

Priscila Rebeca Ferreira de Souza dos Santos

**O impacto da pandemia da COVID-19 nos
nascimentos no estado do Rio de Janeiro**

Niterói - RJ, Brasil

20 de dezembro de 2023

Priscila Rebeca Ferreira de Souza dos Santos

**O impacto da pandemia da COVID-19 nos nascimentos no
estado do Rio de Janeiro**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientadora: Prof. Dra. Márcia Marques de Carvalho

Niterói - RJ, Brasil

20 de dezembro de 2023

Priscila Rebeca Ferreira de Souza dos Santos

**O impacto da pandemia da COVID-19 nos nascimentos
no estado do Rio de Janeiro**

Monografia de Projeto Final de Graduação sob o título “*O impacto da pandemia da COVID-19 nos nascimentos no estado do Rio de Janeiro*”, defendida por Priscila Rebeca Ferreira de Souza dos Santos e aprovada em 20 de dezembro de 2023, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Márcia Marques de Carvalho
Departamento de Estatística – UFF

Prof. Dra. Ludmilla da Silva Viana Jacobson
Departamento de Estatística – UFF

Prof. Dr. Victor Eduardo Leite de Almeida Duca
Departamento de Estatística – UFF

Niterói, 20 de dezembro de 2023

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

S237i Santos, Priscila Rebeca Ferreira de Souza dos
O impacto da pandemia da COVID-19 nos nascimentos no estado do Rio de Janeiro / Priscila Rebeca Ferreira de Souza dos Santos. - 2023.
52 f.: il.

Orientador: Márcia Marques de Carvalho.
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2023.

1. Natalidade. 2. COVID-19. 3. SARIMA. 4. Previsão. 5. Produção intelectual. I. Carvalho, Márcia Marques de, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

Resumo

A Pandemia da COVID-19 transformou a realidade humana em vários aspectos não só sanitários quanto sociais e econômicos; parte dessas transformações tendem a ser passageiras, outras permanentes. Atualmente as pesquisas se concentram nos efeitos colaterais que o vírus deixou na sociedade, visando compreender as mudanças no padrão de vida conhecido. Esse trabalho tem como objetivo analisar os efeitos da Pandemia no número de nascimentos no Rio de Janeiro para entender quais são os efeitos dela no desenvolvimento no estado, uma vez que a baixa taxa de natalidade influencia não só em aspectos populacionais, mas nos econômicos e de saúde, como já se vê em países mais antigos da Europa. Para alcançar esses objetivos, foram utilizados dados do IBGE e da Secretaria de Saúde do Estado para a previsão do comportamento da taxa de natalidade caso não houvesse o fator Pandemia através de modelos SARIMA e então a comparação dessa previsão com os números reais. De 2020 a 2022, 26.606 crianças deixaram de nascer no Estado do Rio de Janeiro, o que representa 4,7% da previsão do modelo estimado neste trabalho. Houve também a indicação da relação entre a idade da mãe e a diminuição do percentual de nascimento de 2017 a 2022.

Palavras-chave: Natalidade, COVID-19, SARIMA, previsão.

Dedicatória

Para você.

Agradecimentos

Agradeço a Deus, por ter sido ímpar (e par) para me dar a calma necessária, o tempo e todo o resto.

Agradeço aos meus pais pelo apoio, incentivo e amor que me fizeram continuar aqui. À minha família de sangue ou não, com destaque especial aos meus avós, a Ester e a Dayse por me ouvirem falar e me apoiarem em todos os sentidos possíveis. Agradeço a todos os meus amigos que estiveram dizendo que estava tudo bem quando eu achava que não. Agradeço a Wila pelo início do percurso, à Hillary, Pedro, Caio, Luiza e Beatriz pelo meio dele e ao Luiz pelo fim. A todos os professores e cooperadores que fazem da Universidade Federal Fluminense o que ela é, em especial à Sandra Brignol por me ensinar sobre humanidade e à Márcia Carvalho por acreditar em mim mesmo sendo humana.

Obrigada a todos que de alguma forma foram decisivos para a conclusão deste percurso e para quem um dia eu disse “seu nome um dia vai estar no meu TCC”.

Sumário

1	Introdução	p. 12
1.1	Motivação	p. 12
1.2	Objetivos	p. 14
2	Materiais e Métodos	p. 16
2.1	Base de Dados.	p. 16
2.2	Indicadores Demográficos	p. 17
2.2.1	Taxa Bruta de Natalidade	p. 18
2.2.2	Taxa de Fecundidade Geral.	p. 18
2.2.3	Taxa Específica de Fecundidade.	p. 18
2.2.3	Taxa de Fecundidade Total	p. 19
2.3	Modelo SARIMA.	p. 20
2.3.1	Séries Temporais.	p. 20
2.3.2	Tendência.	p. 21
2.3.3	Sazonalidade.	p. 21
2.3.4	Ruído Branco	p. 22
2.3.5	Estacionariedade.	p. 22
2.3.5.1	Teste da Raíz Unitária	p. 23
2.3.7	Metodologia Box & Jenkins.	p. 24
2.3.6	Modelos SARIMA.	p. 24
2.3.7.1	Identificação	p. 25
2.3.7.1.1	Função de Autocorrelação.	p. 26
2.3.7.1.2	Função de Autocorrelação Parcial	p. 27

2.3.7.2	Estimação	p. 27
2.3.7.2.1	Teste de Shapiro Wilk	p. 28
2.3.7.3	Verificação	p. 28
2.3.7.4	Previsão	p. 29
2.3.8	Medidas de Seleção do Modelo.	p. 30
2.3.8.1	Critério de Informação de Akaike	p. 30
2.3.8.2	Raiz Quadrada do Erro Quadrático Médio.	p. 31
2.3.8.3	Erro Médio Percentual Absoluto	p. 32
3	Resultados	p. 33
3.1	Análise Descritiva	p. 33
3.2	Indicadores Demográficos	p. 35
3.3	Série Temporal	p. 37
3.3.1	Análises Preliminares	p. 37
3.3.2	Identificação	p. 39
3.3.3	Estimação e Análise de Resíduos.	p. 44
3.3.4	Previsão	p. 46
4	Conclusão	p. 49
5	Referências	p. 51

Lista de Figuras

Figura 1: Nascidos vivos no estado do Rio de Janeiro pelo local de residência da mãe	p. 33
Figura 2: Número de nascimentos por faixa de idade da mãe	p. 34
Figura 3: Proporção de filhos por mulheres segundo o nível de instrução entre 2017 e 2022.	p. 35
Figura 4: Taxa Específica de Fecundidade das mulheres do estado do Rio de Janeiro entre 2017 e 2022.	p. 36
Figura 5: Nascimentos no estado do Rio de Janeiro de 2010 a 2022	p. 37
Figura 6: Gráfico polar dos nascimentos no estado de 2010 a 2019.	p. 38
Figura 7: Decomposição da série dos nascimentos no estado de 2010 a 2019.	p. 39
Figura 8: Gráfico de dispersão da FAC e suas modificações	p. 41
Figura 9: Gráfico da série derivada e da sua FAC e FACP.	p. 42
Figura 10: Características dos resíduos do modelo $SARIMA(2, 1, 0)(3, 1, 2)_{12}$	p. 45
Figura 11: Gráfico da série com a previsão de 36 meses à frente.	p. 46
Figura 12: Gráfico do número de nascimentos previstos x número de nascimentos reais durante a pandemia da COVID-19 no estado do Rio de Janeiro	p. 47

Lista de Tabelas

Tabela 1: Características da FAC e FACP em diferentes modelos teóricos.	p. 26
Tabela 2: Relação entre o número de nascidos vivos e o nível de escolaridade da mãe. .	p. 34
Tabela 3: Taxa Bruta de Natalidade e Taxa de Fecundidade Geral de 2017 a 2022 para cada mil habitantes do Rio de Janeiro	p. 36
Tabela 4: Taxa de Fecundidade Total de 2017 a 2022 do Rio de Janeiro	p. 37
Tabela 5: Possíveis valores para as ordens p , q , P e Q do modelo SARIMA.	p. 43
Tabela 6: AIC, MAPE e RMSE dos modelos testados.	p. 44
Tabela 7: Número real de nascimentos de 2020 a 2022 e a previsão do modelo	p. 48

1. Introdução

1.1. Motivação

A família da Síndrome Respiratória Aguda Grave causada pelo novo Coronavírus (SARS-CoV-2), popularizada em decorrência da pandemia da COVID-19, foi descoberta em 1937 e nomeada “coronavírus” por seu formato lembrar o de uma coroa (Lima, 2020). A primeira detecção da nova variável da doença foi feita na cidade de Wuhan, na China, em dezembro de 2019. Desde então, segundo dados da Organização Mundial de Saúde (OMS), foram mais de 766 milhões de casos confirmados e aproximadamente 6,9 milhões de óbitos no mundo.

No dia 11 de março de 2020 a OMS decretou que a, até então, Emergência de Saúde Pública de Importância Internacional (ESPII) causada pela nova variação do coronavírus seria, dali em diante, caracterizada como uma pandemia pela distribuição geográfica da mesma. A partir desse momento, as sequelas acumuladas em decorrência do vírus no cotidiano não se concentraram apenas no âmbito sanitário. A sociedade, pela instabilidade sanitária, adotou novas características e tomou nova forma econômica, social e demográfica pelas consequências das medidas adotadas para o combate da doença; parte dessas características tornaram-se permanentes e esse estudo pretende verificar se a taxa de nascimentos está nessa categoria.

De acordo com a teoria da transição demográfica, o processo de modernização das sociedades e o desenvolvimento econômico tem relação direta com as taxas de natalidade e de mortalidade. Fato é que em meados da década de 60 aconteceu no Brasil a leve, porém notável e irrevogável queda nos níveis das taxas de natalidade e fecundidade, marcando assim o início do processo da segunda fase da transição demográfica no país (Vasconcelos et al., 2012). Desde aquela época a taxa bruta de natalidade registrou queda de aproximadamente 36.3%, sendo de 44.0 no ano de 1970 e passando para 16.0 em 2010 (Vasconcelos et al., 2012). De forma análoga, a Taxa de Fecundidade Total (TFT) saltou de 5.8 para 1,9 no mesmo período. Deve-se ressaltar que este último dado está abaixo do nível suficiente de asseguramento da reposição populacional de 2,1 (Ramos, 1987).

As consequências da baixa taxa de natalidade já são vistas em diversos países Europeus, onde a proporção média de idosos projetada para 2025 é de 25% (Ramos, 1987). A falta da mão de obra em idade ativa faz com que os países desenvolvidos tenham sistemas provisórios, tais como o incentivo à imigração, para a manutenção dos sistemas que sustentam seus habitantes.

Pela urgência apresentada na possível futura falta da população economicamente ativa que geraria inconsistência no sistema previdenciário, de saúde, entre outros, torna-se pertinente o interesse em averiguar quais foram os impactos da pandemia da COVID-19 na quantidade de crianças que nasceram no estado do Rio de Janeiro entre 2020 e 2022, uma vez que a taxa de nascimentos já apresentava queda antes da pandemia.

Foi realizada a revisão bibliográfica de estudo que abordaram o tema e o artigo desenvolvido no Brasil apontou queda no total de nascidos vivos no país desde 1950 (DE SOUZA et al. 2021). Além disso, no trabalho de Coutinho et al. (2020) foi observado que a idade média da fecundidade está subindo no país, embora não como em outros países com baixa taxa de fecundidade. Embora isso, segundo o autor, é razoável considerar um aumento nessa média durante o cenário da pandemia.

Já o estudo realizado em países da Europa (LUPPI et al. 2020), aponta que, dentre as pessoas que já tinham o planejamento de terem filhos, cerca de 26% desistiram de seu planejamento. O estudo aponta planos de fecundidade retificados ou abandonados em todos os países em diferentes proporções, sendo destaque a Itália, onde a proporção das mulheres que desistiram da maternidade é maior do que a de todos os outros países estudados, seguido pela Espanha, enquanto a proporção das italianas que adiaram os planos é a menor do conjunto. Ao todo, 74,5% das italianas entrevistadas alegaram que não tinham planos de engravidar no período do estudo. Ainda de acordo com Luppi et al. (2020), a desistência da maternidade ocorreu em diferentes grupos de diferentes maneiras: De acordo com o trabalho, a idade (menos de 30 anos), o nível de instrução, a estabilidade econômica e as características demográficas têm relação com o planejamento de conceber. Dentre as italianas, a idade das mulheres e o nível de instrução obtiveram associação com a reconsideração dos planos de engravidar. No Reino Unido, a associação se deu com o medo na diminuição de seus rendimentos futuros e, na Alemanha, com o quão afetada a região habitada pela mulher foi pela COVID-19.

