

## 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.

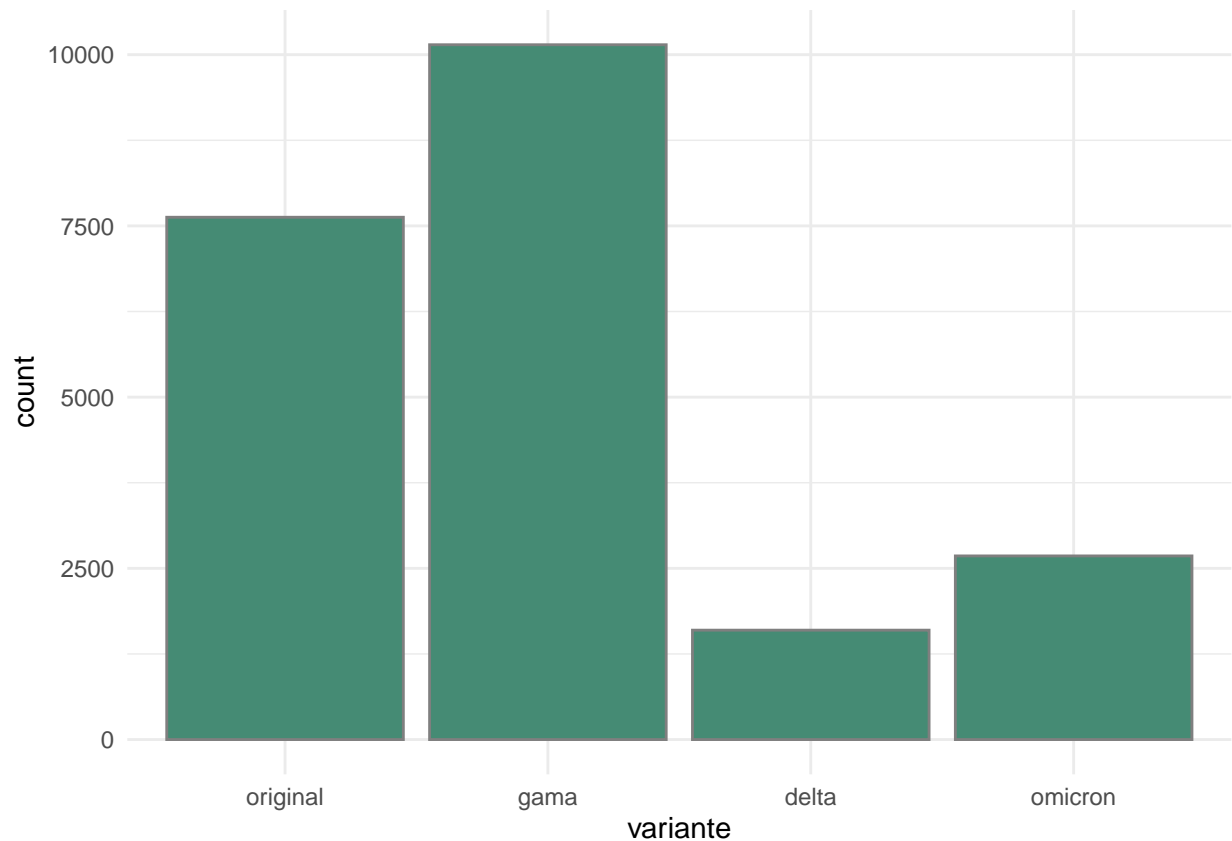
## Informações resumo das variáveis principais

```
## Data Frame Summary
## dados
## Dimensions: 22049 x 8
## Duplicates: 19860
## 
## -----
## Variable      Stats / Values    Freqs (% of Valid)   Graph                Missing
## -----
## variante      1. original       7626 (34.6%)         IIIIII              0
## [factor]      2. gama           10145 (46.0%)        IIIIIIIII          (0.0%)
##               3. delta          1597 ( 7.2%)         I
##               4. omicron        2681 (12.2%)         II
## 
## vacina_cov     1. não            14062 (81.9%)        IIIIIIIIIIIIIIIII  4877
## [character]   2. sim            3110 (18.1%)         III                 (22.1%)
## 
## evolucao       1. cura           18140 (89.9%)        IIIIIIIIIIIIIIIII  1882
## [character]   2. obito          2027 (10.1%)         II                  (8.5%)
## 
## suport_ven     1. não            10018 (51.7%)        IIIIIIIIII         2662
## [factor]      2. não invasivo   6760 (34.9%)         IIIIII             (12.1%)
##               3. invasivo       2609 (13.5%)         II
## 
## uti            1. não            14062 (71.4%)        IIIIIIIIIIIIIIIII  2366
## [character]   2. sim            5621 (28.6%)         IIIII              (10.7%)
## 
## febre          1. não            7492 (40.3%)         IIIIIIII           3441
## [character]   2. sim            11116 (59.7%)        IIIIIIIIIII        (15.6%)
## 
## tosse          1. não            5135 (26.4%)         IIIII              2591
## [character]   2. sim            14323 (73.6%)        IIIIIIIIIIIIIIIII (11.8%)
```

```
##
## fadiga      1. não      9896 (70.9%)      IXXXXXXXXXXXXX      8099
## [character] 2. sim      4054 (29.1%)      IXXXXX              (36.7%)
## -----
```

