



Universidade de Brasília
Departamento de Estatística

Modelos de séries temporais de contagem
aplicados a óbitos por Covid-19 no Distrito Federal

Marcelo Batalha Cunha

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

Marcelo Batalha Cunha

**Modelos de séries temporais de contagem
aplicados a óbitos por Covid-19 no Distrito Federal**

Orientador(a): Prof. Eduardo Yoshio Nakano

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Agradecimentos

Gostaria de expressar minha profunda gratidão às pessoas que foram fundamentais em minha jornada acadêmica e na realização deste trabalho.

Em primeiro lugar, agradeço de coração aos meus pais, Flávia Batalha e Frederico Jangola, e à minha avó, Linda Batalha, pelo apoio incondicional que sempre me deram. Sua presença constante e incentivo foram essenciais em cada etapa importante da minha vida.

Também quero estender meus agradecimentos aos meus colegas de graduação, cuja companhia e colaboração foram inestimáveis. Em especial, gostaria de mencionar Júlia Birbeire, Sabrina França, Luana Feijão e Stephany Lima.

Além disso, gostaria de expressar meu profundo agradecimento ao meu namorado, Érico Kitahara. Sua compreensão, incentivo e encorajamento foram essenciais para minha motivação e sucesso neste trabalho.

Por fim, não posso deixar de expressar minha gratidão ao Professor Doutor Eduardo Yoshio Nakano, meu orientador, pela sua orientação precisa, apoio constante e disponibilidade ao longo do desenvolvimento deste trabalho. Sou imensamente grato também a todo o corpo docente que contribuiu para a minha formação, enriquecendo minha trajetória acadêmica.

A todos vocês, meus sinceros agradecimentos por terem sido parte integrante deste percurso, fornecendo o suporte e incentivo que me levaram a alcançar este momento significativo em minha vida acadêmica.

Resumo

Este trabalho tem como objetivo ajustar um modelo GLARMA para prever o número de óbitos semanais por Covid-19 no Distrito Federal e comparar os resultados com os obtidos através de um modelo ARIMA. Os dados abrangem o período de março de 2020 a março de 2023 e foram agregados de forma semanal. Os resultados mostraram que apesar do modelo ARIMA apresentar melhor previsão quando a contagem é alta, o modelo GLARMA se destaca quando as contagens são baixas. De fato, nesta última situação o modelo ARIMA pode apresentar valores negativos.

Palavras-chaves: Covid-19, desempenho preditivo, GLARMA, ARIMA, séries temporais, previsão

Lista de Tabelas

1	Visão geral dos dados observados.	14
2	Coefficientes lineares, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.	22
3	Coefficientes lineares, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.	23
4	Coefficiente Binomial Negativa, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.	23
5	Resultado dos testes LR e Wald de verossimilhança.	24
6	Valores previstos, GLARMA Binomial Negativa e ARIMA com covariáveis.	29
7	Erro absoluto médio.	29

Lista de Figuras

1	Contagem de óbitos por semana em decorrência de Covid-19, DF, 03/2020-03/2023.	15
2	Contagem de novos casos de Covid-19 por semana, DF, 03/2020-03/2023.	16
3	Porcentagem de indivíduos vacinados contra Covid-19 com a primeira dose ou dose única, DF, 03/2020-03/2023.	17
4	Novos casos vs contagem de óbitos por Covid-19 por onda, DF, 03/2020-03/2023.	18
5	Novos casos vs contagem de óbitos por Covid-19 por porcentagem de vacinação, DF, 03/2020-03/2023.	18
6	Número de óbitos semanais ajustado pelo modelo GLARMA(3,0) Poisson considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.	20
7	Diagnóstico do modelo GLARMA(3,0) Poisson ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.	20
8	Número de óbitos semanais ajustado pelo modelo GLARMA(1,2) Binomial Negativa considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.	21
9	Diagnóstico do modelo GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.	22
10	Número de óbitos semanais ajustado pelo modelo GLARMA(1,2) Binomial Negativa considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.	23
11	Previsão do número de óbitos semanais a partir de 23/12/2021, modelo GLARMA Binomial Negativa.	25
12	Diferença entre valores previstos e observados a partir de 23/12/2021, modelo GLARMA Binomial Negativa.	25
13	Previsão do número de óbitos semanais a partir de 23/12/2021, modelo ARIMA com covariáveis.	27

14	Diferença entre valores previstos e observados a partir de 23/12/2021, modelo ARIMA com covariáveis.	27
15	Previsão do número de óbitos semanais a partir de 23/12/2021, ARIMA com covariáveis e GLARMA Binomial Negativa.	28
16	Diferença entre valores previstos e observados a partir de 23/12/2021, ARIMA com covariáveis e GLARMA Binomial Negativa.	29

Sumário

1 Introdução	8
2 Metodologia	9
2.1 Análise de séries temporais	9
2.1.1 Estacionariedade	9
2.2 ARMA	9
2.3 Modelo ARIMA	10
2.4 Modelo GARMA	10
2.5 Modelo GLARMA	11
2.6 Erro absoluto médio (MAE).	12
2.7 Conjunto de dados	13
3 Resultados	14
3.1 Análise descritiva.	14
3.1.1 Visão geral	14
3.1.2 Óbitos por semana	14
3.1.3 Casos por semana	15
3.1.4 Vacinação por semana	16
3.1.5 Número de óbitos vs número de casos	17
3.2 Ajuste do modelo GLARMA	19
3.3 Previsão GLARMA	24
3.4 Comparação com modelo ARIMA	26
4 Conclusão	31
Referências	32

1 Introdução

O objetivo deste trabalho é obter um modelo de predição do número de óbitos por Covid-19 ocorridos no Distrito Federal. Os dados foram obtidos por meio de um estudo retrospectivo, descritivo e longitudinal, que resultou em 11.843 ocorrências de óbitos por coronavírus, registradas pela Secretaria de Saúde do Distrito Federal no período de 25/02/2020 a 26/02/2023.

A variável de interesse deste trabalho é caracterizada como um dado de contagem, pois são os registros da frequência de óbitos em um intervalo de tempo (semanas). Além disso, esses registros são avaliados em intervalos sucessivos, resultando em uma possível dependência entre as observações, sendo caracterizado como uma série temporal.

Um dos modelos mais utilizados para o ajuste de uma série temporal é o modelo ARMA (Autorregressivo de Média Móvel), proposto por Box e Jenkins (1970). No entanto, esse modelo pode não ser adequado para o ajuste de uma série temporal de dados de contagem, visto que o mesmo supõe que a série segue uma distribuição Normal. De fato, nesta situação, o ideal é considerar uma distribuição de probabilidade discreta para inteiros não negativos, como a Poisson ou a Binomial Negativa.

Existem na literatura trabalhos que propõem procedimentos para a modelagem de séries temporais que associam o MLG (modelo linear generalizado) e o modelo ARMA. Dentre outros podemos citar o Modelo Autorregressivo Média Móvel Linear Generalizado – GLARMA (DAVIS; DUNSMUIR; STREETT, 2003) e o Modelo Autorregressivo Média Móvel Generalizado – GARMA (BENJAMIN; RIGBY; STASINOPOULOS, 2003). O GLARMA e o GARMA são extensões de um MLG que permitem modelar dados de contagem com dependência temporal dos dados.

Neste contexto, este trabalho propõe ajustar o modelo GLARMA para prever o número de ocorrências de óbitos no Distrito Federal e comparar os resultados com aqueles obtidos pelo modelo ARIMA. As análises serão feitas por meio do pacote *glarma* do R (R Core Team, 2022).

No decorrer desse estudo, será apresentada na seção 2 a metodologia de modelagem utilizada e a descrição do conjunto de dados e variáveis. Em seguida, na seção 3, está apresentada a análise descritiva das variáveis em estudo, bem como os resultados para o ajuste dos modelos, previsões e comparações. Por fim, na seção 4, está apresentada a conclusão deste trabalho com as considerações finais.

2 Metodologia

2.1 Análise de séries temporais

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes (EHLERS, 2009).

A análise de séries temporais consiste na investigação de variáveis observadas ao longo do tempo, com pretensão de realizar previsões para períodos futuros (SOUSA et al., 2021).

2.1.1 Estacionariedade

A estacionariedade em uma série temporal indica que os dados analisados oscilam, isto é, variam sobre uma média constante, independentemente do tempo, com a variância das flutuações permanecendo a mesma (SOUSA et al., 2021).

2.2 ARMA

A combinação dos modelos estacionários autorregressivo $AR(p)$ e médias móveis $MA(q)$ dá origem ao modelo autorregressivo médias móveis $ARMA(p, q)$, no qual a ordem do modelo está incluída nos parênteses como $ARMA(p, q)$, onde p é a ordem autorregressiva e q é a ordem de média móvel.

Conforme Morettin (2018), admitindo $E[y_t] = 0$, os modelos $ARMA(p, q)$ são dados por:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q},$$

onde a_t é um ruído branco com média zero e variância σ^2 e $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ são constantes.

Assim, admitindo-se $\Phi(B)$ e $\Theta(B)$ como os operadores autorregressivos, também chamados de operadores de defasagem e de médias móveis. A forma compacta dessa representação é:

$$\Phi(B)y_t = \Theta(B)a_t,$$

em que

$$\Phi(B)y_t = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

é o operador autorregressivo,

$$\Theta(B)a_t = \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

é o operador médias móveis e

$$B^n = y_{t-n}$$

é o operador de retardo.