De acordo com Coutinho et al. (2020), dentre as principais variáveis para a diminuição da taxa de fecundidade no cenário pandêmico está o aumento da incerteza com relação ao futuro, agravado pelo cenário econômico que o país vivia antes da pandemia; a rotina familiar causada pelo confinamento; o medo com relação à infecção pelo novo coronavírus na gestação; e a diminuição dos encontros ocasionais que geram gravidezes. Segundo os autores, tais receios podem ser propagados no cenário pós pandêmico e ocasionar a redução permanente da taxa de fecundidade pela nova forma dos progenitores planejarem o conceber levando em conta possíveis crises sanitárias futuras.

Utilizando um olhar retrospectivo, segundo Chandra et al. (2018), a influenza A H1N1 (Gripe Espanhola), uma das últimas pandemias antes da COVID-19, causou, nos estados Unidos, uma queda acentuada no número de nascimentos entre 9 e 10 meses após o pico de mortalidade. Vale ressaltar que no ano seguinte aconteceu um baby boom no país e os autores se dividem quanto à causa do mesmo; enquanto alguns apontam o fim da pandemia como o principal motivo, outros atribuem o fenômeno ao fim da Primeira Guerra Mundial.

1.2 Objetivos

A seguir, serão apresentados os objetivos gerais e específicos do presente trabalho.

Como objetivo geral, pretende-se analisar a fecundidade e os nascimentos utilizando indicadores demográficos, antes e durante a pandemia da Covid-19 no Estado do Rio de Janeiro. Através deste, foram desenvolvidos os seguintes objetivos específicos:

1. Calcular e analisar os indicadores demográficos relativos à fecundidade, por ano, por faixa etária e nível de escolaridade da mãe, para os anos de 2017, 2018, 2019, 2020, 2021 e 2022;
2. Modelar os nascimentos antes da pandemia utilizando modelos SARIMA, nos dados mensais;
3. Obter previsão pontual e intervalar dos nascimentos durante a pandemia;

4. Comparar as previsões com as observadas para calcular o número de crianças que deixaram de nascer durante a pandemia.

Para atingir esses objetivos, será construída uma base de dados com os dados dos sites do DataSUS, da Secretaria do Estado do Rio de Janeiro e do IBGE, e um modelo será estimado no software R Studio.

Este estudo se apresentará em quatro capítulos. No primeiro capítulo o tema será apresentado, tal como uma revisão bibliográfica a apresentação dos objetivos. No Capítulo 2 serão apresentados os métodos analíticos utilizados bem como a natureza dos testes. No Capítulo 3 serão analisados os resultados obtidos nas análises. Finalmente, no Capítulo 4 serão descritas as conclusões obtidas nesta pesquisa.

2. Materiais e Métodos

Neste capítulo será apresentado o processo de coleta e montagem das bases de dados a ser utilizada além da metodologia estatística empregada.

2.1 Bases de Dados

Os dados utilizados neste trabalho serão extraídos do DataSUS, do Site da Secretaria do Estado do Rio de Janeiro e do Site do IBGE.

A PNAD Contínua foi criada em outubro de 2011, com caráter experimental, e, a partir de janeiro de 2012, oficialmente definida em todo o Brasil. É responsável por fornecer estimativas trimestrais sobre o mercado de trabalho tais como características demográficas e de educação e outras informações necessárias para o estudo e desenvolvimento socioeconômico do Brasil. Serão utilizados os dados da PNAD Contínua do ano de 2018.

O Sistema de Informações sobre Nascidos Vivos (SINASC), foi criado em 1990, e tem como objetivo a coleta de dados dos nascidos no Brasil e o fornecimento dos dados de natalidade ao Sistema de Saúde.

O Censo Demográfico desenvolvido pelo IBGE é, na atualidade, a pesquisa mais competente e completa para apresentação das características das condições de vida da população brasileira em âmbito municipal e nos recortes territoriais internos do país. A base de dados é coletada de todo domicílio do Território Nacional a cada 10 anos. O Censo utilizado será relativo ao ano de 2010.

Através do Site da Secretaria do Estado do Rio de Janeiro serão alcançadas informações relativas à pandemia da COVID-19 no Rio de Janeiro.

Abaixo está a relação de variáveis utilizadas, suas características e os sites em que elas foram coletadas.

- **População:** Referindo-se aos anos de 2017 a 2022, serão coletados dados do Censo Demográfico e da PNAD Contínua a respeito das características demográficas da pesquisa;
- **Natalidade:** Coletados no período de 2012 a 2022, os dados serão obtidos através do SINASC e farão referência ao número de nascidos vivos no estado levando em conta o local de residência da mãe;
- **Fecundidade:** Diferente da natalidade, que analisa os nascimentos na população como um todo, a fecundidade faz referência à ocorrência de crianças nascidas vivas por mulheres em idade reprodutiva. A coleta dos dados foi feita no período de 2017 a 2022 e obtidos pelo SINASC;
- **Dados da COVID-19:** Os dados que dizem respeito à pandemia foram obtidos pelo Site da Secretaria do Estado do Rio de Janeiro e dizem respeito aos anos de 2019, 2020, 2021 e 2022.

2.2 Indicadores Demográficos

Como pontapé inicial para as discussões sobre indicadores demográficos, que serão utilizados para a análise dos nascimentos que ocorreram no estado do Rio de Janeiro de 2017 a 2022, é interessante a apresentação do conceito da demografia em si.

Segundo Foz (2021), demografia é, consensualmente, o estudo estatístico das populações humanas, com maior ênfase nos conjuntos de pessoas do que nos indivíduos que as compõem. Em suma, o estudo da demografia caracteriza-se pela análise de dados populacionais, podendo também ser definida como estudo estatístico das populações.

Para a execução deste trabalho, será mantido o foco nos indicadores de natalidade e fecundidade, onde termo “natalidade” diz respeito à quantificação do número de nascimentos da população total de uma dada região, enquanto “fecundidade” descreve o número médio de filhos de uma mulher durante seu ciclo reprodutivo. Os próximos itens têm como objetivo pontuar os instrumentos usados para medição da população utilizada para a análise.

2.2.1 Taxa Bruta de Natalidade

A Taxa Bruta de Natalidade (TBN) relaciona o número de nascidos vivos e o total da população no período de tempo t . A TBN é calculada de acordo com o que é apresentado na equação 2.1.

$$TBN_t = \frac{Nascimentos_t}{População_t} \times 1000 \quad (2.1)$$

Onde, para cada 1000 pessoas, é dado o número de nascidos vivos em determinada população.

2.2.2 Taxa de Fecundidade Geral

Como o nome sugere, a taxa de fecundidade leva em conta apenas a parte fértil da população (mulheres com idade entre 15 e 49 anos) em uma tentativa de corrigir a TBN para uma visão mais específica do cenário da demografia do local. A fórmula matemática que representa essa taxa é calculada segundo descrito na equação 2.2.

$$TFG_t = \frac{Nascimentos_t}{Pop. Fem 15-49 anos_t} \times 1000 \quad (2.2)$$

Onde, para cada mil mulheres em idade fértil, é dado o número médio de nascidos vivos.

2.2.3 Taxa Específica de Fecundidade

Criada com o intuito de corrigir as diferenças etárias das mulheres de diferentes localidades, a Taxa Específica de Fecundidade (TEF) refere-se à divisão entre o número de nascidos

vivos de mães em uma determinada classe de idades. Pode ser expressa matematicamente como apresentado na equação 2.3.

$$TEF_{[x,x+n)} = \frac{N_{[x,x+n)}}{Pop. Fem_{[x,x+n)}} \quad (2.3)$$

Onde:

A interpretação do resultado normalmente multiplica-o por mil, e assim é dito que a cada mil mulheres entre $[x, x + n)$ anos, o número de nascidos vivos é o resultante da função.

2.2.4 Taxa de Fecundidade Total

O principal indicador utilizado neste estudo será a Taxa de Fecundidade Total, que diz respeito ao número médio de filhos que uma mulher teria no fim do seu período reprodutivo. Matematicamente, esse resultado pode ser obtido através da fórmula apresentada na equação 2.4.

$$TFT = n \sum_X TEF_{[x,x+n)} \quad (2.4)$$

Onde a TEF é a Taxa Específica de Fecundidade apresentada acima e o parâmetro n é o espaçamento do intervalo de idade verificado.

2.3 Modelo SARIMA

Os nascimentos, bem como a mortalidade, são elementos demográficos que podem ser estimados a partir de modelos estatísticos baseados em tendências observadas no passado. A partir disso, a seguir serão apresentados as definições e conceitos desses modelos.

2.3.1 Séries Temporais

Uma Série Temporal é um conjunto de observações, discretas ou contínuas, analisadas no tempo em que cada observação é dependente de outras situadas em instantes próximos. O intuito principal dessas análises é modelar o comportamento dessas dependências, portanto é importante que a ordem temporal das observações seja respeitada. Podem ser discretas, quando a coleta de dados é feita de tempo em tempo (na maioria das vezes tempos equiespaçados), ou contínuas, quando a coleta é feita continuamente ao longo do tempo; univariadas, quando existe uma única variável de interesse ou multivariada, quando existe interesse em mais de uma variável.

Dos objetivos principais desse tipo de análise destacam-se a descrição dos dados, a predição através da projeção dos dados e o controle de processos.

O modelo multiplicativo é usado quando a magnitude do padrão sazonal na amostra aumenta à medida que se aumenta o tamanho da própria amostra, já o aditivo é utilizado quando não existe tal correlação. A equação $Z_t = T_t + S_t + \varepsilon_t$ é utilizada caso o modelo seja aditivo e utiliza-se $Z_t = T_t \times S_t \times \varepsilon_t$ se o modelo for multiplicativo.

Sendo:

- Z_t = série temporal;
- T_t = tendência,
- S_t = componente sazonal;
- ε_t = ruído branco, também chamado componente de aleatoriedade.

Os componentes desta fórmula serão discutidos nos próximos tópicos.

2.3.2 Tendência

A tendência é a descrição comportamental de uma série quando livre de ruídos brancos quando percebida a longo prazo. De acordo com Morettin e Tolo (1987), supondo que a sazonalidade não esteja presente, a série temporal $Z(t)$ será composta da tendência mais o ruído branco e, dentre os principais métodos de estimação estão:

- Ajustar uma função suave no tempo, tais como função exponencial e polinomial;
- Estima-la em um ponto suavizando a série ao redor desse ponto específico;
- Usar a reta de mínimos quadrados ponderados para suavizar os pontos da série.

Um outro recurso para a remoção da tendência é tomar a diferença entre uma observação da série e sua antecessora; como descrito na equação 2.5.

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.5)$$

2.3.3 Sazonalidade

Intuitivamente, a sazonalidade pode ser descrita como uma repetição de padrões em períodos idênticos de uma série temporal.

Esta componente pode ser de dois tipos: Sazonalidade Aditiva, onde a série apresenta flutuações sazonais aproximadamente constantes sem ser considerado o nível global da série, ou Sazonalidade Multiplicativa onde o nível global importa e influencia no tamanho das flutuações sazonais.

2.3.4 Ruído Branco

Um ruído branco é uma sequência de variáveis aleatórias independentes (onde as variáveis não são correlacionadas entre si) e identicamente distribuídas (i.i.d) com média 0 e variância finita. Ou seja:

- $Cov\{\varepsilon_s, \varepsilon_t\} = 0 \forall s \neq t$
- $\varepsilon_i \sim N(0, \sigma^2)$, onde ε_i é i. i. d.

2.3.5 Estacionariedade

Um processo diz-se estacionário se o mesmo se desenvolve no tempo com as características média, variância e autocorrelação inalteradas, ao ponto de que a escolha de origem do mesmo não importa. Além disso, pode ser dividida entre estacionariedade forte ou de segunda ordem.