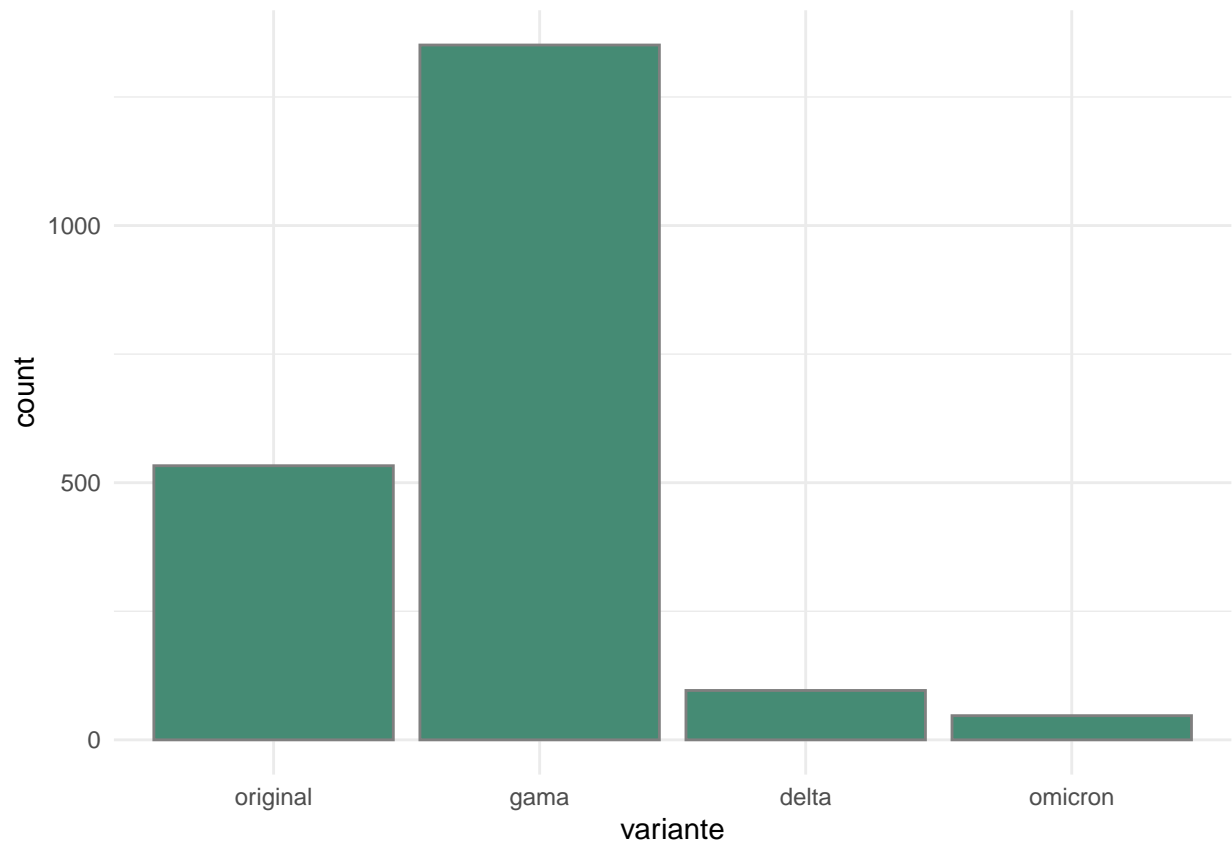
## Gráficos de frequência

Casos para cada uma das variantes

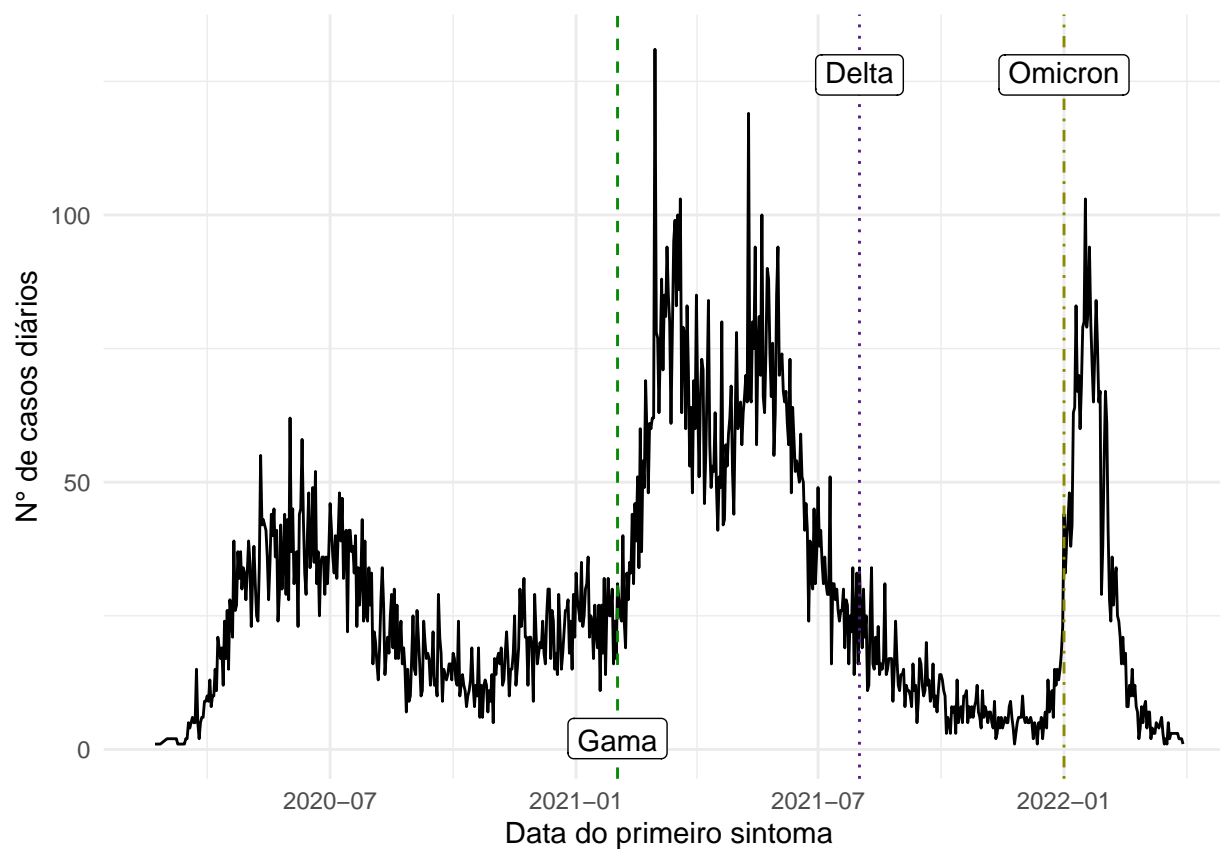


Podemos ver que, para a população de gestantes se puerperas, a variante gama foi a que teve maior número de casos. No tempo podemos tirar algumas informações mais concretas sobre essa questão. Para isso iremos gerar o gráfico a seguir.