2.3 Modelo ARIMA

O modelo autorregressivo integrado médias móveis, ARIMA, é uma extensão do modelo ARMA(p, q) e incorpora a diferenciação para tornar a série temporal estacionária. Assim, seja $w_t = \Delta^d y_t$ uma série temporal, é possível representar w_t por um modelo ARMA(p, q), ou seja, $\Phi(B)w_t = \Theta(B)a_t$. Se w_t for uma diferença de y_t , então y_t é uma integral de w_t , dessa forma, y_t segue um modelo autorregressivo integrado médias móveis ARIMA(p, d, q):

$$\Phi(B)\Delta^d y_t = \Theta(B)a_t,$$

onde p e q são as ordens de $\Phi(B)$ e $\Theta(B)$, respectivamente (MORETTIN, 2018).

2.4 Modelo GARMMA

No modelo autorregressivo média móvel generalizado (GARMMA), a distribuição condicional de cada observação y_t , dado a informação anterior como $W_t = \{x_t, \dots, x_1, y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$, pertence à mesma família exponencial, isto é, tem distribuição de densidade condicional dada por:

$$f(y_t|W_t) = \exp \left\{ \frac{y_t W_t - b(W_t)}{k} + c(y_t, k) \right\}, \quad (1)$$

onde W_t é o parâmetro canônico, também chamado de variável de estado, k é o parâmetro de dispersão, com $b(\cdot)$ e $c(\cdot)$ sendo funções específicas que definem uma família exponencial particular e x é um vetor de r variáveis explicativas.

Assim como em modelos lineares generalizados (MLG), μ , está relacionado ao preditor η_t e função g duplamente diferenciável e monótona, chamada de função de ligação. Diferente dos MLGs usuais, aqui o preditor linear tem um componente adicional, Z_t , que permite a inclusão dos termos autorregressivos e médias móveis de forma aditiva no preditor. Assim, o modelo geral para μ_t é dado por $g(\mu_t) = \eta_t = W_t = x_t^\top \beta + Z_t$, com:

$$Z_t = \sum_{j=1}^p \phi_j A(y_{t-j}, x_{t-j}, \beta) + \sum_{j=1}^q \theta_j M(y_{t-j}, \mu_{t-j}), \quad (2)$$

onde $A(\cdot)$ e $M(\cdot)$ são funções representando os termos autorregressivos e médias móveis com parâmetros desconhecidos, x_t é o vetor de regressores no tempo t , β é o vetor dos coeficientes de regressão estimados e ϕ e θ para os componentes autorregressivos e médias móveis respectivamente.

Benjamin, Rigby e Stasinopoulos (2003) afirmam que o modelo (2) é muito geral para aplicações práticas, então, considera-se o seguinte submodelo flexível e parcimonioso:

$$W_t = x_t^\top \beta + \sum_{j=1}^p \phi_j [g(y_{t-j}) - x_t^\top \beta] + \sum_{j=1}^q \theta_j [g(y_{t-j}) - W_{t-j}]. \quad (3)$$

As equações (1) e (3) juntas definem o modelo GARMA(p, q) (BENJAMIN; RIGBY; STASINOPOULOS, 2003).

2.5 Modelo GLARMA

Para descrever o modelo GLARMA da forma que está implementado no pacote *glarma* do *R*, é definida uma forma simplificada da família exponencial na qual o parâmetro de dispersão $k = 1$ e, para acomodar a distribuição binomial de maneira simples, define-se a função b para ser escalada por um fator a_t . Assim a forma simplificada de (1) é:

$$f(y_t|W_t) = \exp\{y_t W_t - a_t b(W_t) + c_t\},$$

onde a_t e c_t são sequências de constantes possivelmente dependentes das observações y_t .

Além disso, para os modelos GLARMA implementados no pacote *glarma* do *R* a variável de estado W_t é definida como:

$$W_t = x_t^\top \beta + O_t + Z_t,$$

onde há um termo adicional de *offset* incluído, pois Dunsmuir e Scott (2015) cita que essa adição pode ser importante para muitas aplicações.

Para previsões de um passo à frente de Y_{n+1} a partir da última observação no ponto n , dados os valores dos regressores x_{n+1} no ponto $n + 1$, então a distribuição condicional de Y_{n+1} pode ser prevista simplesmente prevendo a variável de estado W_{n+1} . Dessa forma, a operação pode ser feita usando a estimativa do parâmetro no modelo. Assim,

$$\begin{aligned} \hat{W}_{n+1} &= x_{n+1}^\top \hat{\beta} + \hat{Z}_{n+1}, \\ \hat{Z}_{n+1} &= \sum_{j=1}^p \hat{\phi}_j (\hat{Z}_{n+1-j} + \hat{e}_{n+1-j}) + \sum_{j=1}^q \hat{\theta}_j \hat{e}_{n+1-j} \end{aligned}$$

e

$$\hat{e}_{n+1-j} = \frac{Y_{n+1-j} - \mu_{n+1-j}}{\sigma_{n+1-j}}.$$

2.6 Erro absoluto médio (MAE)

O erro absoluto médio (MAE - *Mean Absolute Error*) é uma estatística comumente utilizada para avaliar e comparar o desempenho de previsões. Ele é obtido da seguinte forma:

$$\text{MAE} = \frac{1}{h} \sum_{i=1}^h |y_{n+i} - \hat{y}_{(n+i)|n}|,$$

onde h representa o horizonte de previsão, y_{n+i} o valor observado no tempo $n + i$ e $\hat{y}_{(n+i)|n}$ o valor previsto para o tempo $n + i$ dada a informação anterior até o tempo n .

2.7 Conjunto de dados

Os dados utilizados neste trabalho tem como fonte as informações de óbitos e número de casos de Covid-19 do Painel de Casos de Doença pelo Coronavírus (Ministério da Saúde, 2020) e as informações de vacinação de primeira dose ou dose única no Distrito Federal, obtidas no Vacinômetro - Covid-19 (Secretaria de Saúde do Distrito Federal, 2021).

Para adequar os dados aos modelos em estudo, realizou-se uma manipulação dos mesmos. Primeiramente, os dados foram agrupados de forma semanal, levando em consideração a contagem de óbitos, casos e doses de vacina aplicadas nos últimos sete dias, contando a partir de domingo. Além disso, foi calculada a porcentagem de vacinação, considerando a quantidade acumulada de doses aplicadas em relação à população projetada pelo IBGE para o ano de 2020.

Foi também adicionada a variável “tempo de onda”, uma vez que o período em que uma variante de Covid-19 se destaca entre as demais, conhecido como “onda de casos”, pode estar relacionado ao número de óbitos. Essa variável conta o número de dias desde o início da onda.

Para ajustar o modelo, as variáveis de número de casos na semana e porcentagem de vacinação foram atrasadas em uma semana. Dessa forma, para prever um passo à frente, é necessário conhecer apenas os valores da semana anterior, não da atual.

Essas manipulações e preparações dos dados foram realizadas com o intuito de ajustar adequadamente os modelos utilizados neste estudo, permitindo uma análise precisa e predições significativas.

3 Resultados

3.1 Análise descritiva

Apresenta-se a seguir uma breve análise descritiva sobre os dados de óbitos por Covid-19 ocorridos no DF no período de março de 2020 a março de 2023.

3.1.1 Visão geral

Abaixo, na Tabela 1, são apresentadas algumas estatísticas observadas no período em estudo.

Tabela 1: Visão geral dos dados observados.

Onda	Início	Fim	Duração (dias)	Óbitos	Casos	Porcentagem de óbitos	Casos por dia	Porcentagem média de vacinação
Wuhan	26/02/2020	31/10/2020	248	3633	209369	1,7%	844,2	0,0%
Beta	01/11/2020	31/08/2021	303	6391	259575	2,5%	856,7	16,7%
Delta	01/09/2021	31/12/2021	121	1071	49934	2,1%	412,7	75,6%
Omicron	01/01/2022	30/04/2022	119	543	176759	0,3%	1485,4	82,2%
Omicron B1A	01/05/2022	26/02/2023	301	205	201171	0,1%	668,3	85,7%

Na Tabela 1 é possível observar que a onda com maior porcentagem de óbitos (e também contagem) foi a Beta, com um total de 6391 óbitos, o qual representa 2,5% do total de casos observados durante o período da onda.

Na onda da variante Delta, apesar da média da porcentagem de vacinação relativamente alta, constata-se também alta porcentagem de óbitos, sendo a segunda maior dentre as observadas.

Por fim, as variantes Omicron e Omicron B1A apresentaram as menores porcentagens de óbitos. Entretanto, ressalta-se que, durante a onda da variante Omicron foi observado o maior número de novos casos por dia.

3.1.2 Óbitos por semana

A Figura 1 apresenta a contagem de óbitos em decorrência de coronavírus, por semana, no período de março de 2020 a março de 2023.