Processos estocásticos podem ser descritos como fortes (ou estritos) se todas as distribuições têm as mesmas probabilidades. Ou seja, se a distribuição conjunta de $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$ é a mesma de $Z_{t_1-k}, Z_{t_2-k}, \dots, Z_{t_n-k}$, para todo k e para todos os tempos t_1, t_2, t_n .

O processo estocástico definido acima tem padrões difíceis de serem alcançados, portanto, para análises de séries temporais é aceitável que um processo seja estacionário de segunda ordem (ou, fracamente estacionário).

Um processo estocástico $\{Z_t, t \in T\}$ é estacionário de segunda ordem se:

- $E[Z_t] = \mu, \forall t \in T;$
- $E[Z_t^2] < \infty, \forall t \in T;$ e

- $\gamma(t_1, t_2) = Cov[Z_{t_1}, Z_{t_2}]$ é uma função de $|t_1 - t_2|$.

É importante ressaltar que processos estocásticos estritamente estacionários não podem ser considerados automaticamente incluídos no grupo de processos fracamente estacionários, pois não necessariamente satisfazem ao segundo item das condições de fraca estacionariedade.

2.3.5.1 Teste da raiz unitária

Para verificação da estacionariedade de uma série, considere a série $Z_t = \alpha + \rho Z_{t-1} + \varepsilon_t$ onde ε_t é um ruído branco.

Deseja-se testar se $\rho = |1|$ pois, em caso afirmativo, pode-se escrever Z_t como $Z_t - Z_{t-1} = \alpha + \varepsilon_t \Rightarrow \Delta Z_t = \alpha + \varepsilon_t$. Logo, Z_t não será estacionário e ΔZ_t será.

As hipóteses para esse teste são:

- $H_0: \rho = 1$ (presença de raiz unitária) $\Rightarrow Z_t$ não é estacionário
- $H_1: \rho < 1$ (ausência de raiz unitária) $\Rightarrow Z_t$ é estacionário

A expressão utilizada para esse teste será a ADF (Augmented Dickey-Fuller) que é usada com a função transformada onde $\Delta Z_t = \alpha + (\rho - 1)Z_{t-1} + \varepsilon_t = \alpha + \beta Z_{t-1} + \varepsilon_t$, e testa as hipóteses:

- $H_0: \beta = 0$, Z_t não é estacionário
- $H_1: \beta < 0$, Z_t é estacionário

Tal função pode ser escrita como está apresentado na equação 2.6.

$$\Delta Z_t = \beta_1 + \beta_2 t + \delta z_{t-1} + \sum_{i=1}^n a_i \Delta Z_{t-i} + \varepsilon_t \quad (2.6)$$

Para a qual:

- ΔZ_t : primeira derivada de Z_t ;
- β_1 : intercepto do modelo;
- β_2 : coeficiente de tendência;
- δ : é o coeficiente de presença de raiz unitária;
- n : é o número de defasagens na série.

2.3.6 Metodologia Box & Jenkins

A metodologia de Box & Jenkins proporciona o ajuste de modelos de previsão ARIMA(p,d,q), ou SARIMA(p,d,q)(P,D,Q) α , com o objetivo de que o melhor modelo seja baseada nos próprios dados de forma que a diferença dos valores observados e os ajustados das séries estudadas sejam um ruído branco. A construção desse modelo é feita a partir de um processo de três passos, identificação, estimação e verificação, que serão apresentados abaixo.

2.3.7 Modelos SARIMA

A seguir serão explicados os parâmetros e componentes dos Modelos Sazonais Auto regressivos Integrados de Médias Móveis com Sazonalidade, ou, como popularmente conhecidos, modelos SARIMA.

O SARIMA (p, d, q)(P, D, Q) α , onde α é o período sazonal (de quanto em quanto tempo o padrão sazonal se repete), é um modelo estatístico linear para análise de séries temporais com a presença da sazonalidade, anteriormente apresentada, e além dela, contém o componente do Processo Auto regressivo (AR) (que usa a informação das observações anteriores para prever o próximo valor da série), e o componente do Processo de Médias Móveis (MA) (que utiliza a média ponderada dos erros passados do modelo para prever o próximo valor da série).

Os parâmetros (p, d, q) são a parte não sazonal do modelo e os parâmetros (P, D, Q) são a parte sazonal. O componente “p” é o número de parâmetros auto regressivos, “q” o número de parâmetros de médias móveis, e “d” o número de diferenciações para que a série seja estacionária.

Sua representação é definida por:

$$\Delta_S^D \Delta^d Z_t = c + \varphi_1 \Delta^d Z_{t-1} + \dots + \varphi_p \Delta^d Z_{t-p} + \phi_1 \Delta_S^D Z_{t-1} + \phi_P \Delta_S^D Z_{t-P} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_Q \varepsilon_{t-Q}$$

2.3.7.1 Identificação

De acordo com Morettin e Tolo (2006), a etapa de identificação tem como objetivo principal encontrar os melhores valores p, d e q além de fazer inferências precoces dos parâmetros utilizados na estimação. Ainda em concordância com Morettin e Tolo (2006), são três os passos para a identificação; o primeiro passo é a verificação da estabilidade da variância e, conseqüentemente, da estacionariedade da série; tal verificação é feita através de análises empíricas ou através do teste da raiz unitária.

O segundo passo faz referência à diferenciação da série, onde ela é diferenciada até que se torne estacionária; a quantidade de diferenciações é estabelecida analisando os gráficos da FAC e FACP, se eles têm decaimento exponencial para zero nos primeiros instantes, indica que o número de diferenciações, d foi o suficiente; por fim, o terceiro passo consiste em identificar o processo do Modelo Autoregressivo de Média Móvel (ARMA)(p, q) obtido com o segundo passo através da análise das autocorrelações e das autocorrelações parciais estimadas, cujos comportamentos devem imitar os comportamentos das respectivas quantidades teóricas.

Para a mais fácil identificação de modelos potenciais, ajusta-se modelos de baixa ordem e utiliza-se critérios para a escolha dos mais adequados. Na Tabela 1 descreve-se o comportamento da FAC e da FACP teóricas desses modelos de baixa ordem que podem ajudar nessa identificação.

Tabela 1: Características da FAC e FACP em diferentes modelos teóricos

Modelo	Autocorrelação	Autocorrelação Parcial
AR(p)	Decai exponencialmente	Decai para zero quando $k > p$
MA(q)	Decai para zero quando $k > q$	Decai exponencialmente
ARMA(p,q)	Cortada na defasagem q	Cortada na defasagem p

2.3.7.1.1 Função de Autocorrelação

A Função de Autocorrelação (FAC) é um instrumento utilizado para averiguar os impactos das defasagens da série em seu valor. Para tanto, verifica-se a correlação entre Z_t e Z_{t-k} e o resultado é chamado de autocorrelação de k-ésima ordem. Utilizada para verificar a estacionariedade da série e para identificar os possíveis valores das ordens “q” e “Q” do modelo SARIMA, a FAC é dada pela equação 2.8.

$$\rho_k = \frac{Cov(Z_t, Z_{t-k})}{Var(Z_t)} \quad (2.8)$$

E tem por propriedade que:

- $\rho_0 = 1$;
- A correlação entre Z_t e Z_{t+k} é a mesma que entre Z_t e Z_{t-k} ;
- $-1 \leq \rho_k \leq 1$

2.3.7.1.2 Função de Autocorrelação Parcial

A Função de Autocorrelação parcial é utilizada para verificar a relação entre observações diferentes sem influência das intermediárias. Ou seja, deseja-se verificar exclusivamente a correlação entre Z_t e Z_{t-k} sem o efeito das intermediárias $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$. A função também é utilizada para encontrar os possíveis valores para as ordens “P” e “p”.

A FACP é dada por:

$$\rho_{kk} = Cov(Z_t, Z_{t-k} | Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}) \quad (2.9)$$

2.3.7.2 Estimação

O segundo passo, o passo de estimação, consiste na obtenção dos valores para os parâmetros dos modelos resultantes da identificação. Tais estimativas podem ser feitas, entre outros, pelo Método da Máxima Verossimilhança.

Para tanto, será necessário a verificação da normalidade das observações utilizando o teste de Shapiro-Wilk, além disso, é preciso que os resíduos tenham o padrão de um ruído branco, com média zero e variância constante, para a verificação desse resultado será construído o correlograma dos resíduos.

2.3.7.2.1 Teste de Shapiro-Wilk

O teste de Shapiro e Wilk (1965) tem como finalidade identificar a normalidade de uma distribuição. Para verificar se uma amostra segue uma distribuição normal, utiliza-se a estatística teste apresentada na equação 2.10 para as hipóteses:

- H_0 : A distribuição da amostra é normal
- H_1 : A amostra não é proveniente de uma distribuição normal

$$W = \frac{\left(\sum_{i=1}^n a_i x_i \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.10)$$

Onde:

- x_i é a i -ésima estatística de ordem
- \bar{x} é a média da amostra.
- a_i são constantes dadas pelo vetor $(a_1, a_2, \dots, a_n)^T$

2.3.7.3 Verificação

Segundo Morettin e Tolo (2006), a verificação consiste, basicamente, na examinação do modelo com o intuito de descobrir se ele faz boa representação dos dados através da análise dos resíduos. A indicação do bom ajuste do modelo é que a diferença dos valores observados e ajustados tenha as características de um ruído branco. Além disso, é necessário que as observações sejam i.i.d (independentes e identicamente distribuídas). Caso o teste mostre a não satisfação da suposição de ruído branco, será necessário voltar à etapa de identificação.

Existem diferentes maneiras utilizadas para o diagnóstico do modelo ajustado. Neste trabalho será utilizado o Teste de Ljung e Box (1978), que é uma variação do Teste de Box-Pierce (1970), que testa se o conjunto de autocorrelação dos resíduos estimados têm valor alto o suficiente para ser considerado não nulo. Além disso, será utilizado um gráfico correlograma dos resíduos composto pelos coeficientes de autocorrelação.

Para o Teste de Ljung e Box, o modelo é apropriado se $Q(K)$ tem distribuição qui-quadrada com $K-p-q$ graus de liberdade com fórmula apresentada na equação 2.11.

$$Q(K) = n(n + 2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \quad (2.11)$$

Com:

K = número de defasagens da função de autocorrelação;

p e q = ordens do modelo ajustado;

r = as autocorrelações estimadas.

2.3.7.4 Previsão

Após a verificação do modelo e constatação de um bom ajuste, na última etapa faz-se a previsão da série temporal, estimando Z_{t+h} , $h \geq 1$ e supondo conhecidas as observações até o instante de origem das previsões, ou seja, os elementos até o instante t , a saber, $\dots, Z_{t-2}, Z_{t-1}, Z_t$.

Para isso, será pressuposto que $W_t = (1 - B)^d Z_t = \Delta^d Z_t$ é estacionário, invertível e que já foram estimados os parâmetros do modelo.

Pela previsão dos erros quadráticos mínimos, a estimação do modelo Z_{t+h} , $h \geq 1$ poderá ser escrita como a esperança condicional de Z_{t+h} dado os valores anteriores; sendo assim:

$$\hat{Z}_{t+h} = E[Z_{t+h} | Z_t, Z_{t-1}, \dots] \quad (2.12)$$

O intervalo de confiança poderá ser construído quando aplicados os testes e verificado que os resíduos têm média zero, variância constante e que $Cov\{\varepsilon_s, \varepsilon_t\} = 0, \forall s \neq t$. Além disso, adicionalmente será suposto que os resíduos tenham distribuição normal. Com isso, tendo $N(\widehat{Z}_{t+h}, V(h))$ como distribuição condicional de Z_{t+h} , tem-se:

$$U = \frac{Z_{t+h} - \widehat{Z}_t(h)}{[V(h)]^{\frac{1}{2}}} \sim N(0, 1) \quad (2.13)$$

Sendo $V(h)$ estimado por σ_t^2 .

Através disso, é possível calcular o intervalo de confiança através da equação 2.14.

$$\widehat{Z}_t(h) \pm 1,96[\widehat{V}_z(h)]^{\frac{1}{2}} \quad (2.14)$$

2.3.8 Medidas De Seleção Do Modelo

Para a seleção do melhor e mais adequado modelo, serão utilizadas medidas de aderência. O modelo utilizado será o que tiver os menores resultados nos testes a seguir:

2.3.8.1 Critério de Informação de Akaike (AIC)

O critério de informação de Akaike (AIC), proposto por Akaike (1974), mensura a qualidade do modelo estatístico. E pode ser expresso através da equação 2.15.

$$AIC = -2 \ln(L_p) + 2[(p + 1) + 1] \quad (2.15)$$

com os parâmetros:

- L_p : é a função de MV;
- p : é o número de parâmetros.