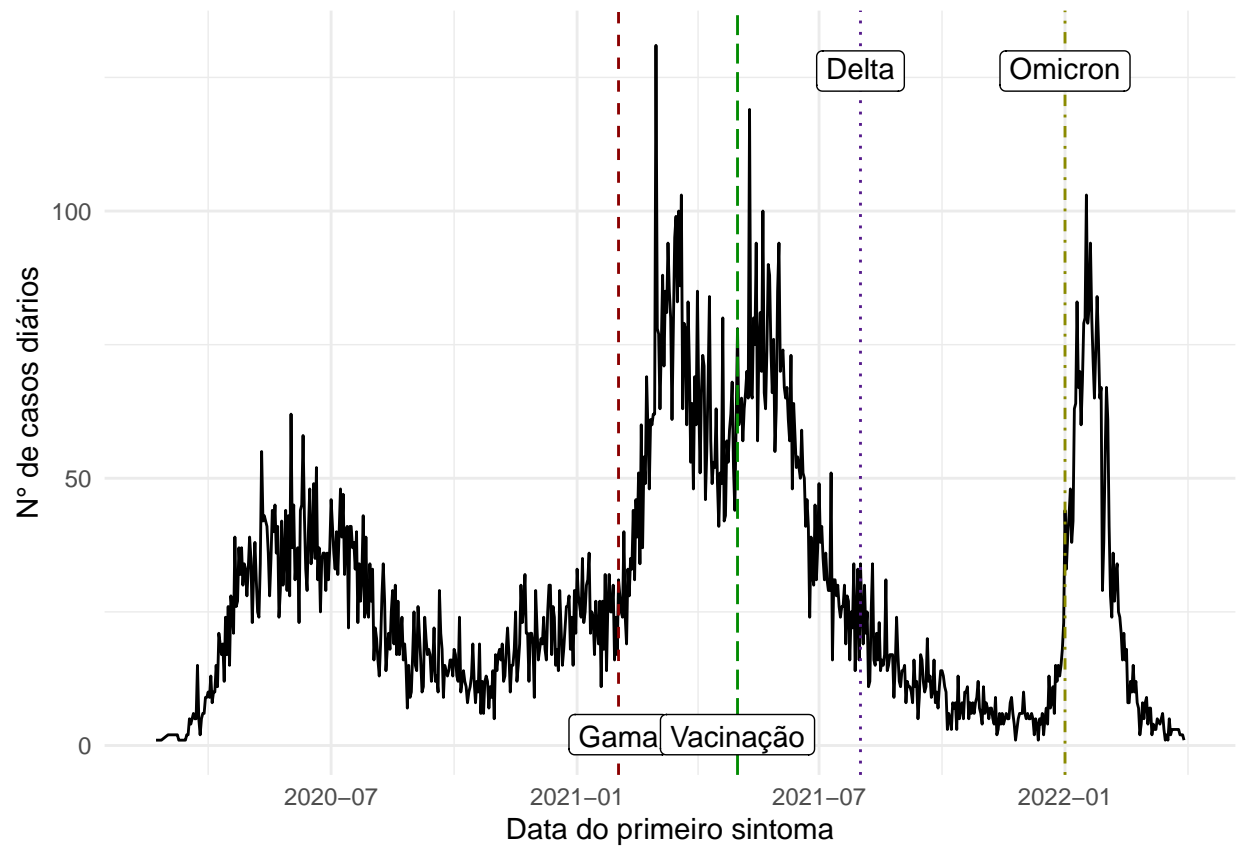
## Óbitos para cada uma das variantes



## Evolução no tempo - casos

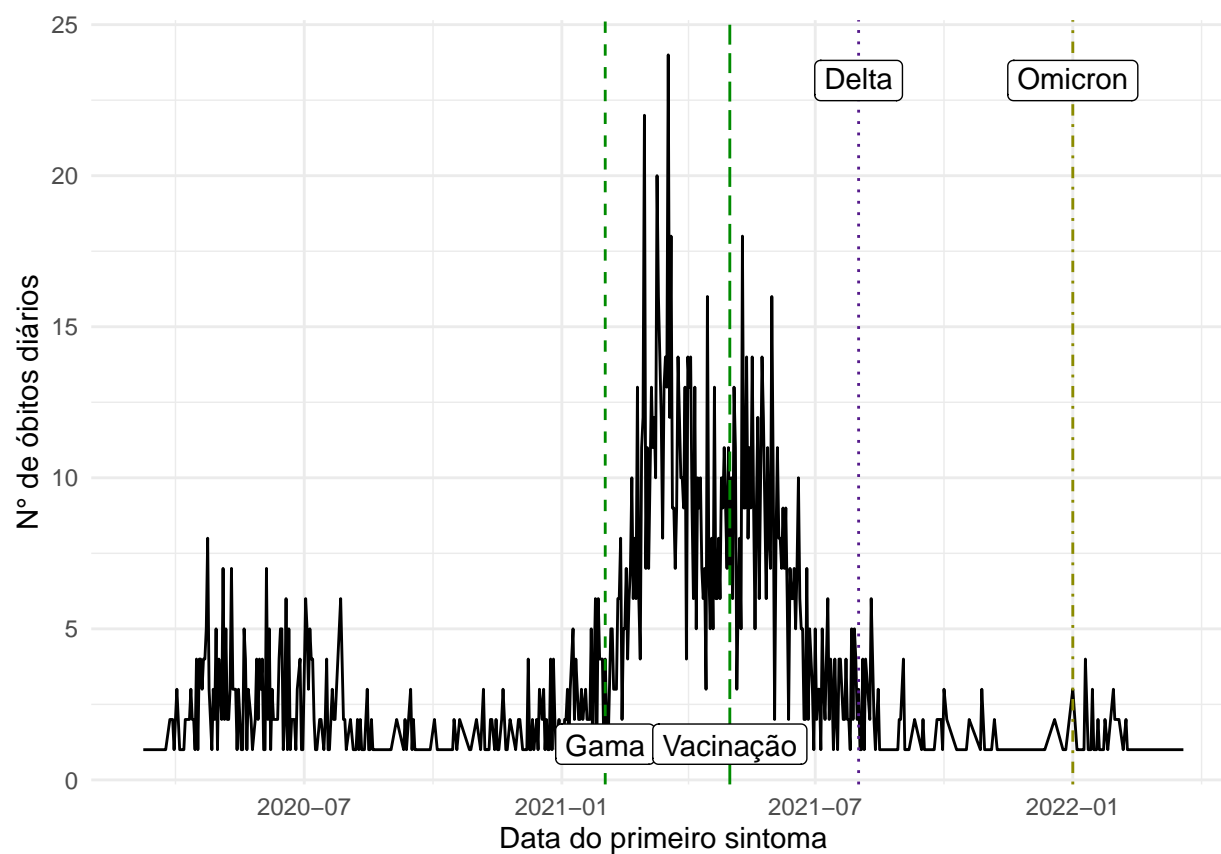


No gráfico acima podemos ver que a variante gama além de ter o maior pico de casos diários, teve uma constância maior em relação ao período de outras variantes como a delta que não teve um contágio alto e a omicron que