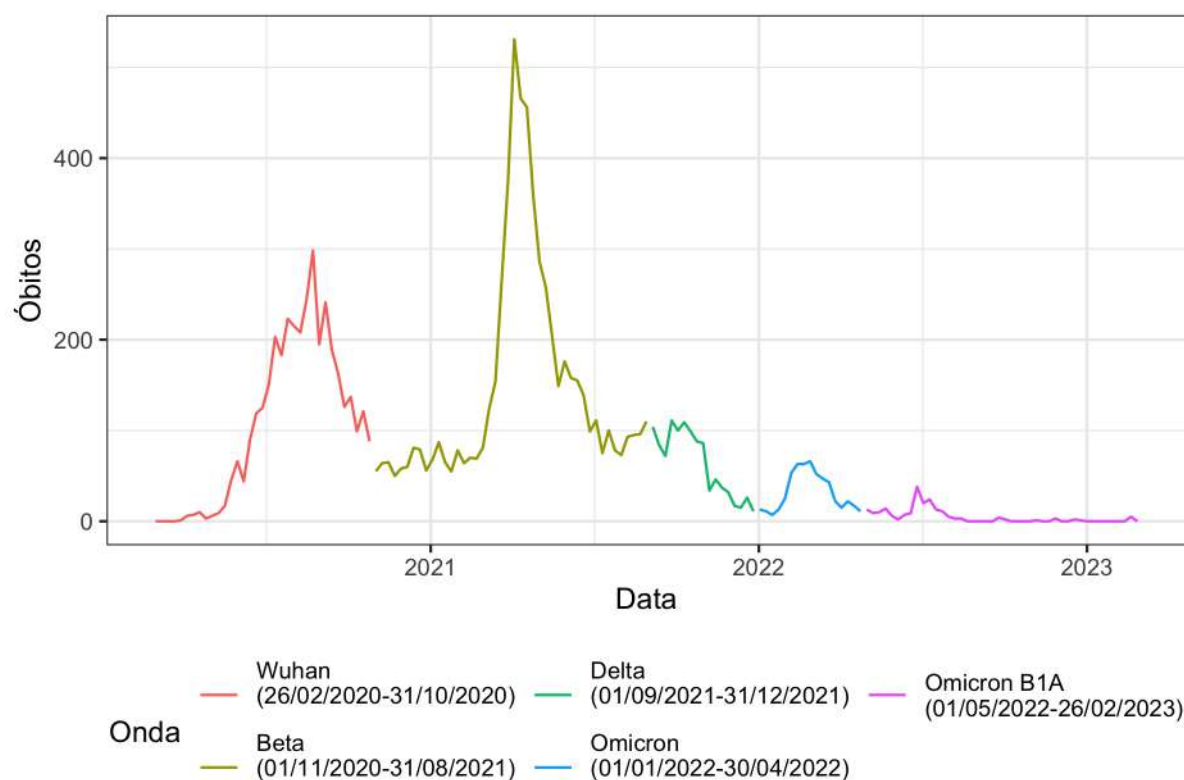


Figura 1: Contagem de óbitos por semana em decorrência de Covid-19, DF, 03/2020-03/2023.

Observa-se, na Figura 1, que durante a onda da variante Beta foi registrada a maior contagem de óbitos por semana, seguida pela onda da variante Wuhan, Delta, Omicron e Omicron B1A.

3.1.3 Casos por semana

Na Figura 2 está representado o número de novos casos de Covid-19, por semana, separados por onda.

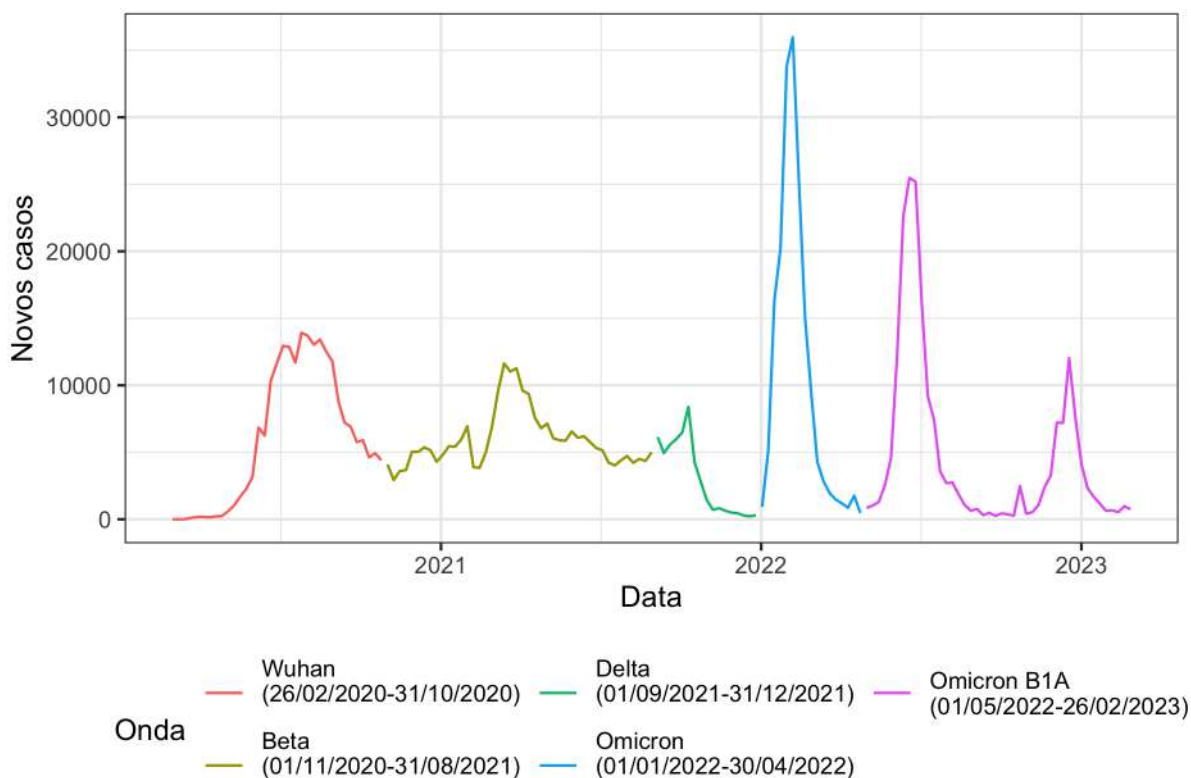


Figura 2: Contagem de novos casos de Covid-19 por semana, DF, 03/2020-03/2023.

Observa-se, na Figura 2, que as ondas das variantes Omicron e Omicron B1A, apesar de representarem as menores contagens de óbitos, são as que apresentam os maiores picos de novos casos por semana. Como visto na Tabela 1, durante a onda da variante Omicron foram registrados, em média, 1485 casos por dia, valor 73% superior a onda com segunda maior média (Beta).

3.1.4 Vacinação por semana

Está representada na Figura 3 a porcentagem de indivíduos vacinados no DF com a primeira dose ou dose única contra a Covid-19.

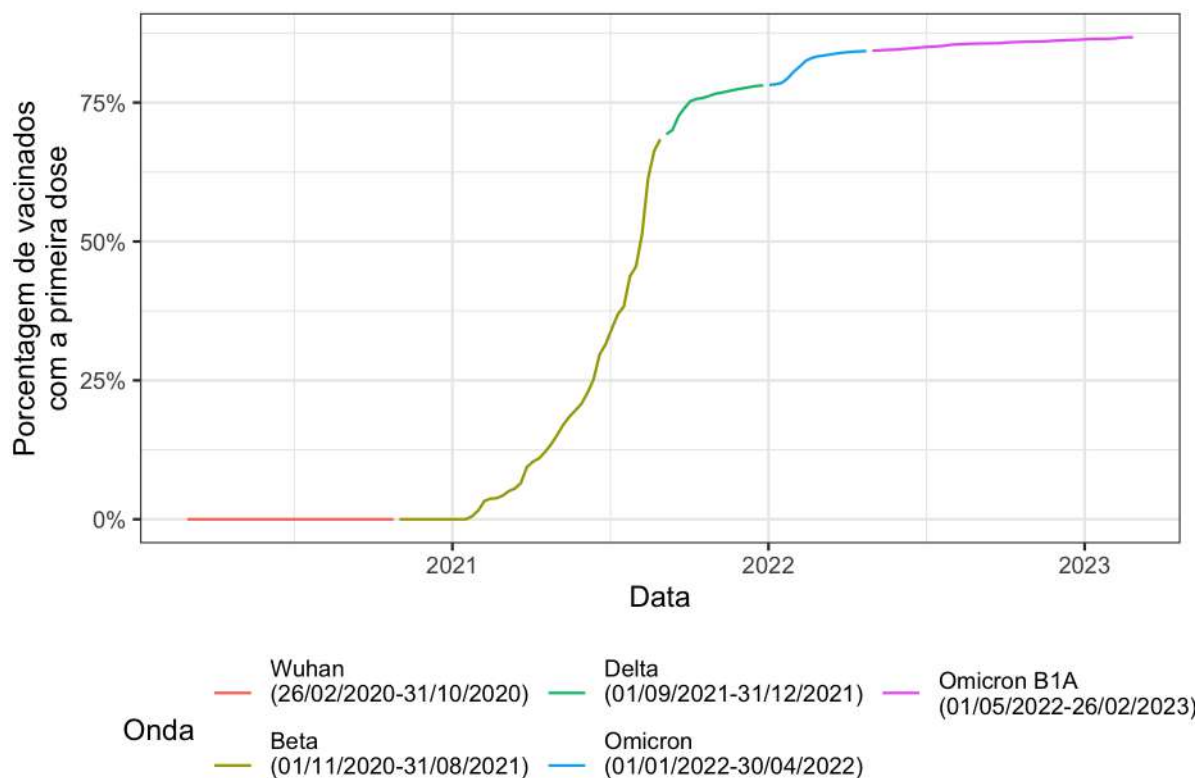


Figura 3: Porcentagem de indivíduos vacinados contra Covid-19 com a primeira dose ou dose única, DF, 03/2020-03/2023.

Visto que a vacinação contra a Covid-19 só começou em 2021 no DF, constata-se na Figura 3 que durante toda a onda da variante Wuhan a porcentagem de vacinação permanece em 0%.

Ademais, percebe-se que grande parte das doses de vacina foram aplicadas durante a onda da variante Beta, indo de 0% a 68,4% dos indivíduos vacinados ao fim da onda, totalizando 2.062.652 doses aplicadas ao todo.