2.3.8.2 Raiz Quadrada do Erro Quadrático Médio

A Raiz Quadrada do Erro Quadrático Médio (RMSE), é o desvio padrão dos valores residuais. Seus resultados, assim como os do MSE (que avalia a diferença entre o valor observado e o valor ajustado de maneira quadrática), são sempre maiores ou iguais a 0.

Sua expressão é descrita na expressão 2.16.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z_t - \hat{Z}_t)^2}{n}} \quad (2.16)$$

com os parâmetros:

- Z_t : valor observado;
- \hat{Z}_t : valor ajustado;
- n : tamanho da série.

2.3.8.3 Erro Médio Percentual Absoluto

O Erro Médio Percentual absoluto (MAPE) representa, percentualmente, o quanto o modelo erra sem que os pontos negativos compensem os positivos. Seus valores variam de 0 a 100.

$$MAPE = \sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \times \frac{100}{n} \quad (2.17)$$

com os parâmetros:

- Z_t : valor observado;
- \hat{Z}_t : valor ajustado;
- n : tamanho da série.

3. Resultados

3.1 Análise Descritiva

De 2017 a 2022, no estado do Rio de Janeiro, aconteceram cerca de um milhão duzentos e vinte mil nascimentos, de acordo com o SINASC. A Figura 1 apresenta o número de nascidos vivos no estado do Rio de Janeiro por local de residência da mãe no período. Através dela é possível identificar uma queda acentuada no número de nascimentos ao longo dos anos. O número foi reduzido de pouco mais de 223 mil para 180 mil; o que representa uma queda de 19,23%.

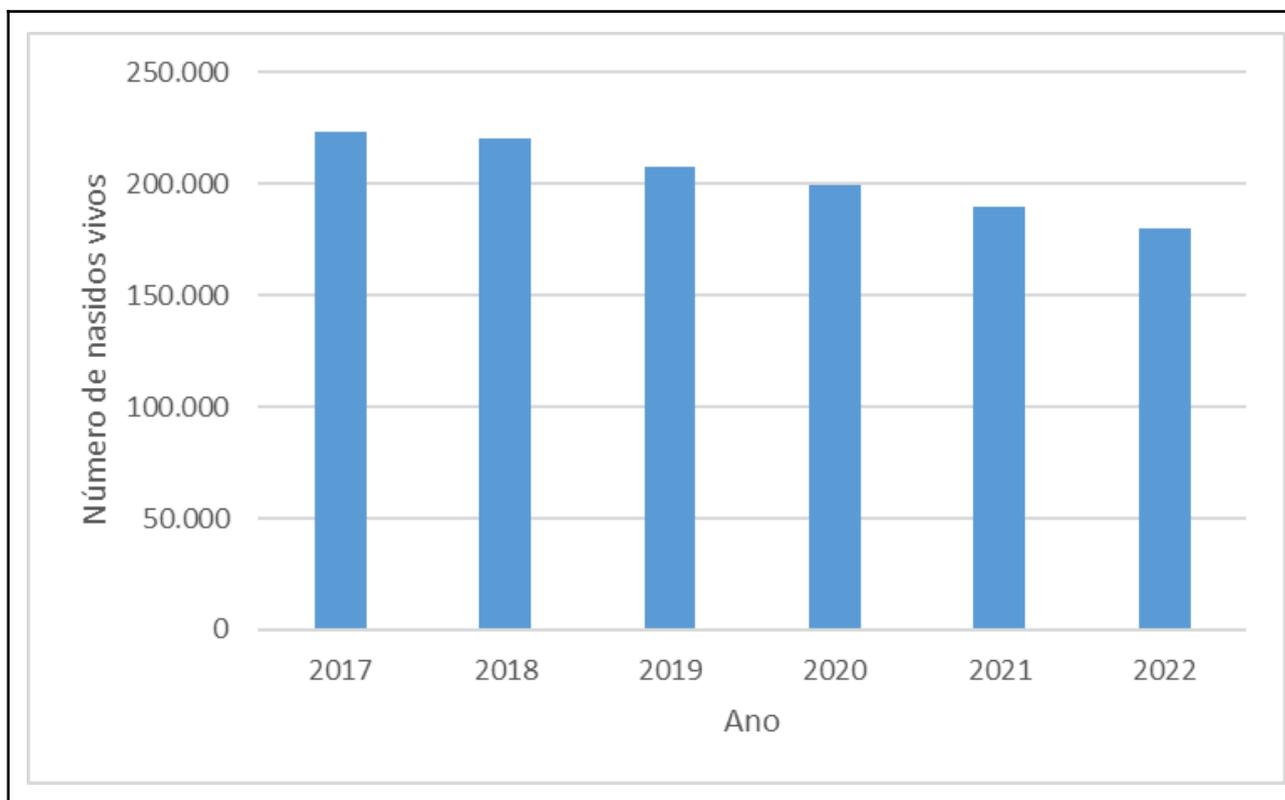


Figura 1: Nascidos vivos no estado do Rio de Janeiro pelo local de residência da mãe.

A Figura 2 apresenta o número de nascimentos de 2017 a 2022 categorizado pela idade da mãe a partir de 15 anos. Observa-se, através dela, a predominância na ocorrência de maternidade em mulheres entre 20 e 29 anos de idade em todos os anos estudados. Além disso, é interessante ressaltar que o padrão de todas as faixas etárias é o decréscimo no número de nascimentos ao longo dos anos, com exceção das mulheres com mais de 40 anos. Esse decaimento aparece de forma mais expressiva nas mulheres mais jovens.

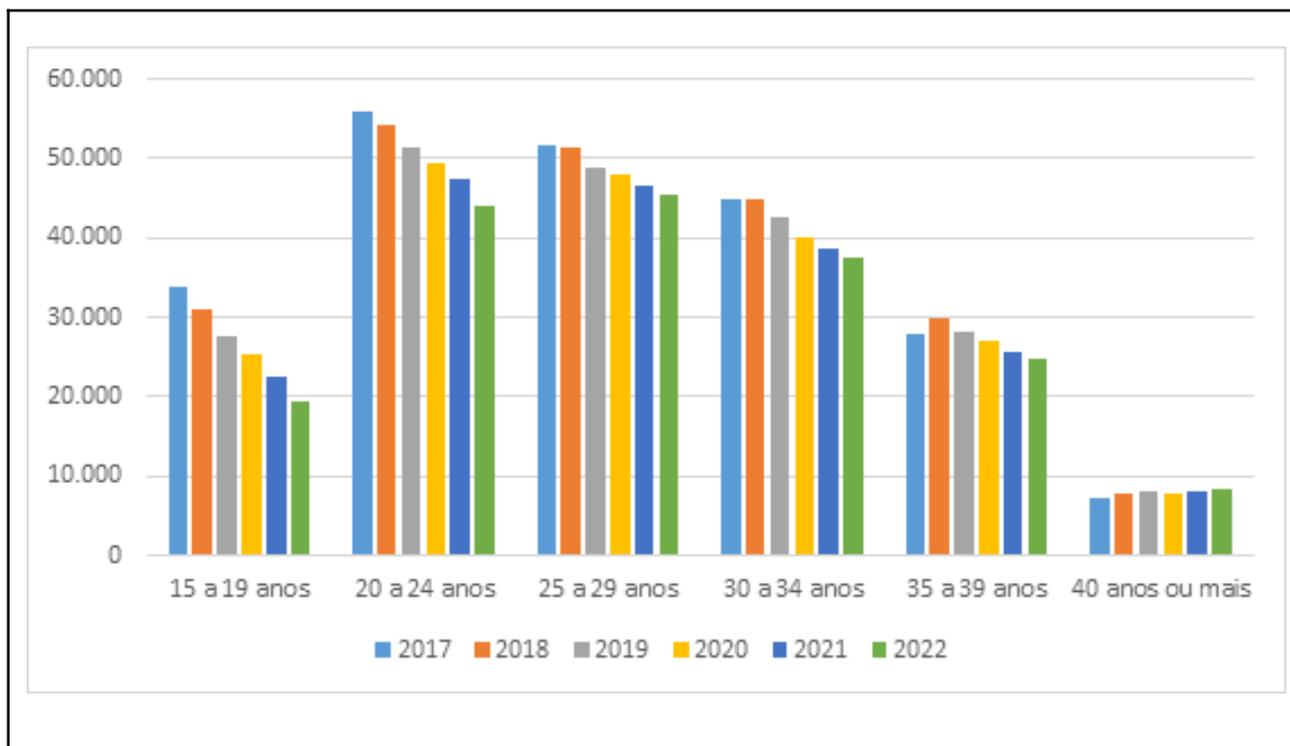


Figura 2: Número de nascimentos por faixa de idade da mãe.

Para identificar a relação entre os anos de estudo da mulher e o número de nascimentos foi desenvolvida uma tabela de distribuição de frequências. A Tabela 2 mostra a relação entre o número de nascidos vivos e o nível de escolaridade da mãe.

Tabela 2: Relação entre o número de nascidos vivos e o nível de escolaridade da mãe.

Instrução da mãe	2017	2018	2019	2020	2021	2022
Nenhuma	291	282	251	242	172	168
1 a 3 anos	3.368	2.913	2.560	2.241	1.949	1.608
4 a 7 anos	37.464	35.655	32.389	28.826	25.392	22.093
8 a 11 anos	133.141	132.773	125.259	122.484	117.001	111.402
12 ou mais	45.250	45.622	43.626	41.370	41.093	41.709
Total	219.514	217.245	204.085	195.163	185.607	176.980

Os dados, também obtidos através do SINASC, mostram que as mulheres com nível escolar entre 8 e 11 anos (ensino médio incompleto) compõem a faixa com o maior número de

concepções. Atrás delas estão as mulheres com o ensino médio completo. Na Figura 3 é possível identificar esses padrões.

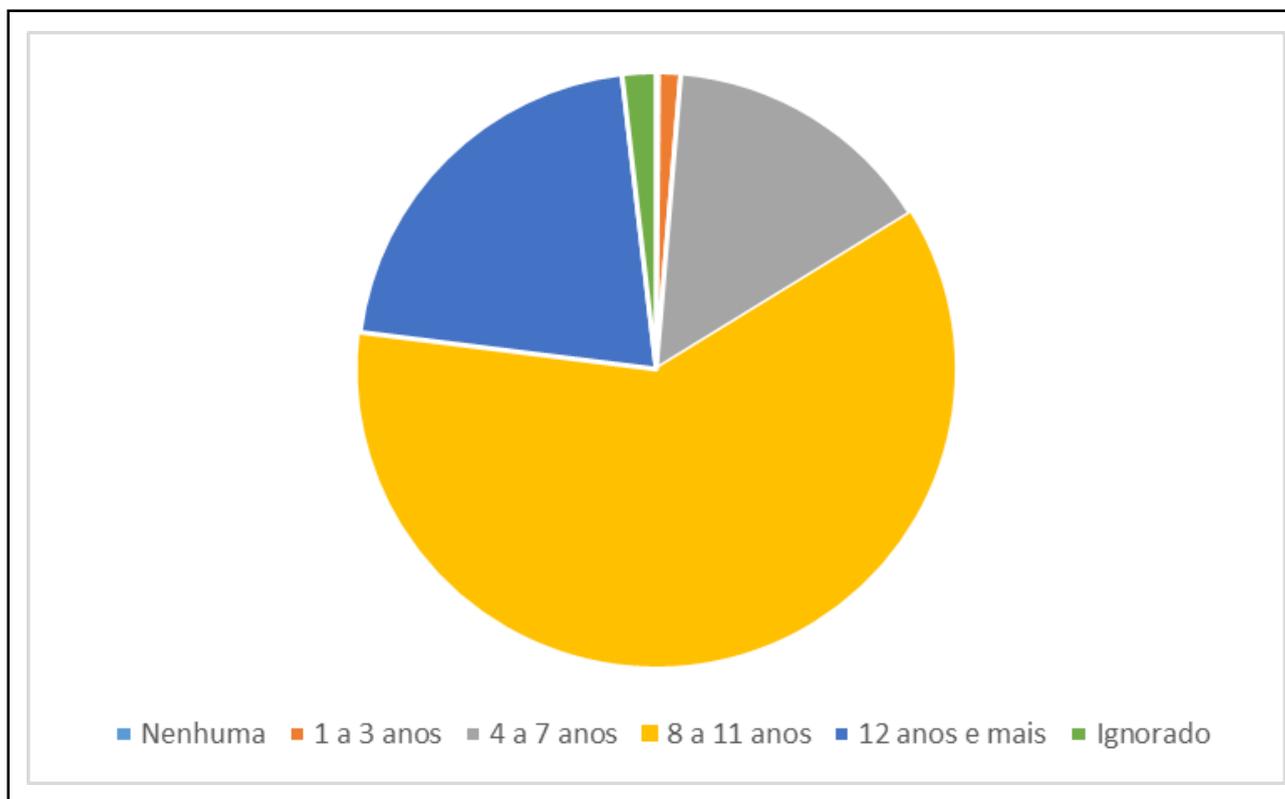


Figura 3: Proporção de filhos por mulheres segundo o nível de instrução entre 2017 e 2022.