teve um pico relevante de casos diários mas por um intervalo bem menor. Um dos fatores 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:



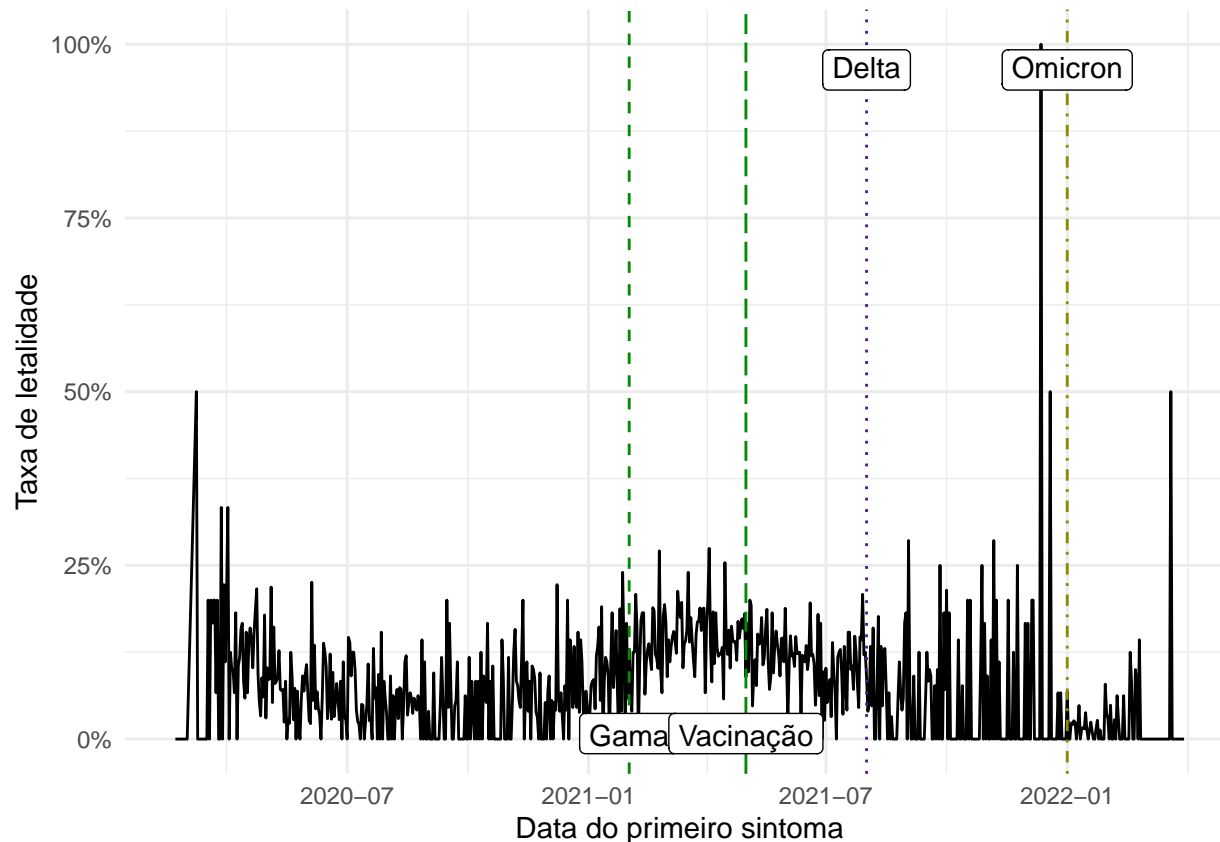
Nesse caso fica claro que após um certo período do início da vacinação tivemos um decrescimento dos casos de SRAG por COVID-19 voltando a ter pico de casos somente com a variante omicron.

