2,
    geom = "label"
  ) +
  geom_vline(xintercept = in_omicron,
             colour = "yellow4",
             linetype = 4) +
  annotate(
    x = in_omicron,
    y = +Inf,
    label = "Omicron",
    vjust = 2,
    geom = "label"
  ) +
  geom_vline(xintercept = as.Date("01-05-2021", format = "%d-%m-%Y"),
             colour = "green4",
             linetype = 5) +
  annotate(
    x = as.Date("01-05-2021", format = "%d-%m-%Y"),
    y = -Inf,
    label = "Vacinação",
    vjust = -0.5,
    geom = "label"
  ) +
  labs(
    x = "Data do primeiro sintoma",
    y = "N° de casos diários"
  )
```



Nesse caso fica claro que após um certo período do início da vacinação tivemos um decrescimento dos casos de SRAG por COVID-19 voltando a ter pico de casos somente com a variante omicron.

Evolução no tempo - óbitos

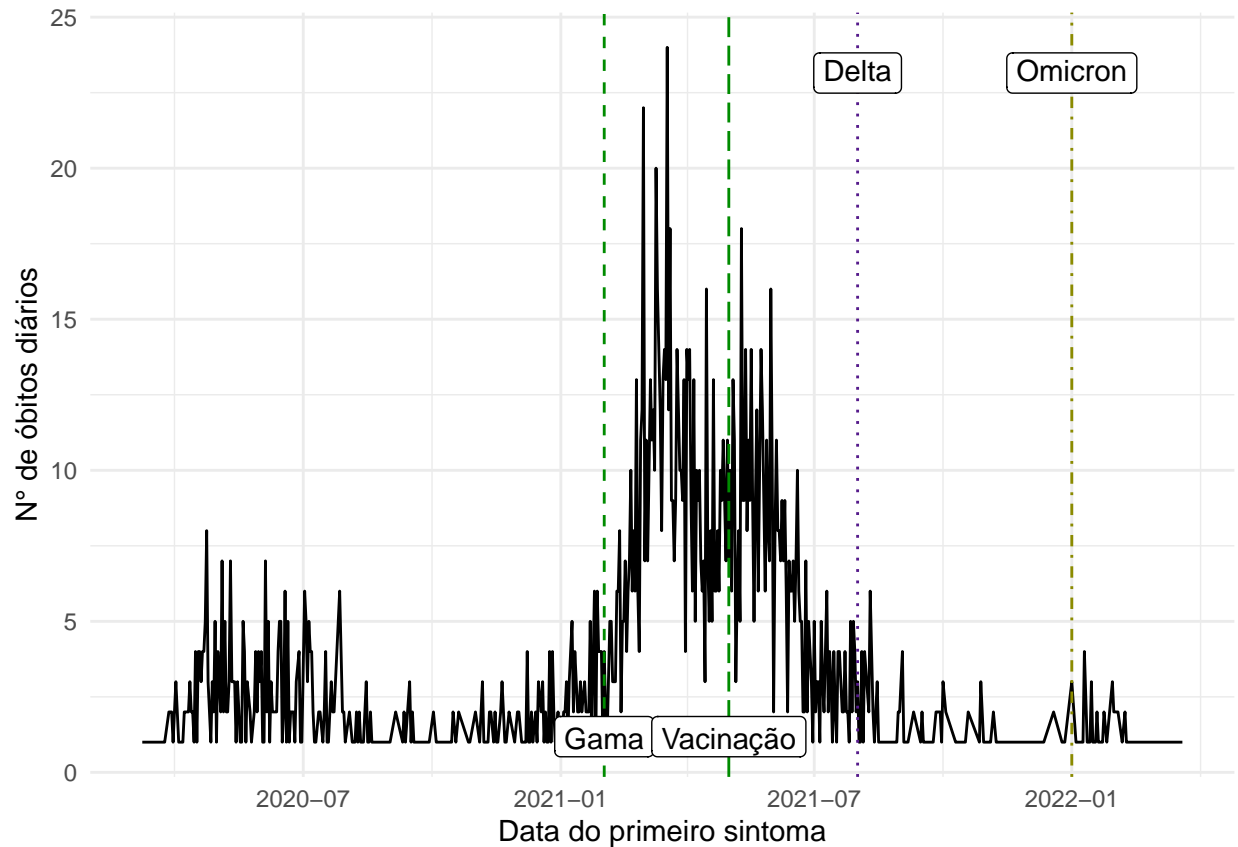
```
dados %>%
  filter(evolucao == "obito") %>%
  group_by(DT_SIN_PRI) %>%
  summarise(obitos = n()) %>%
  ggplot(aes(DT_SIN_PRI, obitos)) +
  geom_line() +
  geom_vline(xintercept = in_gama,
             colour = "green4",
             linetype = 2) +
  annotate(
    x = in_gama,
    y = -Inf,
    label = "Gama",
    vjust = -0.5,
    geom = "label"
  ) +
  geom_vline(xintercept = in_delta,
             colour = "purple4",
             linetype = 3) +
```