3.2 Indicadores Demográficos

A seguir serão apresentados os indicadores demográficos de natalidade e fecundidade a fim de obter a relação de nascidos vivos antes e durante a pandemia.

Na Tabela 3 é apresentada a Taxa Bruta de Natalidade e a Taxa de Fecundidade Geral de 2017 a 2022 para cada mil habitantes.

Para cada mil habitantes, o número de nascidos vivos caiu de pouco mais de 13 para 10,27. É possível, através dos resultados, novamente identificar a queda da taxa de nascidos vivos no estado do Rio de Janeiro. Para uma análise de maior precisão, foram feitos os cálculos da Taxa de Fecundidade geral que, conforme apresentado, expõe que para cada mil mulheres com idade entre 15 e 49 anos do Rio de Janeiro a TFG passou de pouco mais de 44 para menos de 36.

Tabela 3: Taxa Bruta de Natalidade e Taxa de Fecundidade Geral de 2017 a 2022 para cada mil habitantes do Rio de Janeiro.

Ano	TBN	TFG
2017	13,0915	44,3345
2018	12,8503	43,6978
2019	12,0476	40,7502
2020	11,4663	38,9447
2021	10,8725	38,0264
2022	10,2693	35,7856

Para verificar preliminarmente a relação entre a idade da mulher e o número de nascimentos, este trabalho buscou calcular a Taxa de Fecundidade Total. Para isso foi desenvolvida a Taxa Específica de Fecundidade no Rio de Janeiro entre mulheres de 15 a 49 anos, como mostra a Figura 4.

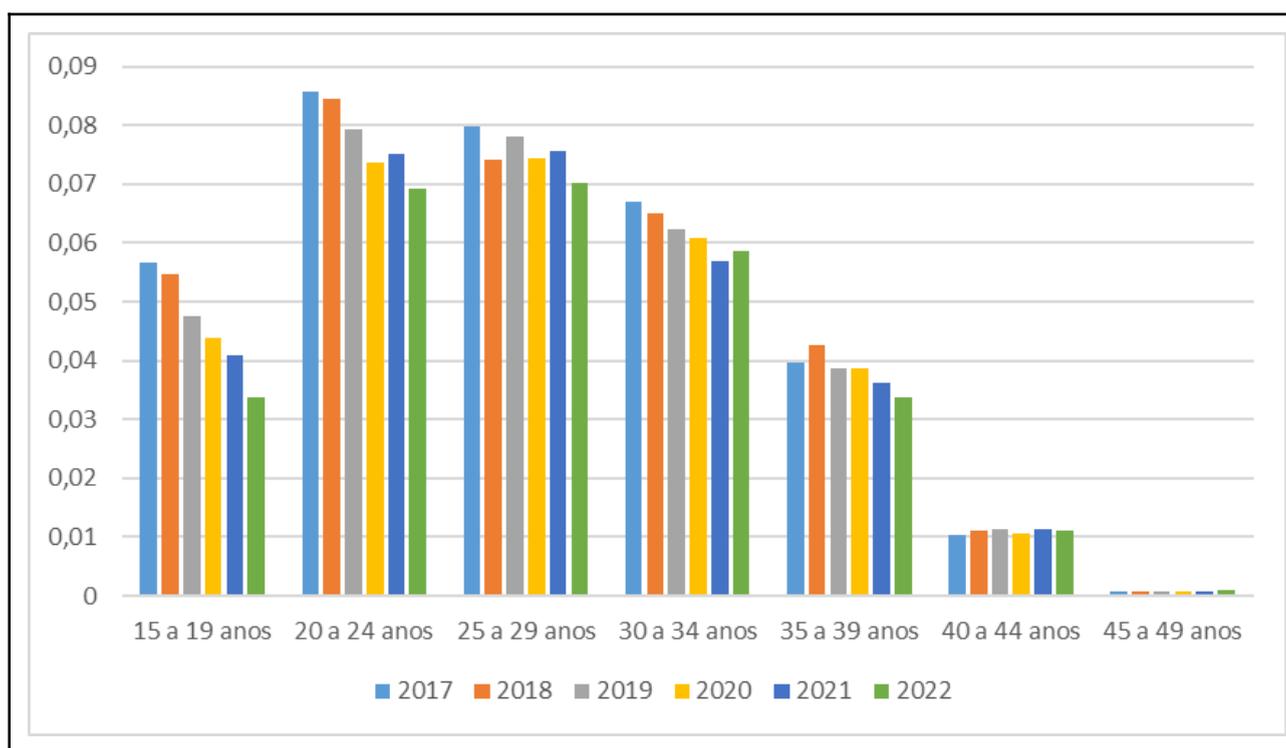


Figura 4: Taxa Específica de Fecundidade das mulheres do estado do Rio de Janeiro entre 2017 e 2022

Com base na TEF foi construída a Taxa de Fecundidade Total, como é apresentado na Tabela 4. O Rio de Janeiro teve, em 2022, a TFT inferior a 1,4. Ou seja, em média, em toda a sua vida, uma mulher fluminense terá 1,4 filhos. Tal estimativa está muito abaixo do nível de reposição de 2,1.

Para fins de comparação, em 2000, no Brasil, essa taxa era de aproximadamente 2,39, de acordo com o IBGE.

Tabela 4: Taxa de Fecundidade Total de 2017 a 2022 do Rio de Janeiro.

2017	2018	2019	2020	2021	2022
1,7002	1,6644	1,5893	1,5115	1,4836	1,3879

3.3 Série Temporal

Para estimar o número de nascidos vivos caso a pandemia não ocorresse foram analisados os componentes e o comportamento da série temporal dos nascidos vivos entre 2010 e 2022. Os resultados das análises e a modelagem selecionada foram dispostos a seguir.

3.3.1 Análises Preliminares

Os dados foram coletados do SINASC e apresentam a relação mensal de nascidos vivos no estado do Rio de Janeiro. Na Figura 5 pode ser observado o gráfico da série original.

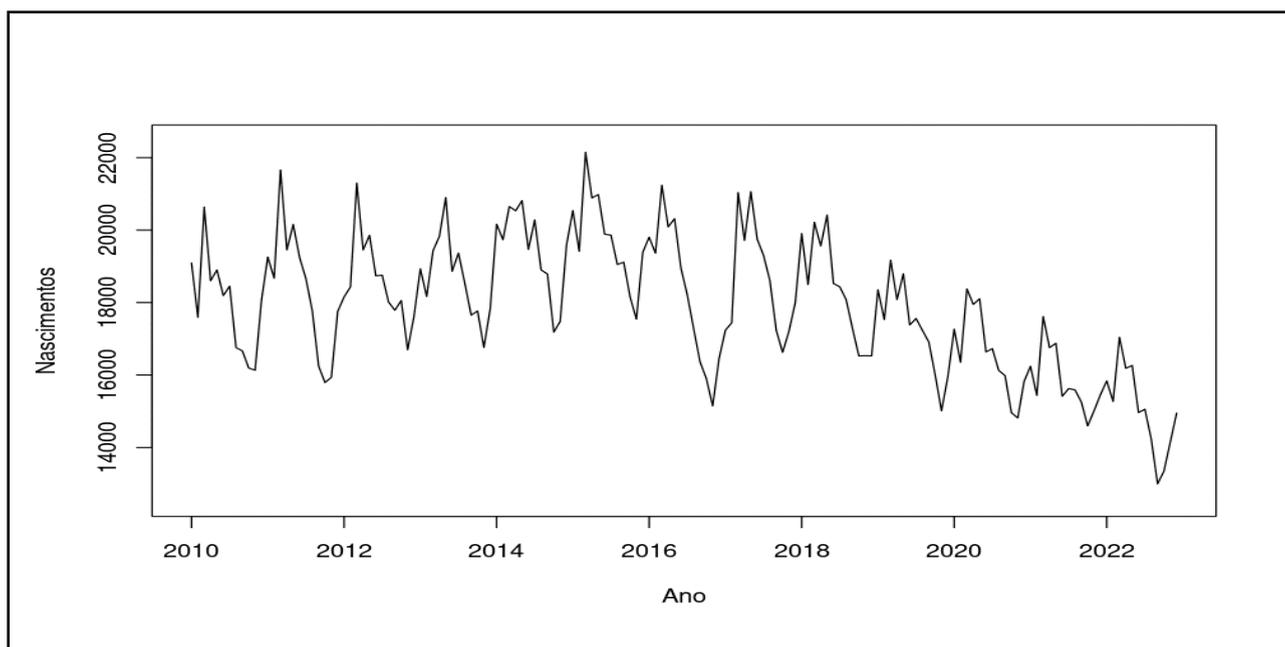


Figura 5: Nascimentos no estado do Rio de Janeiro de 2010 a 2022.

É possível identificar visualmente a presença de sazonalidade pela repetição anual de padrão. Além disso, observa-se uma tendência decrescente desde 2015, que concorda com todos os estudos mencionados no primeiro capítulo.

Para alcançar o objetivo deste estudo (projetar o número de crianças nascidas caso a pandemia não existisse), foi filtrada uma série apenas com os dados até dezembro de 2019 e as características dessa nova série serão apresentadas a seguir.

Para melhor identificação, dos padrões sazonais da série, a Figura 6 é a representação de um gráfico polar, que apresenta o comportamento mensal da série ao longo dos anos.

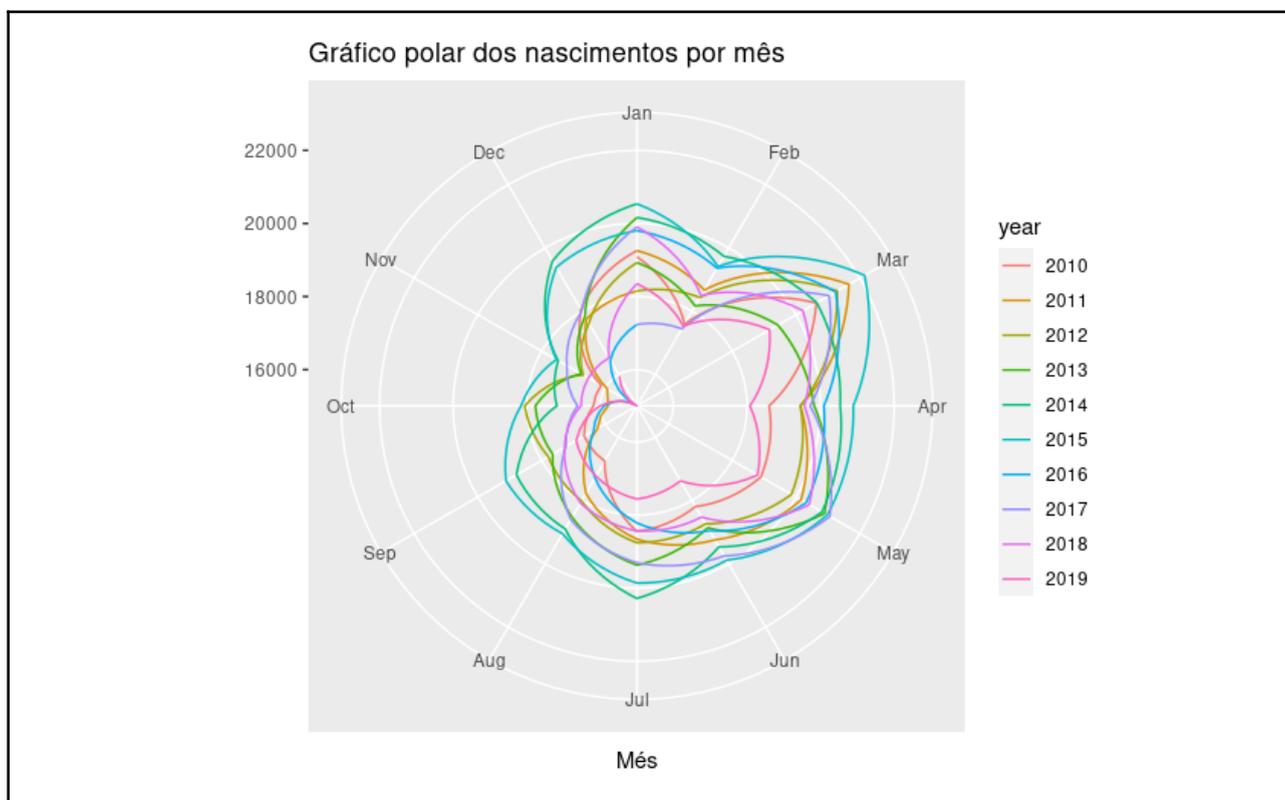


Figura 6: Gráfico polar dos nascimentos no estado de 2010 a 2019.