## Evolução no tempo - óbitos



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.

## Evolução no tempo - taxa de letalidade



## Tabela cruzada

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

```
##
## -----
##          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

```
##
## -----
##           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

```
##
## -----
##           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

```
##
## -----
##           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

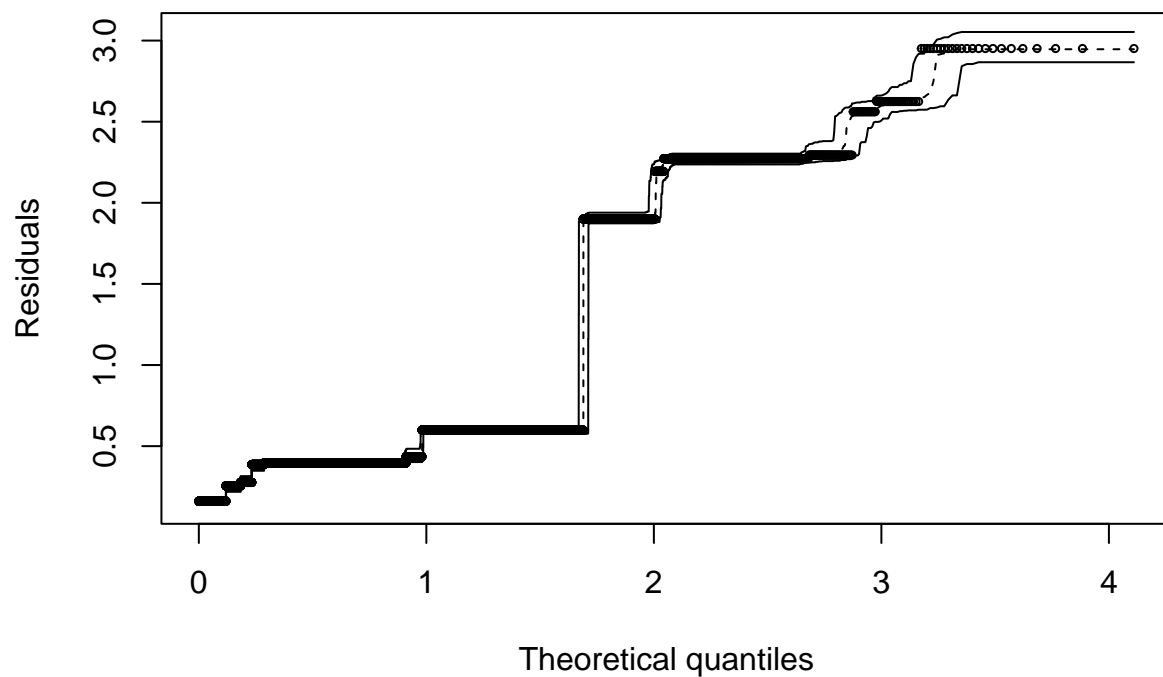
Vemos que para todos os casos acima temos Odds Ratio (OR) acima de 1 para as variantes quando comparadas a gama, ou seja, temos que no período das variantes original, delta e omicron as gestantes e puérperas tinham 2.08, 2.34 e 8.65 mais chances de se curar, respectivamente, se comparadas com o período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

## Modelos

### Evolução

```
##
## Call:
## glm(formula = evolucao ~ vacina_cov + variante, family = "binomial",
##      data = dados)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5993  -0.4348  -0.3966  -0.2769   2.9510
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.50303    0.04505 -55.558 < 2e-16 ***
## vacina_covsim  -0.93076    0.11898  -7.823 5.16e-15 ***
## variantegama    0.87702    0.06017  14.576 < 2e-16 ***
## variantedelta   0.19196    0.12983   1.479  0.139
## varianteomicron -0.90761    0.16403  -5.533 3.14e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 9653.6  on 15805  degrees of freedom
## Residual deviance: 9146.3  on 15801  degrees of freedom
## (6243 observations deleted due to missingness)
## AIC: 9156.3
##
## Number of Fisher Scoring iterations: 6
```

```
##
## The Hosmer-Lemeshow goodness-of-fit test
##
## Group Size Observed Expected
## 1 1520 24 19.53
## 2 828 22 26.47
## 3 623 18 23.44
## 4 709 52 51.03
## 5 7046 533 533.00
## 6 627 62 56.56
## 7 4453 731 731.97
##
## Statistic = 3.72062
## degrees of freedom = 5
## p-value = 0.5903
##
## Binomial model
```