3.1.5 Número de óbitos vs número de casos

A seguir, nas Figuras 4 e 5, está representada a relação entre o número de novos casos e contagem de óbitos por semana separados por onda e porcentagem de vacinação, respectivamente.

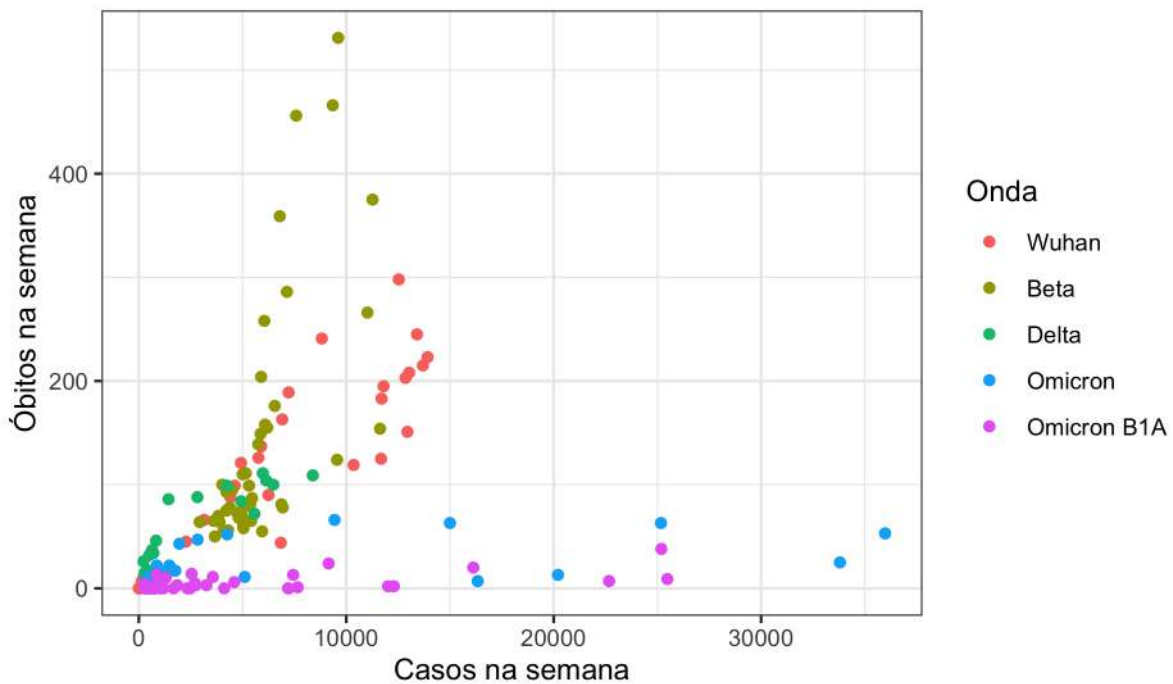


Figura 4: Novos casos vs contagem de óbitos por Covid-19 por onda, DF, 03/2020-03/2023.

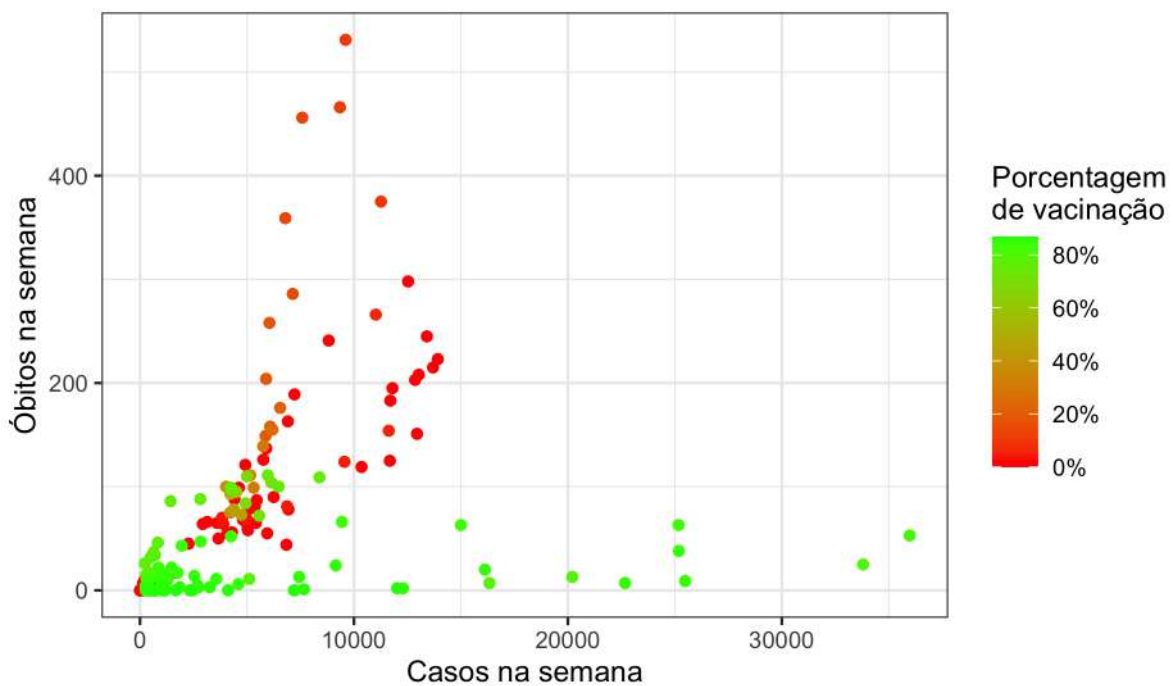


Figura 5: Novos casos vs contagem de óbitos por Covid-19 por porcentagem de vacinação, DF, 03/2020-03/2023.

Observando as Figuras 4 e 5, rapidamente percebe-se que as maiores contagens de óbitos por semana ocorreram durante as ondas Wuhan e Beta e quando a porcentagem de vacinação era menor ou igual a 20%. Além disso, nota-se na Figura 5 que com a

porcentagem de vacinação acima de 60% a maioria das observações ficam abaixo de 100 óbitos por semana.

Ademais, na Figura 4, observa-se que as ondas das variantes Omicron e Omicron B1A estão muito dispersas no eixo horizontal, com grandes números de novos casos observados, entretanto, apresentam número de óbitos relativamente baixo, quando comparado à outras ondas.

3.2 Ajuste do modelo GLARMA

Seguindo para o ajuste do modelo GLARMA, inicialmente foram consideradas as covariáveis: número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

Para a escolha dos parâmetros autorregressivo e médias móveis (p, q) , Dunsmuir e Scott (2015) afirma que ainda não existem métodos ou análises que deem boa orientação para a escolha, como há para as séries Gaussianas através da análise dos resíduos. Diante disso, para a seleção desses parâmetros o modelo foi testado com diferentes combinações de valores de p e q , com $p, q \leq 5$, e foi selecionada a combinação com menor valor de Critério de Informação de Akaike (AIC) (Akaike (1974)).

Além disso, inicialmente foi usada a distribuição Poisson e os parâmetros obtidos foram $p = 3$ e $q = 0$, com $AIC = 5206,34$. Os resultados do ajuste estão expressos nas Figuras 6 e 7.

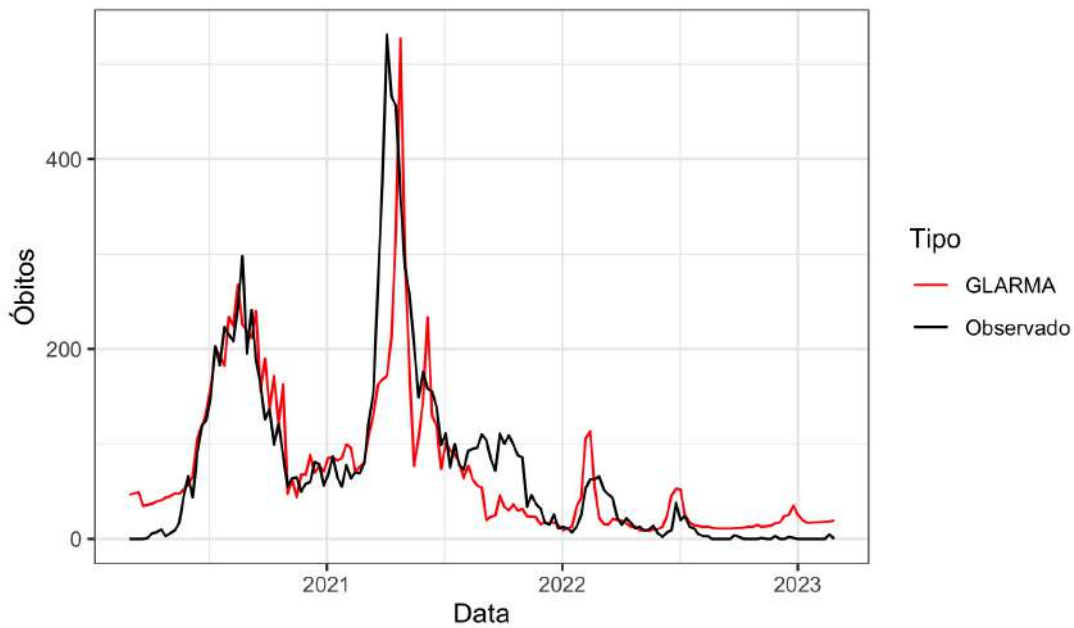


Figura 6: Número de óbitos semanais ajustado pelo modelo GLARMA(3,0) Poisson considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