Através da imagem é possível identificar que existe um padrão para cada um dos meses do ano. A maior incidência de novos nascimentos é percebida no mês de março, com média de 20.749 partos, enquanto novembro tem a menor média, 16.444 nascimentos). Através dos padrões mensais identificados é confirmada a sazonalidade da série.

A fim de identificar separadamente os componentes da série, a mesma foi desconstruída em tendência, sazonalidade e resíduos, como é apresentado na Figura 7.

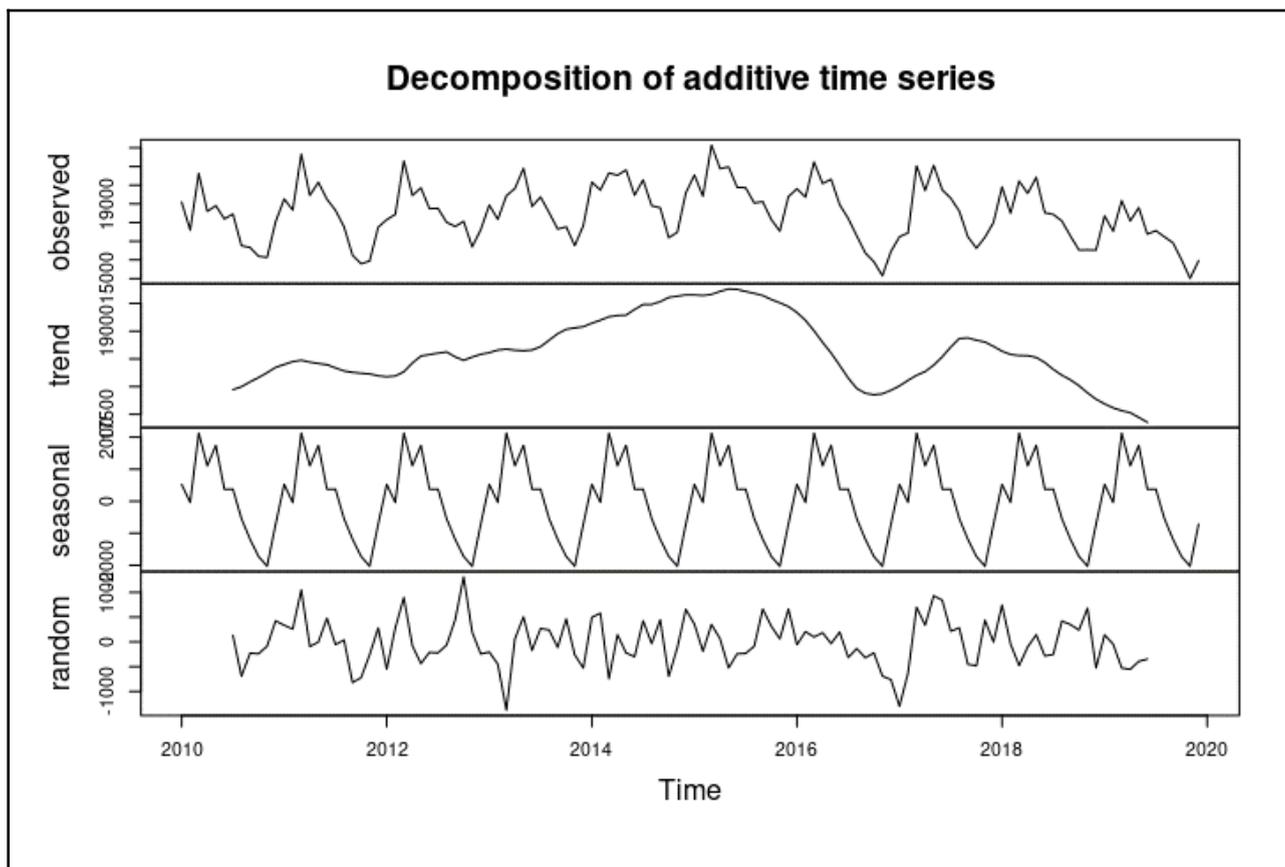


Figura 7: Decomposição da série dos nascimentos no estado de 2010 a 2019.

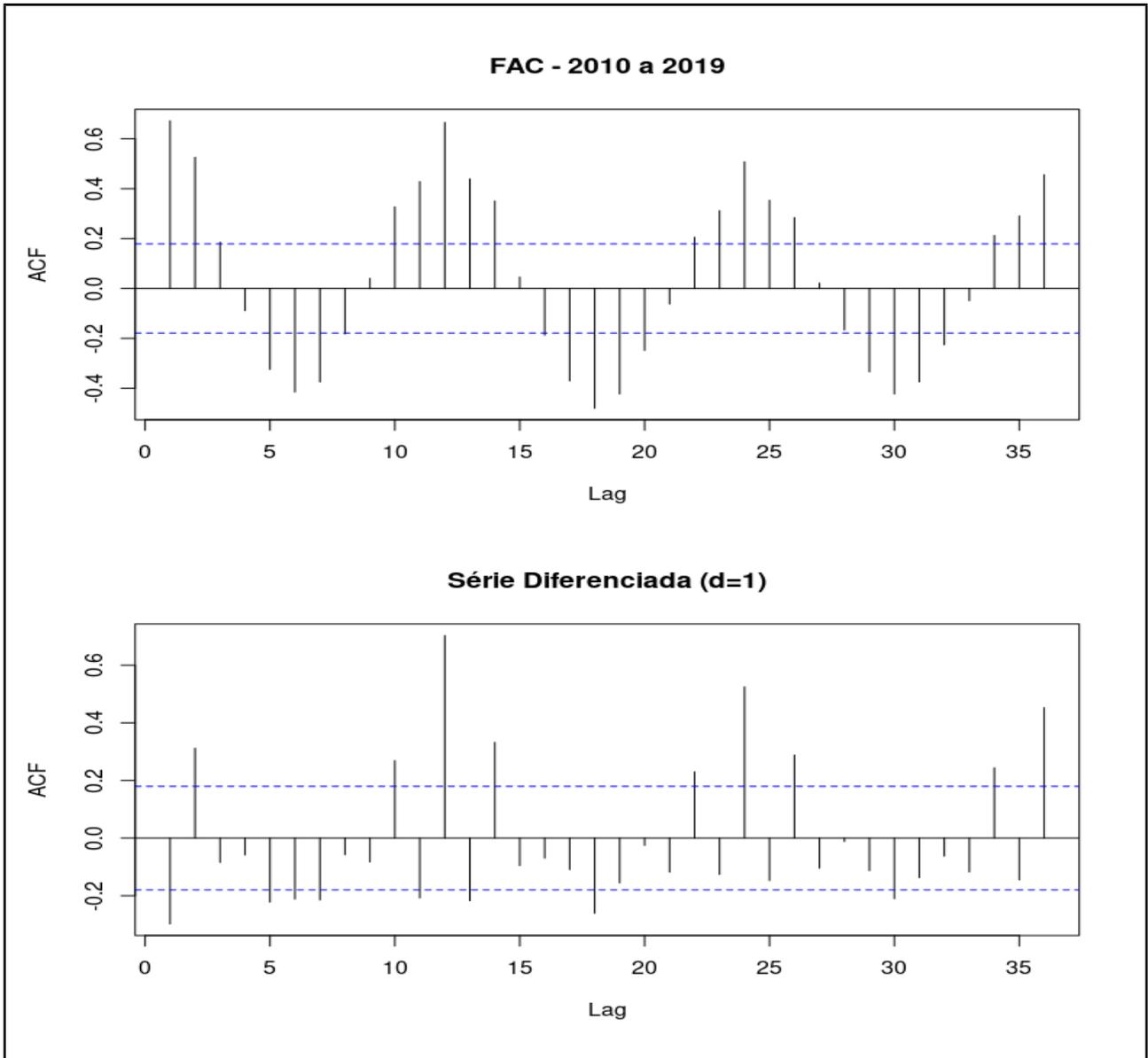
Identifica-se, na Figura 7, o gráfico de decomposição da série, onde, no primeiro quadro é apresentada a série observada, no segundo está a tendência da série, mais uma vez apresentando um comportamento crescente nos primeiros anos analisados, seguido por outro decrescente a partir do ano de 2015. O terceiro quadro contém a sazonalidade observada na série, com os meses de março e maio com os maiores picos e o mês de novembro com a menor elevação. No último quadro da Figura 7 encontram-se os resíduos.

A partir das componentes apresentadas observou-se a necessidade da utilização de um modelo SARIMA para alcançar os objetivos do estudo.

3.3.2 Identificação

Para utilizar a metodologia de Box & Jenkins na confecção do modelo SARIMA, o primeiro passo é a identificação, que consiste na verificação da estacionariedade da série e na identificação dos candidatos a parâmetros do modelo. Para tanto, foi utilizado o teste ADF e os gráficos da FAC,

tanto para identificar a estacionariedade quanto para apontar possíveis valores para os coeficientes p , q , P e Q do modelo SARIMA.



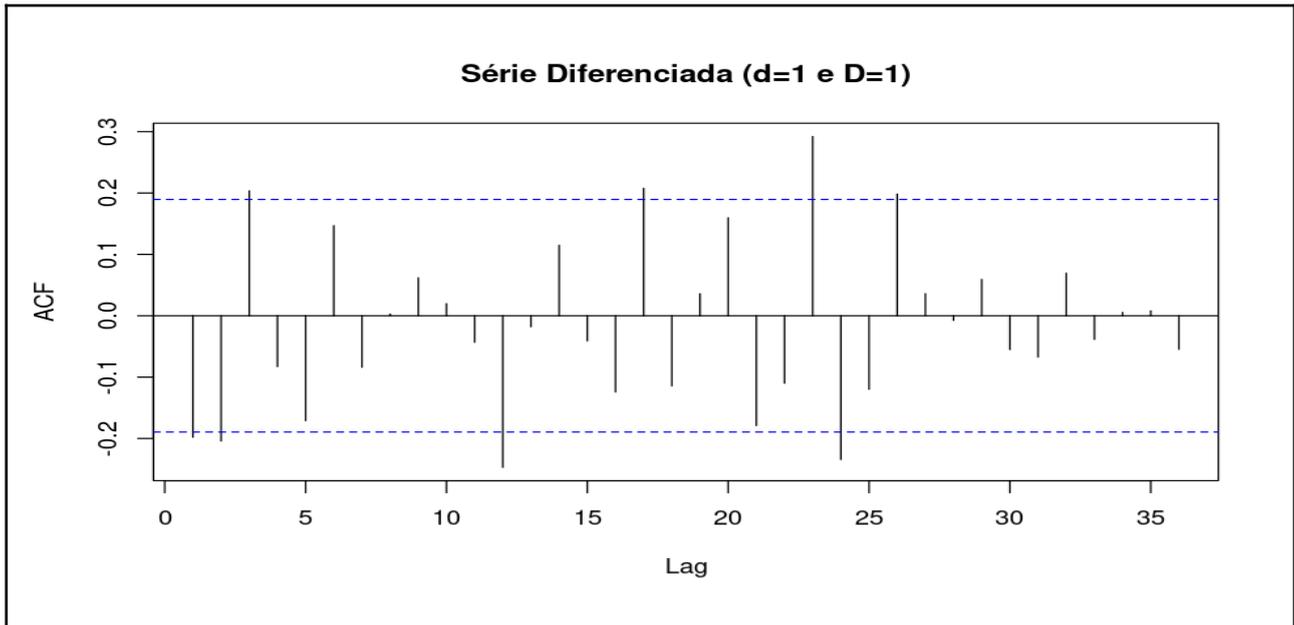


Figura 8: Gráfico de dispersão da FAC e suas modificações.