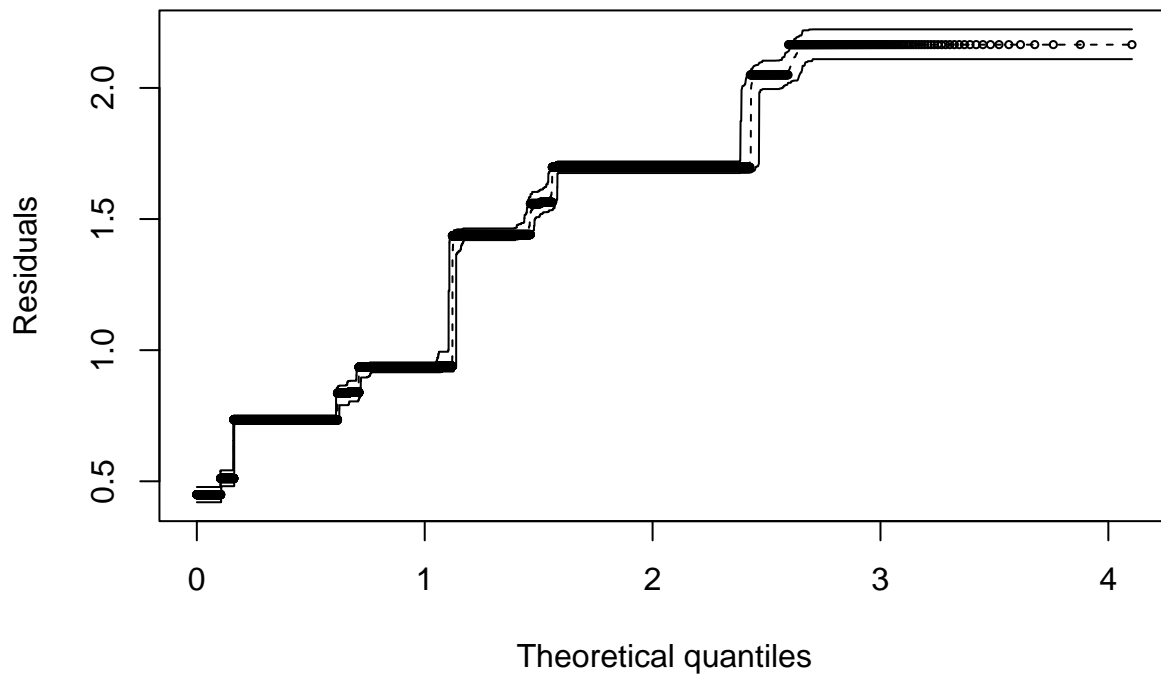
Uti

```
##
## Call:
## glm(formula = uti ~ vacina_cov + variante, family = "binomial",
## data = dados)
##
```

```

## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.9388  -0.8354  -0.7349   1.4364   2.1651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.17109    0.02869 -40.825 < 2e-16 ***
## vacina_covsim  -0.27247    0.06424  -4.241 2.22e-05 ***
## variantegama    0.57023    0.04172  13.668 < 2e-16 ***
## variantedelta   0.58017    0.07299   7.949 1.89e-15 ***
## varianteomicron -0.79929    0.08413  -9.500 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 17674  on 15363  degrees of freedom
## Residual deviance: 17091  on 15359  degrees of freedom
## (6685 observations deleted due to missingness)
## AIC: 17101
##
## Number of Fisher Scoring iterations: 4
##
##      The Hosmer-Lemeshow goodness-of-fit test
##
##      Group Size Observed Expected
##      1 1422      145   136.47
##      2  789       88    96.53
##      3 6727     1592  1592.00
##      4  670      197   197.36
##      5  631      179   187.18
##      6 4483     1588  1587.64
##      7  642      237   228.82
##
##      Statistic = 2.41213
## degrees of freedom = 5
##      p-value = 0.78967
##
## Binomial model

```



### Suporte ventilatório

```
##
## Call:
## glm(formula = suport_ven ~ vacina_cov + variante, family = "binomial",
##      data = dados)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.434  -0.996  -0.613   1.051   1.878
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.44305    0.02520 -17.579  < 2e-16 ***
## vacina_covsim -0.28100    0.05784  -4.858 1.19e-06 ***
## variantegama   1.02879    0.03931  26.172  < 2e-16 ***
## variantedelta  0.53032    0.06809   7.789 6.77e-15 ***
## varianteomicron -0.85241    0.06983 -12.207  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20885  on 15169  degrees of freedom
## Residual deviance: 19393  on 15165  degrees of freedom
```