```

annotate(
  x = in_delta,
  y = +Inf,
  label = "Delta",
  vjust = 2,
  geom = "label"
) +
geom_vline(xintercept = in_omicron,
  colour = "yellow4",
  linetype = 4) +
annotate(
  x = in_omicron,
  y = +Inf,
  label = "Omicron",
  vjust = 2,
  geom = "label"
) +
geom_vline(xintercept = as.Date("01-05-2021", format = "%d-%m-%Y"),
  colour = "green4",
  linetype = 5) +
annotate(
  x = as.Date("01-05-2021", format = "%d-%m-%Y"),
  y = -Inf,
  label = "Vacinação",
  vjust = -0.5,
  geom = "label"
) +
labs(
  x = "Data do primeiro sintoma",
  y = "N° de óbitos diários"
)

```



Podemos ver que temos um comportamento que tem uma relação à evolução no tempo de número de casos, sendo a excessão mais clara a variante omicron pois os óbitos não seguem o pico de casos, tendo como possível explicação a maior quantidade de gestantes e puérperas vacinadas.

Tabela cruzada

Aqui podemos ver as tabelas cruzadas tendo como covariáveis as variantes e como variável resposta a evolução dos casos.

```
with(
  dados,
  ctable(
    x = variante,
    y = evolucao,
    prop      = 'r',
    totals    = FALSE,
    headings  = FALSE,
    chisq     = TRUE
  )
)
```

```
##
## -----
##          evolucao      cura      obito      <NA>
```

```
## variante
## original          6513 (85.4%)    533 ( 7.0%)    580 ( 7.6%)
## gama              7925 (78.1%)    1351 (13.3%)    869 ( 8.6%)
## delta             1317 (82.5%)     96 ( 6.0%)    184 (11.5%)
## omicron            2385 (89.0%)     47 ( 1.8%)    249 ( 9.3%)
## -----
##
## -----
## Chi.squared   df   p.value
## -----
##    451.07      3      0
## -----
```

Pela tabela acima vemos que a variante gama parece ter sido a pior variante para gestantes e puérperas. Vamos fazer algumas análises fixando a variante gama e comparando-a com as demais variantes.

Gama e Original

```
dados_var <- dados %>%
  filter(variante %in% c("gama", "original") &
         !is.na(evolucao))

dados_var$variante <- factor(dados_var$variante,
                             levels = c("original", "gama"))

with(
  dados_var,
  ctable(
    x = variante,
    y = evolucao,
    prop      = 'r',
    totals    = FALSE,
    headings  = FALSE,
    chisq     = TRUE,
    OR        = TRUE
  )
)
```

```
##
## -----
##           evolucao          cura          obito
## variante
## original          6513 (92.4%)    533 ( 7.6%)
## gama              7925 (85.4%)    1351 (14.6%)
## -----
##
## -----
## Chi.squared   df   p.value
## -----
##    191.4788    1      0
## -----
##
```

```
## -----
## Odds Ratio   Lo - 95%   Hi - 95%
## -----
##      2.08      1.87      2.31
## -----
```

Gama e Delta

```
dados_var <- dados %>%
  filter(variante %in% c("gama", "delta") &
         !is.na(evolucao))

dados_var$variante <- factor(dados_var$variante,
                             levels = c("delta", "gama"))

with(
  dados_var,
  ctable(
    x = variante,
    y = evolucao,
    prop      = 'r',
    totals    = FALSE,
    headings  = FALSE,
    chisq     = TRUE,
    OR        = TRUE
  )
)
```

```
##
## -----
##      evolucao      cura      obito
## variante
##      delta      1317 (93.2%)      96 ( 6.8%)
##      gama      7925 (85.4%)     1351 (14.6%)
## -----
##
## -----
## Chi.squared  df  p.value
## -----
##      62.5927    1      0
## -----
##
## -----
## Odds Ratio   Lo - 95%   Hi - 95%
## -----
##      2.34      1.89      2.90
## -----
```

Gama e Omicron

```
dados_var <- dados %>%
  filter(variante %in% c("gama", "omicron") &
         !is.na(evolucao))

dados_var$variante <- factor(dados_var$variante,
                             levels = c("omicron", "gama"))

with(
  dados_var,
  ctable(
    x = variante,
    y = evolucao,
    prop = 'r',
    totals = FALSE,
    headings = FALSE,
    chisq = TRUE,
    OR = TRUE
  )
)
```

```
##
## -----
##           evolucao           cura           obito
## variante
##   omicron           2385 (98.1%)           47 ( 1.9%)
##     gama           7925 (85.4%)          1351 (14.6%)
## -----
##
## -----
## Chi.squared   df   p.value
## -----
##    291.2008    1       0
## -----
##
## -----
## Odds Ratio   Lo - 95%   Hi - 95%
## -----
##      8.65      6.44      11.61
## -----
```

Conclusão

Vemos que para todos os casos acima temos Odds Ratio (OR) acima de 1 para as variantes quando comparadas a gama, ou seja, temos que no período das variantes original, delta e omicron as gestantes e puérperas tinham 2.08, 2.34 e 8.65 mais chances de se curar, respectivamente, se comparadas com o período da variante gama.