No primeiro gráfico é visível a presença de padrões repetitivos a cada 12 lags e nenhum deles é não significativo. Não é possível, também, identificar a diminuição do tamanho dos lags à medida que seus valores aumentam, sendo assim, a série é claramente não estacionária.

O segundo gráfico da Figura 8 é a FAC da série diferenciada uma vez. Percebe-se a quebra de alguns padrões, porém os lags múltiplos de 12 persistem em sua exagerada significância, indicando que os traços sazonais continuam presentes na série.

Foi aplicada, então, nova diferenciação com lag=12 e sua FAC pode ser verificada no terceiro gráfico da Figura 8. De acordo com a representação, embora os dois primeiros múltiplos de 12 ainda apresentam certa significância, nota-se que o lag=36 perdeu sua expressividade. Dessa forma, existem indicativos de que a série está livre de tendência e sazonalidade; pode-se, assim, seguir com processo de modelagem.

Para comprovar se a série está realmente livre de tendência e sazonalidade, foi aplicado o teste ADF na série diferenciada. Os resultados apresentaram p-valor < 0,01, sendo assim, rejeita-se a hipótese nula de não estacionariedade a um nível de significância de 5%. Com isso, é possível afirmar que a série é estacionária. Com essa confirmação, serão verificados os gráficos da FAC e FACP para encontrar os candidatos a parâmetros do modelo.

Na Figura 9 é apresentada a série diferenciada e sua FAC e FACP.

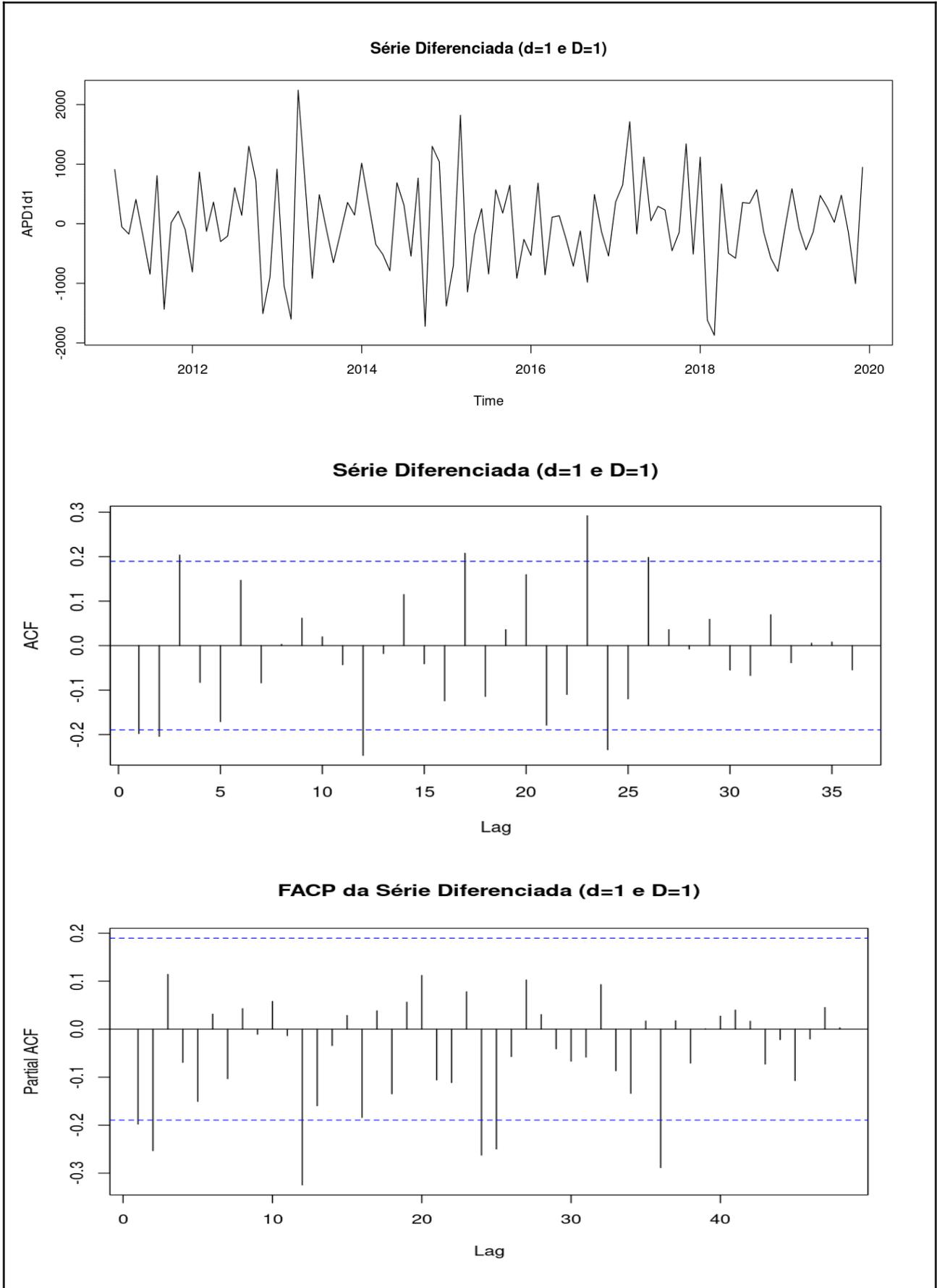


Figura 9: Gráfico da série derivada e da sua FAC e FACP.

Com a FACP e FAC é possível identificar prováveis valores para os parâmetros “p” e “P” (através da FACP) e “q” e “Q” (através da FAC).

Analisando a FAC no segundo gráfico da Figura 9, observam-se os três primeiros lags ultrapassando levemente a linha de significância, indicando $q=3$. Porém, como o avanço dos mesmos é pequeno, será necessário utilizar as medidas de seleção de modelo apresentadas na seção 2.3.8. Com relação aos lags múltiplos de 12 do gráfico da FAC, os lags 12 e 24 mostram-se significativos, isso indica um parâmetro $Q=2$.

Através do gráfico da FACP é possível identificar os dois primeiros lags ultrapassando a borda, porém, assim como no gráfico da FAC, o lag=1 não se afasta tanto da linha de significância, sendo necessária a aplicação dos testes de medida de seleção de modelo para escolha do valor de “p”. Observando os lags múltiplos de 12, os três primeiros ultrapassaram a faixa que indica significância; assim, pode-se apontar o valor de “P” como 3.

Os parâmetros candidatos estão listados na Tabela 5.

Tabela 5: Possíveis valores para as ordens p, q, P e Q do modelo SARIMA.

Ordens	Possíveis valores
p	1 ou 2
q	0 ou 3
P	3
Q	2

Foram analisadas as combinações dos parâmetros listados na Tabela 5 através das medidas de aderência de AIC, MAPE e RMSE. De acordo com os testes aplicados, o melhor modelo foi o $SARIMA(2, 1, 0)(3, 1, 2)_{12}$. Na Tabela 6 foram apresentados alguns dos testes feitos em possíveis modelos inclusive no modelo fornecido pelo comando `auto.arima(SARIMA(4, 1, 1)(2, 1, 1)_{12})` que está disponível no pacote `forecast` (Hyndman, 2018) no programa estatístico RStudio.

Tabela 6: AIC, MAPE e RMSE dos modelos testados.

Modelos	RMSE	MAPE	AIC
$SARIMA (1, 1, 0)(3, 1, 2)_{12}$	459.9966	1.843420	1691.798
$SARIMA (1, 1, 3)(3, 1, 2)_{12}$	460.6917	1.831784	1693.150
$SARIMA (2, 1, 0)(3, 1, 2)_{12}$	455.9940	1.819845	1688.793
$SARIMA (2, 1, 3)(3, 1, 2)_{12}$	495.3872	1.937769	1688.934
$SARIMA (4, 1, 1)(2, 1, 1)_{12}$	515.8656	2.017578	1690.627

3.3.3 Estimação e Análise dos Resíduos

Após a escolha do modelo $SARIMA (2, 1, 0)(3, 1, 2)_{12}$ será feita a análise dos resíduos do modelo. É preciso que os resíduos tenham o padrão de um ruído branco, com média zero e variância constante, além disso, será verificada a normalidade dos resíduos.

A Figura 10 mostra o comportamento dos resíduos do modelo selecionado. Percebe-se no primeiro quadro o comportamento relativamente constante com leves picos em pontos específicos, contudo seus valores circundam a média zero e não existem indicativos de uma tendência positiva ou negativa no gráfico.

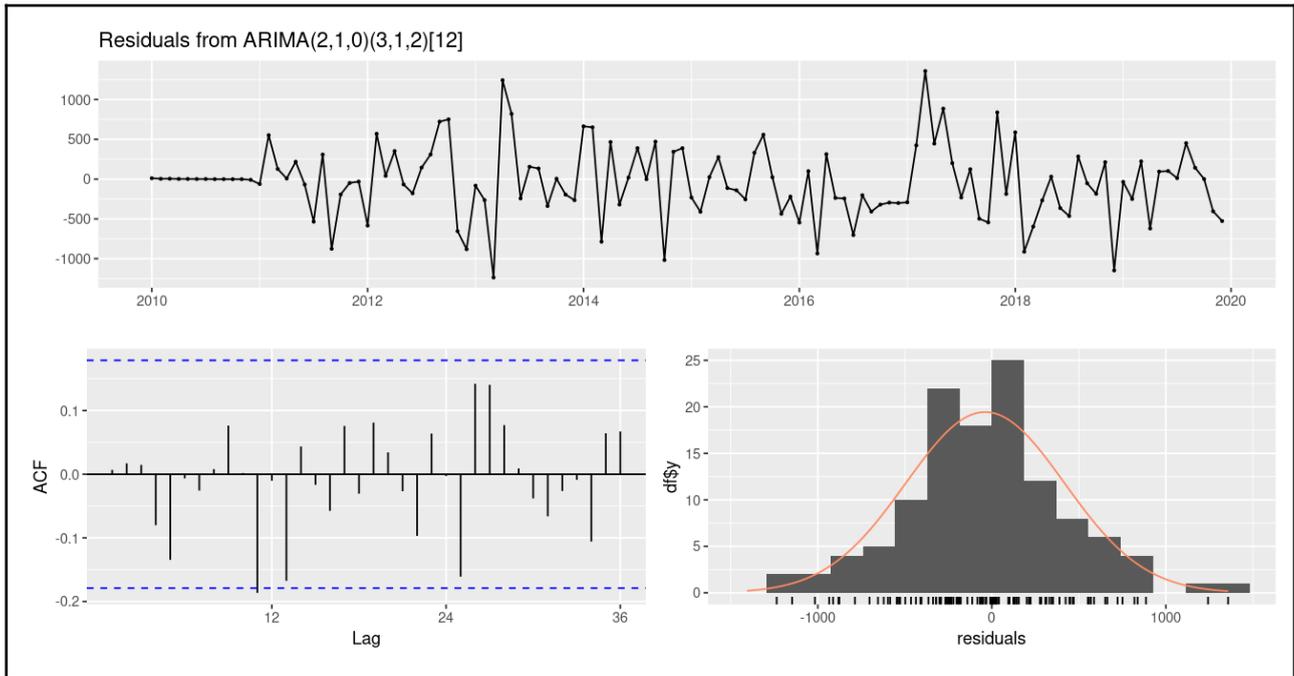


Figura 10: Características dos resíduos do modelo $SARIMA(2, 1, 0)(3, 1, 2)_{12}$

Através da FAC dos resíduos, apresentada no segundo quadro, não é possível observar valores significativamente diferentes de 0 a não ser por um pequeno avanço do lag=11, podendo-se assumir que não são correlacionados.