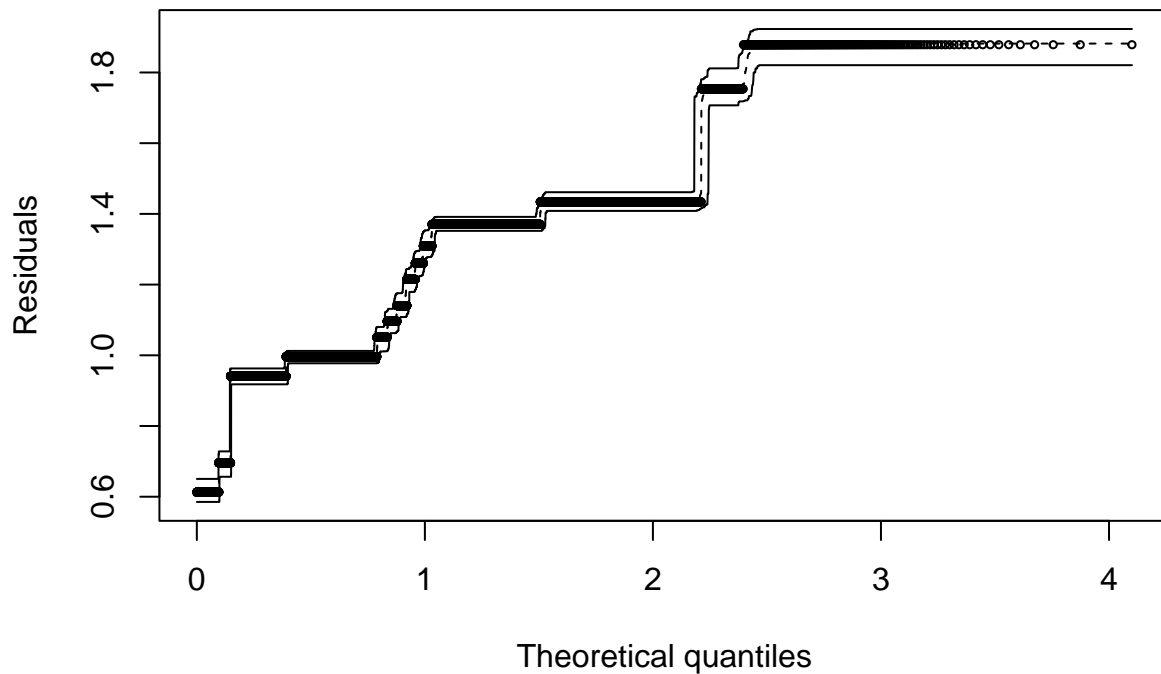
```
## (6879 observations deleted due to missingness)
## AIC: 19403
##
## Number of Fisher Scoring iterations: 4
```

```
##
## The Hosmer-Lemeshow goodness-of-fit test
##
```

| Group | Size | Observed | Expected |
|-------|------|----------|----------|
| 1     | 1414 | 251      | 242.2    |
| 2     | 776  | 158      | 166.8    |
| 3     | 6611 | 2585     | 2585.0   |
| 4     | 629  | 266      | 284.1    |
| 5     | 636  | 350      | 331.9    |
| 6     | 670  | 395      | 385.7    |
| 7     | 4434 | 2839     | 2848.3   |

```
##
##      Statistic = 5.77564
## degrees of freedom = 5
##      p-value = 0.32867
```

```
## Binomial model
```



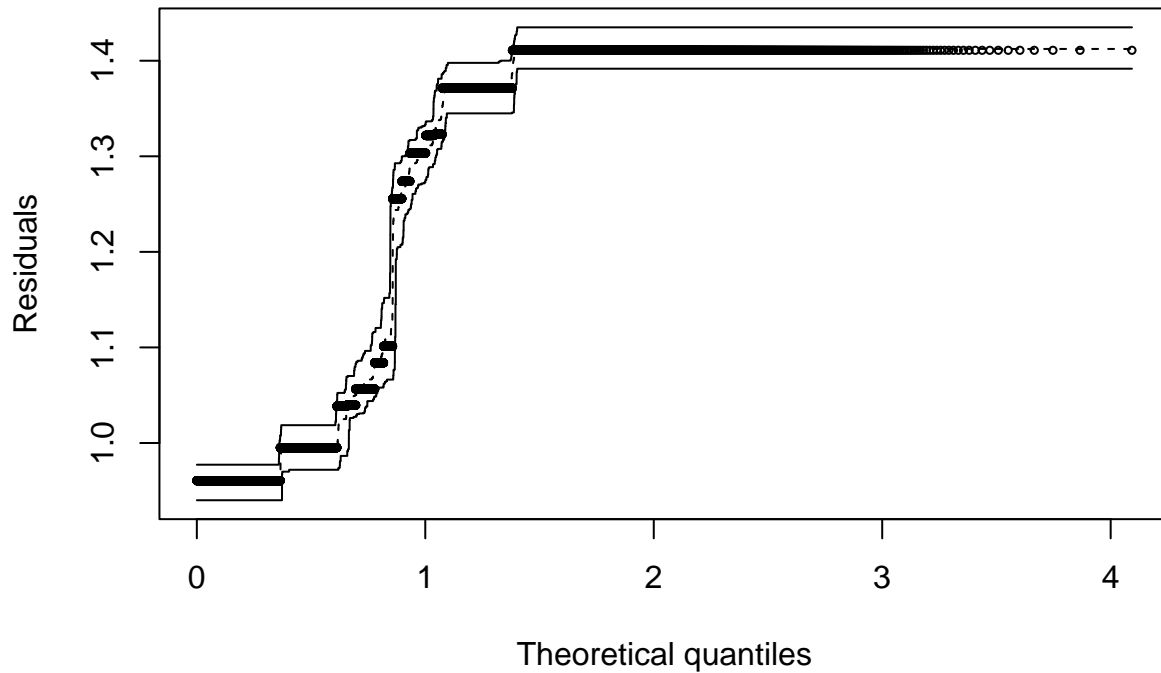
```
Febre
```

```
##
```

```

## Call:
## glm(formula = febre ~ vacina_cov + variante, family = "binomial",
##      data = dados)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4109  -1.3714   0.9606   0.9951   1.3034
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.53393    0.02541  21.009 < 2e-16 ***
## vacina_covsim  -0.10932    0.05600  -1.952  0.05093 .
## variantegama   -0.08874    0.03942  -2.251  0.02438 *
## variantedelta  -0.20067    0.06922  -2.899  0.00374 **
## varianteomicron -0.71591    0.06426 -11.141 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 19863  on 14689  degrees of freedom
## Residual deviance: 19627  on 14685  degrees of freedom
## (7359 observations deleted due to missingness)
## AIC: 19637
##
## Number of Fisher Scoring iterations: 4
##
##      The Hosmer-Lemeshow goodness-of-fit test
##
##      Group Size Observed Expected
##      1 1251      507      535.0
##      2  651      324      296.0
##      3  599      329      332.9
##      4  613      361      357.1
##      5  662      418      386.1
##      6 4269     2570     2601.9
##      7 6645     4189     4189.0
##
##      Statistic = 14.98013
## degrees of freedom = 5
##      p-value = 0.010448
##
## Binomial model