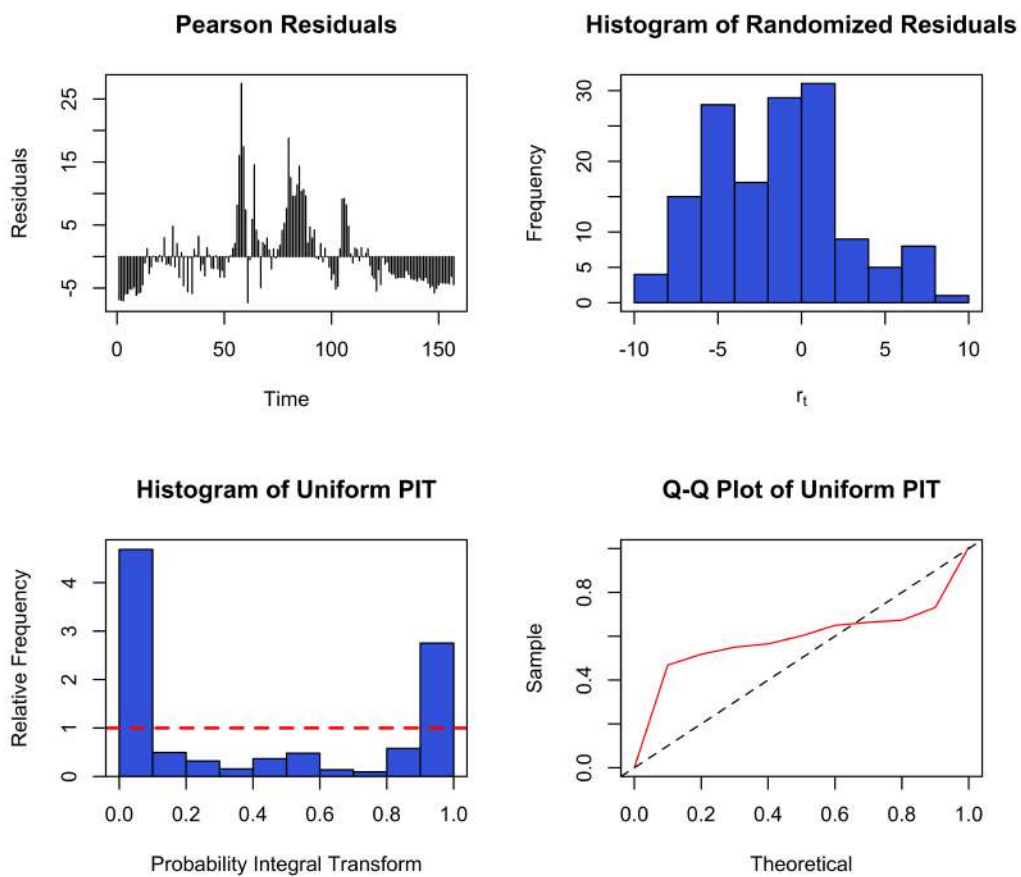


Figura 7: Diagnóstico do modelo GLARMA(3,0) Poisson ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

A análise da Figura 6 sugere um ajuste razoável, com grande diferença na iniciação e no ponto de máximo, e ao analisar os gráficos de diagnóstico gerados pelo modelo na Figura 7, percebe-se rapidamente nos gráficos *Q-Q Plot of Uniform PIT* e *Histogram of Uniform PIT* que a distribuição escolhida não é adequada para os dados e, portanto, será escolhida outra configuração de modelo com a distribuição Binomial Negativa.

Refazendo os testes usando a distribuição Binomial Negativa, foi obtido o modelo GLARMA(1,2) com $AIC = 1323,08$, valor consideravelmente menor ao encontrado anteriormente. Os resultados estão expressos nas Figuras 8 e 9.

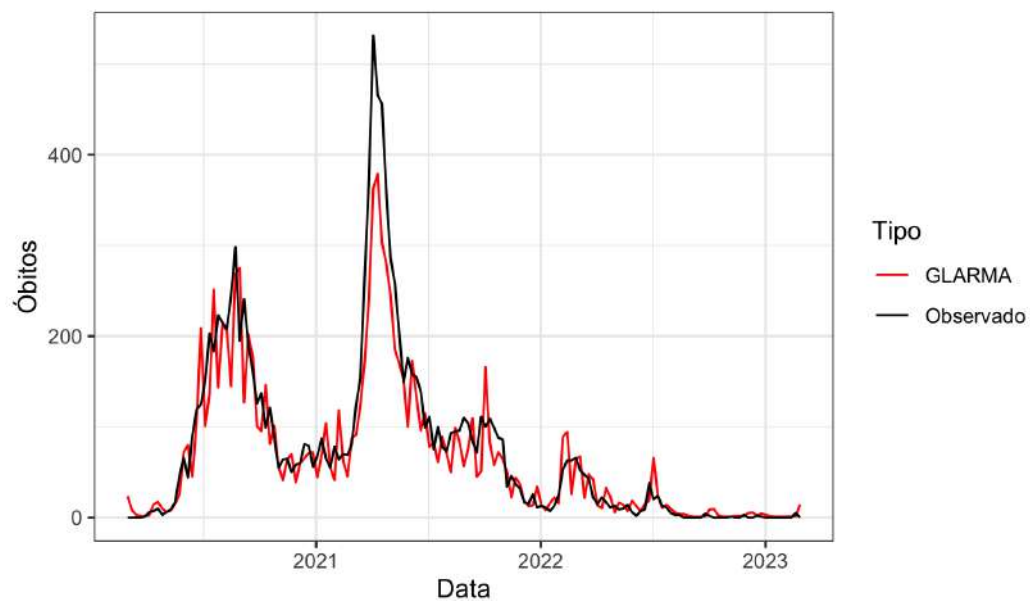


Figura 8: Número de óbitos semanais ajustado pelo modelo GLARMA(1,2) Binomial Negativa considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

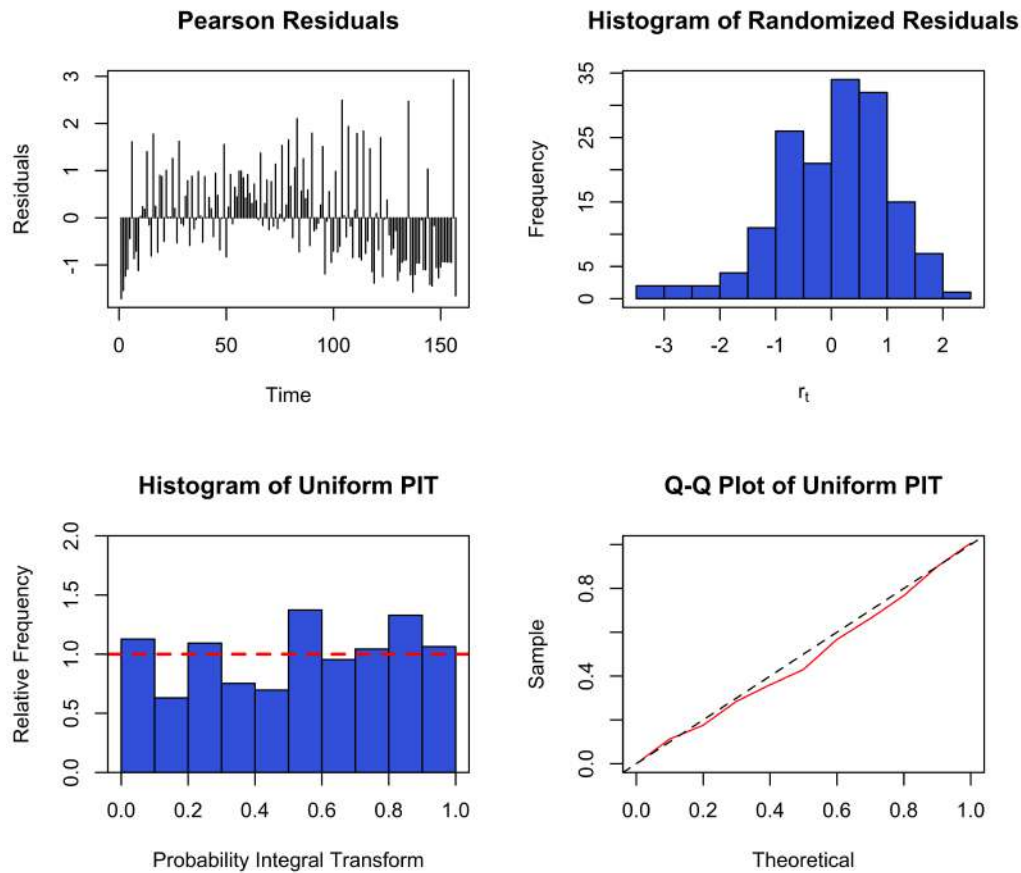


Figura 9: Diagnóstico do modelo GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

Com o novo modelo, facilmente percebe-se pela Figura 9 que a distribuição Binomial Negativa é consideravelmente mais adequada, já que no gráfico de *Histogram of Uniform PIT* a distribuição se assemelha à uma Uniforme e no gráfico *Q-Q Plot of Uniform PIT* os resíduos se comportam de forma adequada em relação aos quantis teóricos. Com isso, é possível gerar um ajuste ainda melhor, observado na Figura 8.

Seguindo com o diagnóstico do modelo, foi utilizada a função *summary* do R para verificar as estimativas do modelo linear das variáveis, exibidos na Tabela 2.

Tabela 2: Coeficientes lineares, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana, taxa de vacinação na última semana e tempo de onda.