```



## Fadiga

```
##
## Call:
## glm(formula = fadiga ~ vacina_cov + variante, family = "binomial",
##      data = dados)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9317  -0.9277  -0.7120   1.4448   1.8594
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.24292    0.04083  -30.443  < 2e-16 ***
## vacina_covsim  -0.01060    0.06628   -0.160  0.87292
## variantegama    0.63321    0.05217  12.137  < 2e-16 ***
## variantedelta   0.39952    0.08422   4.744  2.1e-06 ***
## varianteomicron -0.27967    0.08763  -3.192  0.00141 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 12649  on 10701  degrees of freedom
## Residual deviance: 12386  on 10697  degrees of freedom
```

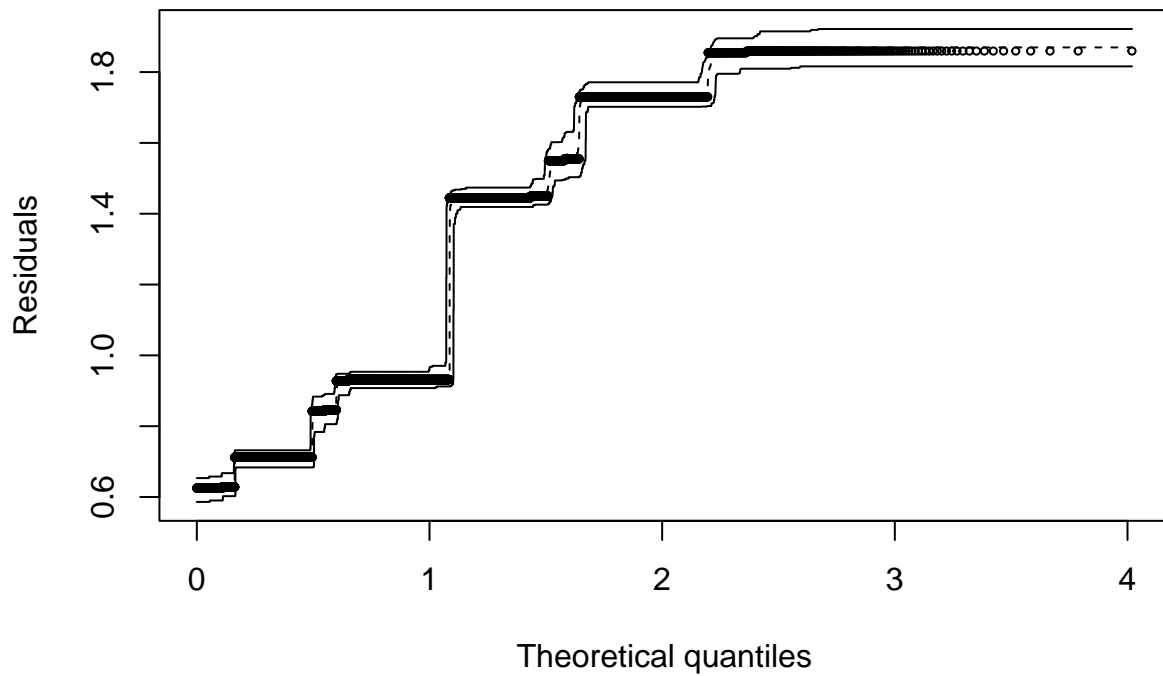
```
## (11347 observations deleted due to missingness)
## AIC: 12396
##
## Number of Fisher Scoring iterations: 4
```

```
##
## The Hosmer-Lemeshow goodness-of-fit test
##
```

| Group | Size | Observed | Expected |
|-------|------|----------|----------|
| 1     | 1127 | 194      | 200.07   |
| 2     | 558  | 106      | 99.93    |
| 3     | 3452 | 773      | 773.00   |
| 4     | 537  | 148      | 160.35   |
| 5     | 554  | 179      | 166.65   |
| 6     | 585  | 223      | 204.58   |
| 7     | 3889 | 1351     | 1369.42  |

```
##
##      Statistic = 6.26941
## degrees of freedom = 5
##      p-value = 0.28088
```

```
## Binomial model
```



Tosse

```
##
```



```

## Call:
## glm(formula = tosse ~ vacina_cov + variante, family = "binomial",
##      data = dados)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7470  -1.3463   0.7002   0.8115   1.0174
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.94163    0.02698  34.896 < 2e-16 ***
## vacina_covsim  -0.02571    0.06047  -0.425   0.6707
## variantegama    0.33933    0.04403   7.706 1.29e-14 ***
## variantedelta   0.13977    0.07583   1.843   0.0653 .
## varianteomicron -0.52712    0.06618  -7.965 1.66e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 17873  on 15233  degrees of freedom
## Residual deviance: 17631  on 15229  degrees of freedom
## (6815 observations deleted due to missingness)
## AIC: 17641
##
## Number of Fisher Scoring iterations: 4
##
##      The Hosmer-Lemeshow goodness-of-fit test
##
##      Group Size Observed Expected
##      1 1323      769    788.5
##      2  690      435    415.5
##      3 6804     4895   4895.0
##      4  635      471    471.1
##      5  644      481    480.9
##      6  697      562    542.4
##      7 4441     3456   3475.6
##
##      Statistic = 7.19253
## degrees of freedom = 5
##      p-value = 0.20671
##
## Binomial model

```