Variável	Estimativa	Erro padrão	z	p-valor
Intercepto	3,379	0,222	15,24	<0,0001
Casos na última semana	0,00008	<0,0001	4,05	<0,0001
Taxa de vacinação na última semana	-1,296	0,395	-3,28	0,001
Tempo de onda	0,0009	0,0006	1,59	0,113

Na Tabela 2, percebe-se que a variável tempo de onda apresentara p-valor alto, o que indica que não há evidências de significância. Diante disso, na tentativa de melhorar o modelo, o ajuste foi refeito sem a variável tempo de onda.

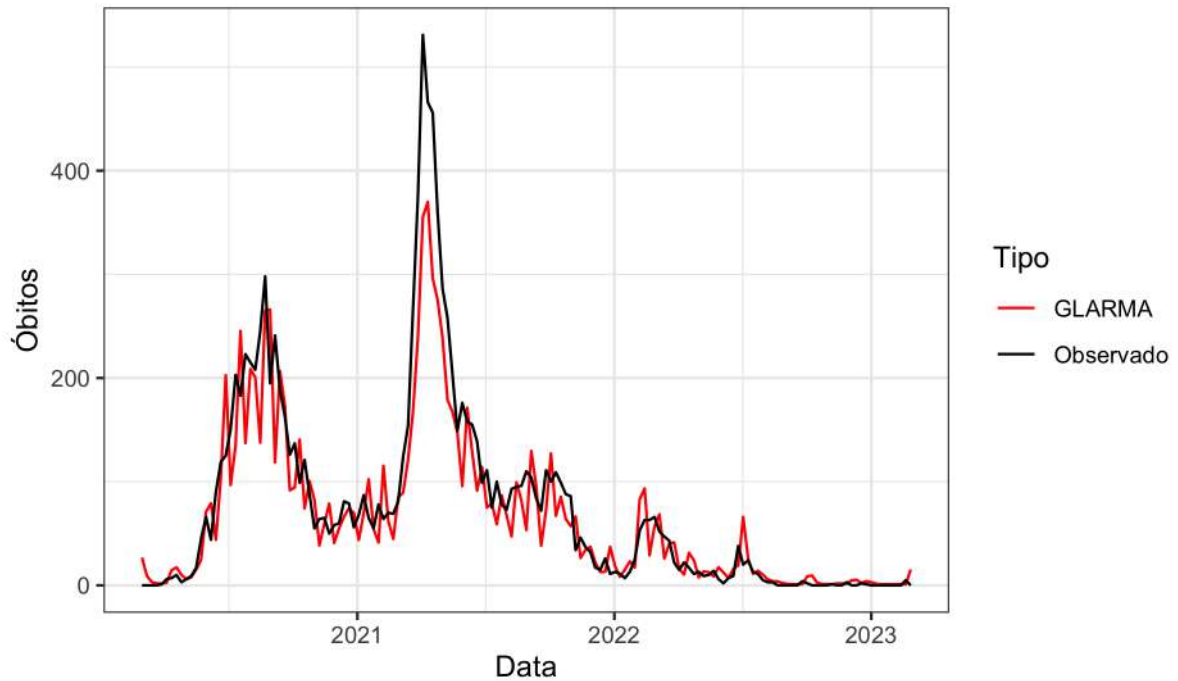


Figura 10: Número de óbitos semanais ajustado pelo modelo GLARMA(1,2) Binomial Negativa considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.

Tabela 3: Coeficientes lineares, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.

Variável	Estimativa	Erro padrão	z	p-valor
Intercepto	3,502	0,212	16,510	<0,0001
Casos na última semana	0,00007	<0,0001	3,498	0,0004
Taxa de vacinação na última semana	-1,064	0,412	-2,583	0,0097

Na Tabela 3 observa-se que sem a covariável tempo de onda, as outras covariáveis permanecem significativas e com p-valores próximos a zero, confirmando a significância das mesmas .

Tabela 4: Coeficiente Binomial Negativa, GLARMA(1,2) Binomial Negativa ajustado considerando as covariáveis número de casos na última semana e taxa de vacinação na última semana.

Parâmetro	Estimativa	Erro padrão	z	p-valor
α	2,70	0,33	8,20	<0,0001

Ademais, na Tabela 4 constata-se que o parâmetro α da distribuição Binomial Negativa é significativa, confirmando que a escolha da distribuição é adequada.

Tabela 5: Resultado dos testes LR e Wald de verossimilhança.

Teste	Estatística	P-valor
LR Test	216,0	<0,0001
Wald Test	369,9	<0,0001

Por fim, ao ajustar o modelo com a função *glarma*, obtem-se também o resultado dos testes de razão de verossimilhança e teste de Wald. O resultado exibido na Tabela 5 mostra que para ambos a hipótese nula é rejeitada, ou seja, não há evidências de que o modelo é um MLG com a mesma estrutura de regressão.

3.3 Previsão GLARMA

Para a previsão com o modelo GLARMA, os dados serão separados em treino e teste, onde os dados de treino serão usados para ajustar o modelo até o tempo t e fazer a previsão de um passo (uma semana), até $t + 1$. Além disso, para a seleção dos sucessivos modelos, foram feitos novamente testes com combinações de parâmetros, sempre com $p, q \leq 5$, com o intuito de escolher o modelo de menor AIC em casa passo.

Inicialmente, foi escolhido como período de treino inicial os dados até 26/12/2021, que é a data final da onda da variante Beta. Dessa forma, o modelo iniciará a previsão para os óbitos de uma nova onda de Covid-19. Os resultados estão exibidos abaixo.

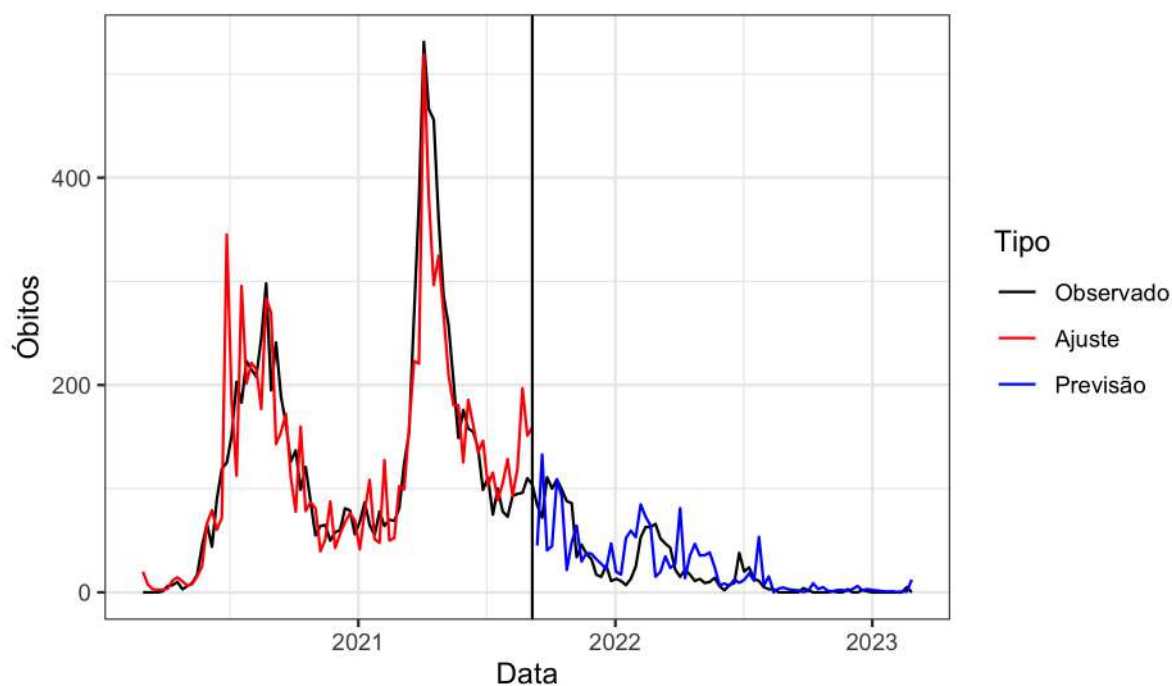


Figura 11: Previsão do número de óbitos semanais a partir de 23/12/2021, modelo GLARMA Binomial Negativa.

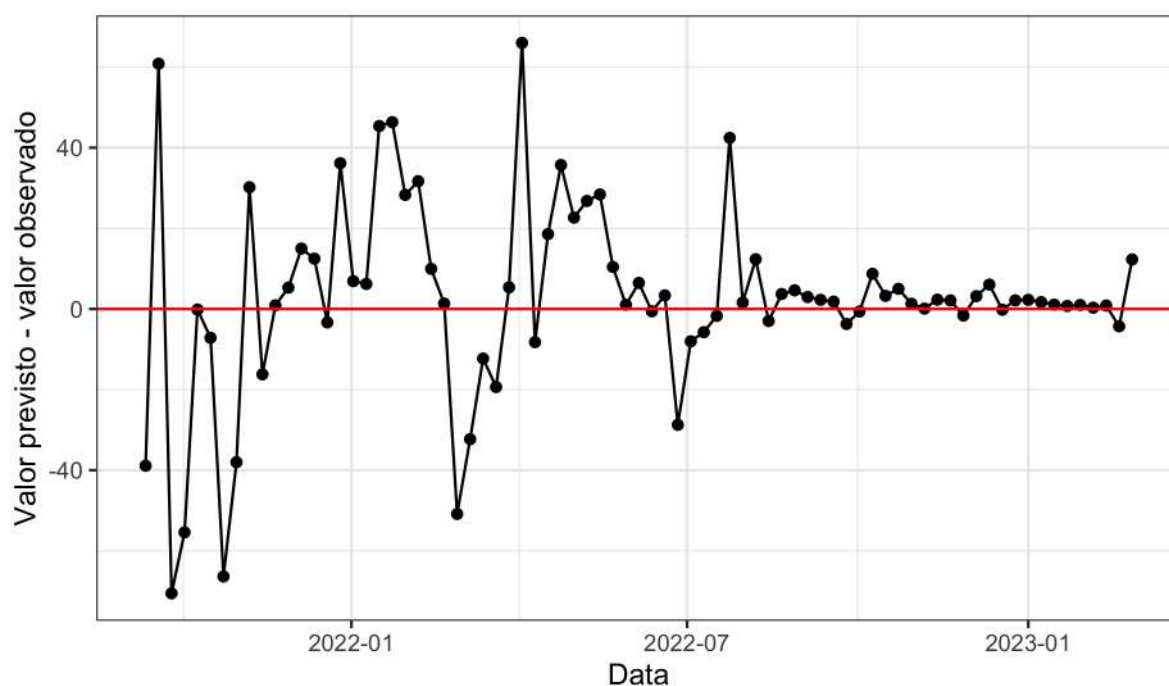


Figura 12: Diferença entre valores previstos e observados a partir de 23/12/2021, modelo GLARMA Binomial Negativa.

Na Figura 11, nota-se que o ajuste inicial para o conjunto de treino tem boa aderência aos valores observados. Em relação aos valores previstos, observa-se grande

oscilação e, em alguns pontos, grande diferença em relação aos valores observados.

Em sequência, na Figura 12, na qual está representada a diferença entre o valor previsto e o observado, percebe-se que os valores começam a diminuir significativamente, em módulo, a partir de julho de 2022. Entretanto, ressalta-se que essa menor diferença entre os valores previstos e observados não é necessariamente significativa, visto que os valores esperados são, também, menores.

3.4 Comparação com modelo ARIMA

Apesar do modelo GLARMA ser mais adequado para a natureza dos dados em estudo, surge a curiosidade de comparar seu desempenho de predição com modelos mais tradicionais.

Apesar de não ser teoricamente adequado para modelagem de dados de contagem, o modelo ARIMA costuma apresentar bom desempenho, mesmo quando todos seus pressupostos não são atendidos, como mostrado em Kane et al. (2014), Petukhova et al. (2018) e diversos outros estudos.

Dito isso, foram realizadas previsões no mesmo intervalo usado para o modelo GLARMA. As previsões foram feitas ajustando o modelo até o tempo t , com a função *auto.arima* do R, com as mesmas covariáveis utilizadas no modelo GLARMA, e prevendo um passo à frente em $t + 1$. Os resultados são apresentados abaixo.

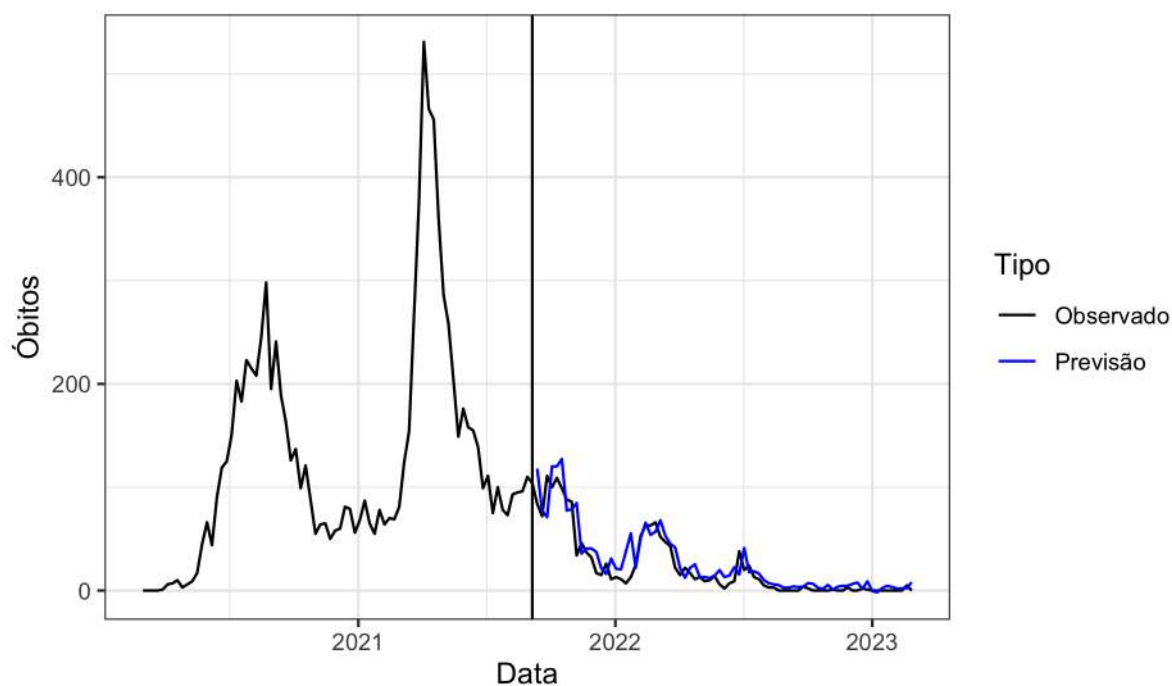


Figura 13: Previsão do número de óbitos semanais a partir de 23/12/2021, modelo ARIMA com covariáveis.

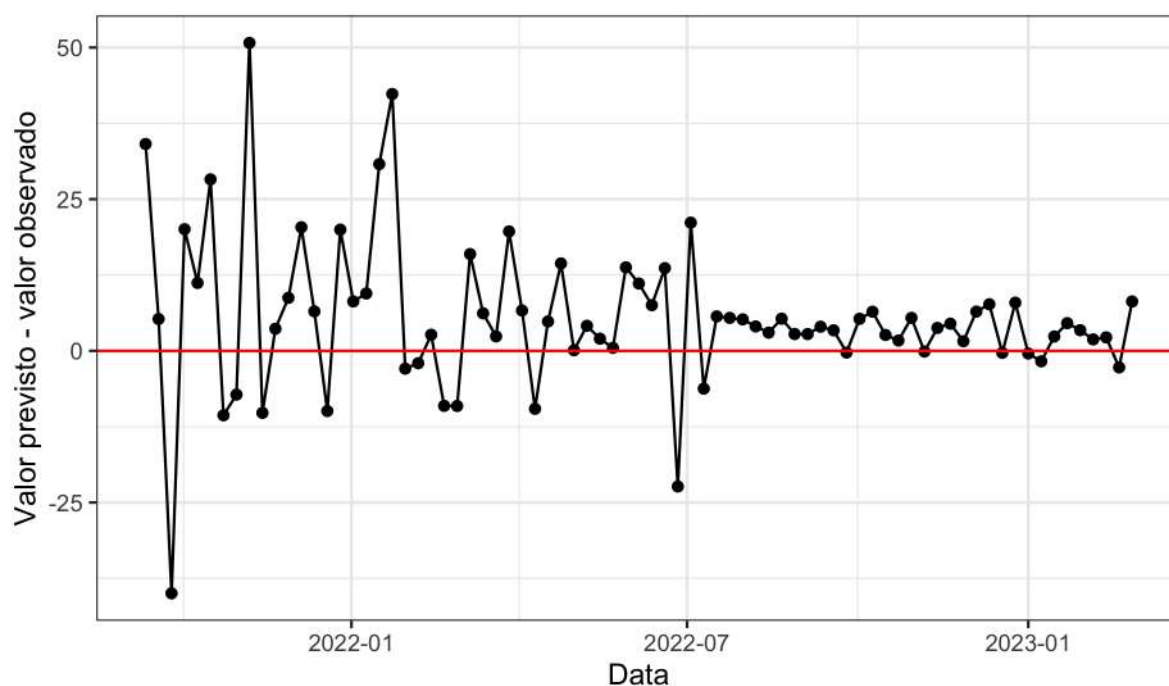


Figura 14: Diferença entre valores previstos e observados a partir de 23/12/2021, modelo ARIMA com covariáveis.

Observando a Figura 13 é possível perceber que a previsão do modelo ARIMA é bem mais suave e não apresenta oscilações muito grandes em torno dos valores observados.

Outrossim, na Figura 14, nota-se que os erros variam em torno de zero e, assim como para o modelo GLARMA, diminuem consideravelmente a partir de julho de 2022, período no qual houve grande queda no número de óbitos.

Por fim, para melhor comparar os resultados obtidos com os dois modelos, os gráficos de previsão e de diferença entre os valores previstos e observados foram refeitos considerando os dois modelos. Estão representados nas Figuras 15 e 16 abaixo.

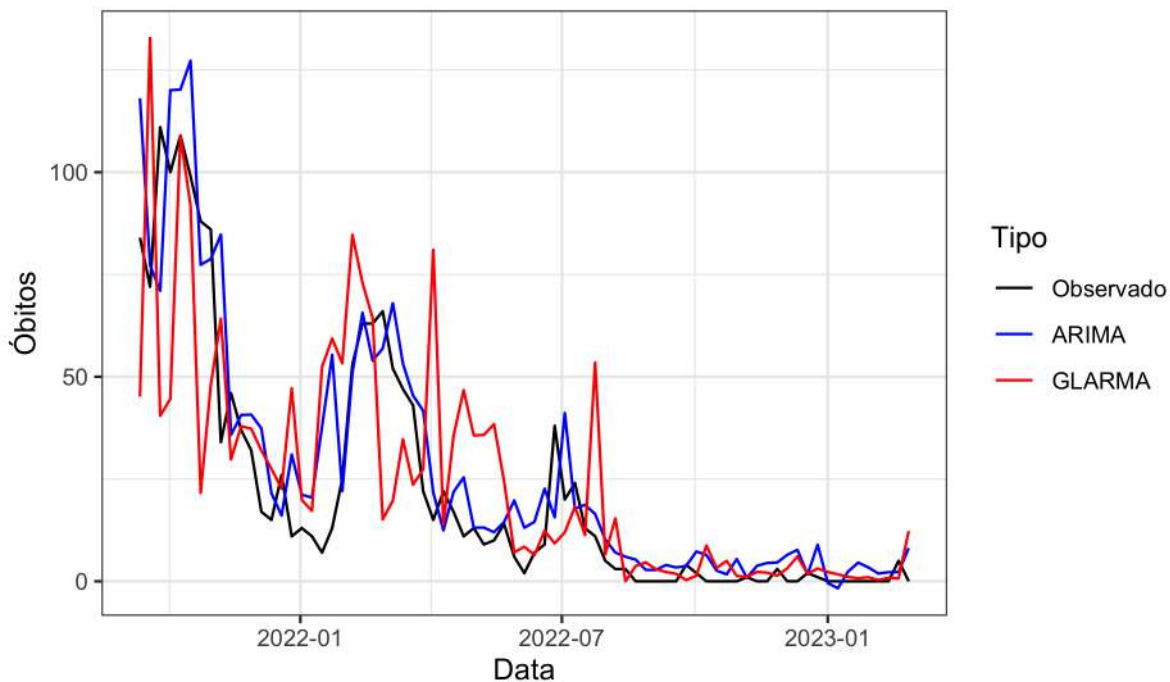


Figura 15: Previsão do número de óbitos semanais a partir de 23/12/2021, ARIMA com covariáveis e GLARMA Binomial Negativa.

Como mencionado anteriormente, as previsões do modelo ARIMA com covariáveis são mais suaves e não apresentam picos ou oscilações bruscas, como no modelo GLARMA.

Entretanto, na Figura 15, percebe-se, com certa dificuldade, que para as contagens de óbitos próximas a zero, o modelo ARIMA prevê, em alguns pontos, valores negativos.

Esse problema não ocorre no modelo GLARMA devido à natureza da distribuição Binomial Negativa, mostrando a relevância da escolha de uma modelo com distribuição adequada aos dados em estudo.

Para melhor visualização das observações com baixa contagem, apresenta-se na Tabela 6 os valores observados e preditos para as últimas 10 observações.

Tabela 6: Valores previstos, GLARMA Binomial Negativa e ARIMA com covariáveis.

Semana	Observado	GLARMA	ARIMA
25/12/2022	1	3,13	8,94
01/01/2023	0	2,26	-0,44
08/01/2023	0	1,70	-1,73
15/01/2023	0	1,07	2,36
22/01/2023	0	0,74	4,56
29/01/2023	0	0,98	3,41
05/02/2023	0	0,32	1,89
12/02/2023	0	0,84	2,22
19/02/2023	5	0,69	2,27
26/02/2023	0	12,29	8,14

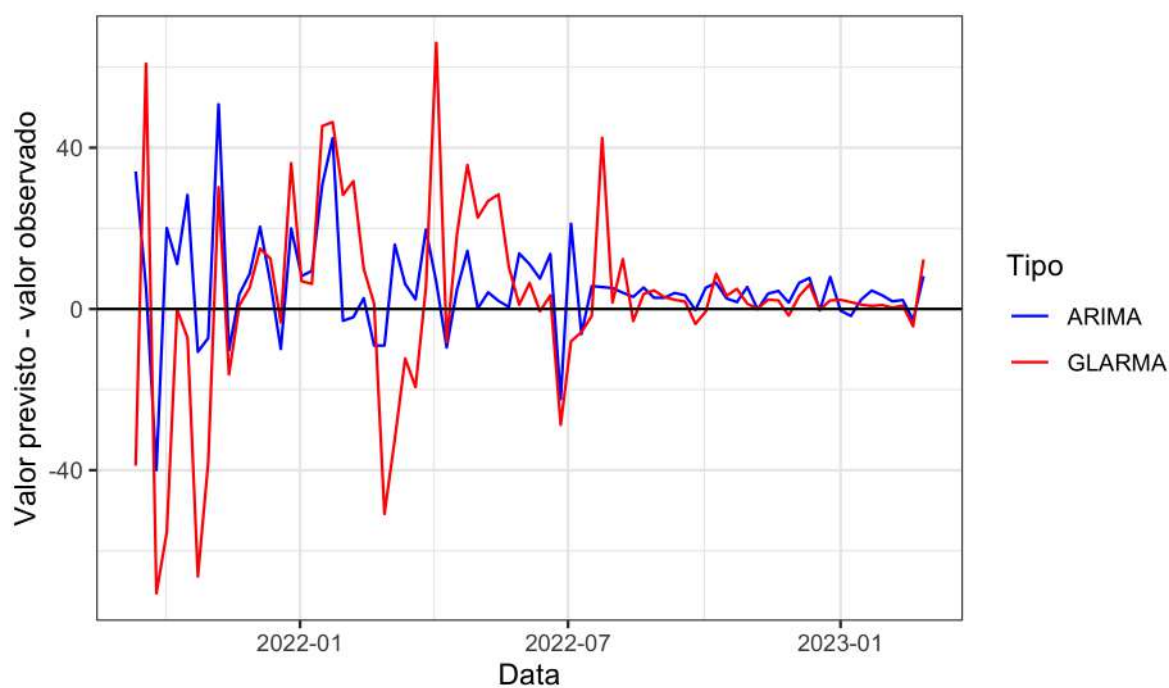


Figura 16: Diferença entre valores previstos e observados a partir de 23/12/2021, ARIMA com covariáveis e GLARMA Binomial Negativa.

Tabela 7: Erro absoluto médio.

Modelo	MAE
GLARMA	15,133
ARIMA	9,202

Finalmente, comparando a diferença entre os valores previstos e observados para

os dois modelos, na Figura 16, percebe-se que a maior discrepância está nas observações anteriores à julho de 2022, onde o modelo GLARMA chega a apresentar erros superiores a 60 em módulo, enquanto o modelo ARIMA chega no máximo a 50,8 em módulo.

Para valores mais baixos, como os das últimas semanas observadas, os modelos aparentam se equiparar em relação aos resíduos, onde o modelo GLARMA leva vantagem, por não prever valores de contagem inferiores a 0.

Por fim, comparando o erro absoluto médio dos dois modelos para o período em questão, constata-se na Tabela 7 que o valor para o modelo GLARMA é aproximadamente 1,6 vezes o valor do modelo ARIMA.

4 Conclusão

Em suma, este trabalho teve como objetivo ajustar um modelo GLARMA para prever o número de óbitos por Covid-19 por semana no Distrito Federal e comparar seus resultados com os de um modelo ARIMA.

Durante o ajuste do modelo GLARMA, foi identificado que a distribuição mais adequada era a Binomial Negativa, e as variáveis significativas incluíam o número de casos na semana anterior e a porcentagem de vacinação na última semana.

Além disso, ao analisar as previsões realizadas pelo modelo GLARMA, observou-se que elas apresentam bruscas oscilações entre alguns pontos, o que ocasiona em erros maiores. Por outro lado, ao ajustar o modelo ARIMA com as mesmas covariáveis e aplicar o mesmo método de previsão de um passo de forma sucessiva, verificou-se que as previsões foram mais coerentes com os dados observados.

Em conclusão, ao comparar o desempenho dos dois modelos, fica evidente que o ajuste utilizando o modelo ARIMA apresentou erros significativamente menores ao longo do período analisado. No entanto, nas últimas semanas de previsão, o modelo GLARMA se destacou, pois obteve erros semelhantes aos do modelo ARIMA, com a vantagem de não prever valores negativos para os dados, que representam uma contagem.

Esses resultados indicam que, embora o modelo ARIMA tenha sido mais preciso em alguns casos, o modelo GLARMA pode ser mais apropriado para prever as últimas semanas de óbitos por Covid-19, pois o mesmo leva em consideração a natureza dos dados e, assim, evita previsões negativas em dados de contagem.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19(6), p. 716–723, 1974.
- BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical association*, Taylor & Francis, Ltd., v. 98(461), p. 214–223, 2003.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. [S.l.]: San Francisco, 1970.
- DAVIS, R. A.; DUNSMUIR, W. T.; STREETT, S. B. Observation driven models for poisson counts. *Biometrika*, Oxford University Press, v. 90(4), p. 777–790, 2003.
- DUNSMUIR, W. T. M.; SCOTT, D. J. The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software*, v. 67(7), p. 1–36, 2015. Disponível em: <https://www.jstatsoft.org/article/view/v067i07>.
- EHLERS, R. S. *Análise de séries temporais*. 2009.
- KANE, M. J. et al. Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. *BMC Bioinformatics*, v. 15(1), p. 276, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4152592/#CR20>.
- Ministério da Saúde. *Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil*. 2020. Disponível em: <https://covid.saude.gov.br/>.
- MORETTIN, P. A. *Análise de Séries Temporais*. [S.l.]: Editora Blucher, 2018.
- PETUKHOVA, T. et al. Assessment of autoregressive integrated moving average (arima), generalized linear autoregressive moving average (glarma), and random forest (rf) time series regression models for predicting influenza a virus frequency in swine in ontario, canada. *PLoS ONE*, v. 13(6), p. 1–17, 2018. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0198313>.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>.
- Secretaria de Saúde do Distrito Federal. *Vacinômetro - Covid-19*. 2021. Disponível em: <https://www.saude.df.gov.br/vacinometro>.
- SOUSA, A. R. dos S. et al. *Análise de Séries Temporais*. [S.l.]: SAGAH EDUCAÇÃO S.A, 2021.