Para confirmar a não presença de autocorrelação dos resíduos, foi aplicado o teste de Box-Pierce-Ljung. O resultado do teste apresentou p-valor=0.9399. Sendo assim, como o p-valor é maior que 0,05, não rejeita-se a hipótese nula, confirmando assim a independência dos resíduos. Sendo assim, é possível seguir para a etapa de previsão a partir do modelo escolhido.

Para o teste da normalidade dos resíduos do modelo foi usado o método de Shapiro-Wilk. Como o p-valor do teste foi de 0,2025 (maior que 0,05), não rejeita-se a hipótese nula a um nível de significância de 5%. Esse resultado é verificado observando o histograma dos resíduos do modelo no terceiro quadro da Figura 10. Com isso, pode-se afirmar que os resíduos do modelo tem distribuição normal. Sendo assim, é possível afirmar que os resíduos do modelo escolhido tem as características de um ruído branco, conforme era desejado para o uso do modelo.

3.3.4 Previsão

Nesta etapa foi utilizado o modelo para prever o número de nascimentos caso não acontecesse a pandemia. A Figura 12 apresenta os números observados até dezembro de 2019 com uma linha preta, a azul representa a previsão pontual do modelo para os anos de 2020, 2021 e 2022 e a sombra cinza apresenta a previsão intervalar com um nível de confiança de 95%.

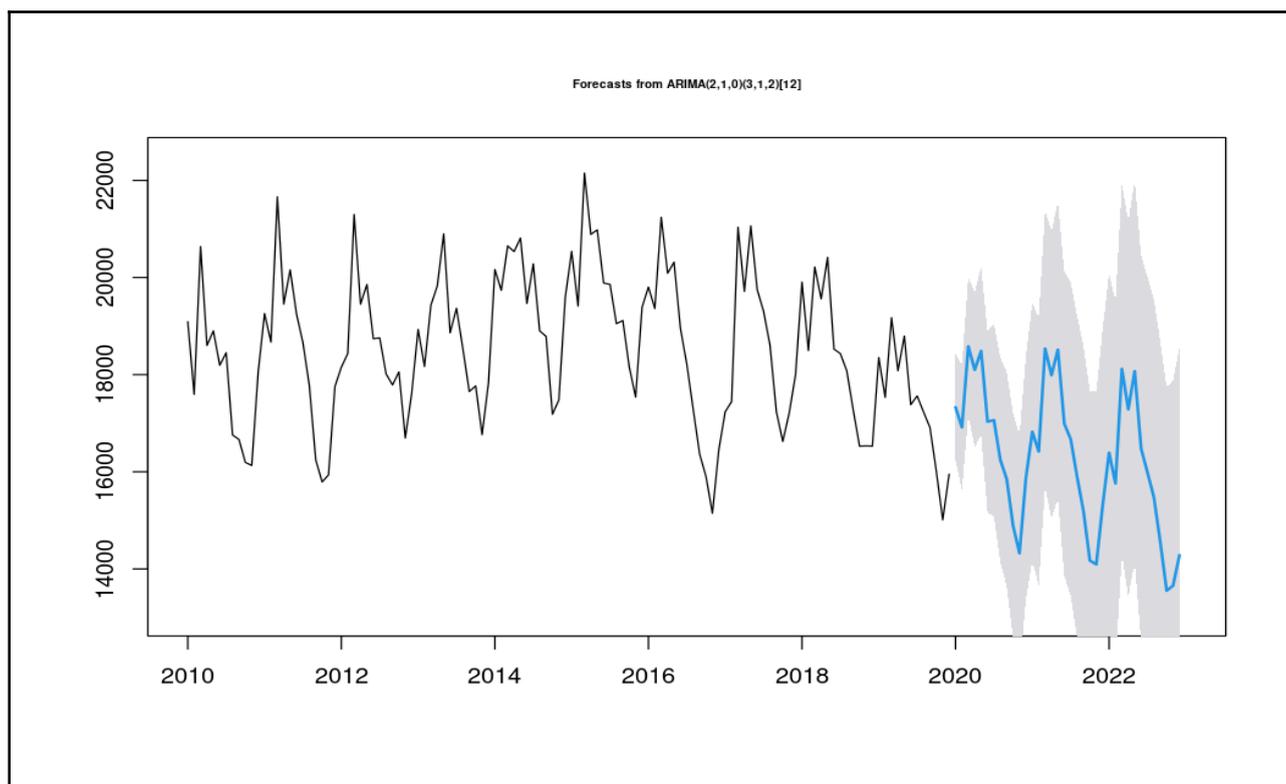


Figura 11: Gráfico da série com a previsão de 36 meses à frente

Pelo gráfico na Figura 11 observa-se que, de acordo com o modelo, o número de nascimentos continuaria caindo ano após ano no estado do Rio de Janeiro. Ressalta-se, nessa previsão, os valores de outubro e novembro de 2022 que ficaram abaixo de 14.000.

Para comparar com o número real de nascimentos nos anos estudados, foi elaborado o gráfico da Figura 12 com a linha em preto representando o número de nascimentos reais de 2010 a 2022 e a linha azul sendo a previsão para os 3 últimos anos dessa série. A imagem deixa clara a diferença nos dois últimos anos analisados.

No ano de 2020 a linha de previsão é quase exatamente coincida ao número real; isso acontece devido ao período gestacional. Em outras palavras, as pessoas que ficaram grávidas no

início da pandemia no Brasil, que aconteceu em março de 2020, só deram à luz em dezembro daquele ano, onde a linha de previsão começa a se afastar da linha de nascimentos, de fato.

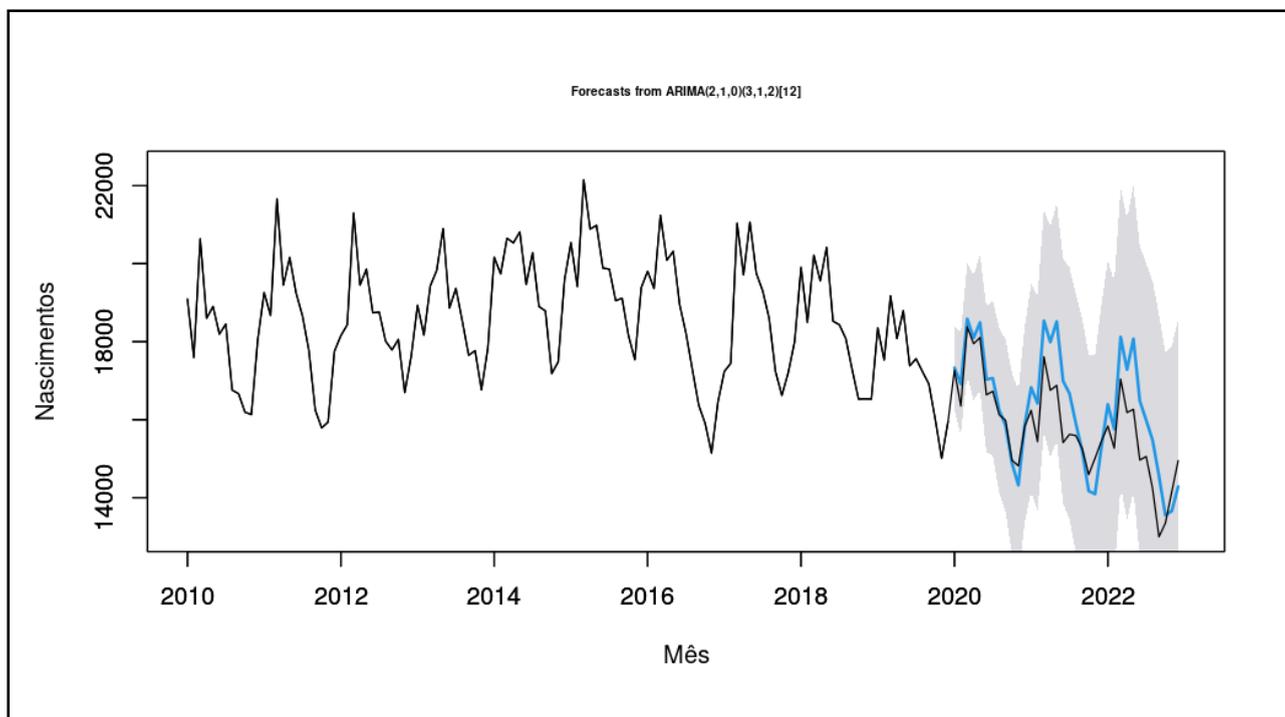


Figura 12: Gráfico do número de nascimentos previstos x número de nascimentos reais durante a pandemia da COVID-19 no estado do Rio de Janeiro

Nos dois últimos anos o número de nascimentos a cada mês mostra-se muito distante da previsão, sendo o meado do ano o período com maior diferença em ambos os anos. Apesar disso, nenhum dos valores ficou fora do intervalo de confiança, como também pode ser observado na Figura 12.

Na Tabela 10 são apresentados os valores da diferença entre a previsão de nascimentos e o número real nos anos de 2020, 2021 e 2022 no estado do Rio de Janeiro. Na tabela, confirma-se a maior relevância nos dois últimos anos da pandemia e, além disso, a maior diferença no meado dos anos estudados.

Tabela 7: Número real de nascimentos de 2020 a 2022 e a previsão do modelo

Ano	Nº de nascimentos		Diferença	
	Previstos	Observados	Absoluta	Relativa (%)
2020	200.675	199.124	-1.551	-0,8
2021	196.594	189.866	-6.728	-3,5
2022	198.615	180.288	-18.327	-10,2
Total	595.884	569.278	-26.606	-4,7

Como apresentado na Tabela 7, 26.606 crianças deixaram de nascer durante a pandemia. Isso equivale a 4,7% do número previsto pelo modelo. É interessante realçar que no ano de 2020 (primeiro ano da pandemia da COVID-19) a diferença entre o número previsto e o número real de crianças nascidas no estado do Rio de Janeiro não é grande, apenas 1.551 nascimentos a menos, visto que os 9 primeiros meses de nascimentos são de bebês cuja as mães estavam grávidas antes do início da pandemia no Brasil (que aconteceu em março de 2020).

4. Conclusão

A diminuição da natalidade é um assunto global é de suma importância devido aos seus impactos na sociedade. O envelhecimento populacional afeta a economia e essa por sua vez afeta a previdência que afeta o sistema de saúde gerando uma onda de incertezas a respeito do futuro para a população deste século.

O trabalho desenvolvido buscou verificar o quanto a taxa de natalidade foi afetada pela pandemia nos anos de 2020 a 2022 no estado do Rio de Janeiro. Sendo possível identificar que a partir do ano de 2021 os resultados se distanciaram do conquistado pelo modelo de previsão, uma vez que a pandemia começou em março de 2020 e o período gestacional dura em torno de 9 meses.

De acordo com os resultados obtidos neste trabalho, todas as taxas relacionadas à natalidade que foram analisadas já vinham sofrendo queda antes da pandemia com as mudanças sociais e econômicas que aconteceram no Brasil nesse período. Ainda assim, o pandemia da Covid-19 mostrou resultados abaixo do previsto mesmo nessas circunstâncias.

Os instrumentos utilizados para a verificação das características da natalidade no estado foram os indicadores demográficos (TBN, TFG, TEF e TFT) que mostraram, de 2017 a 2022, o comportamento dos nascimentos da população fluminense. Para verificar quantos bebês nasceriam caso a pandemia não acontecesse foi modelado um SARIMA.

Verificou-se nas análises empíricas o decaimento contínuo no número de nascimentos desde 2017. Além disso, observou-se que a idade da mãe pode ser inversamente proporcional ao decaimento de nascidos vivos entre 2017 e 2022. Em relação ao nível de escolaridade da mãe, verificou-se que 60,77% dos nascimentos vieram de mulheres com ensino médio incompleto.

Observando-se a TFT da série, foi possível identificar a redução do número de filhos que uma mulher teria no estado do Rio de Janeiro de 1,7002 para 1,3879.

Pelo modelo obtido através do modelo $SARIMA(2, 1, 0)(3, 1, 2)_{12}$, um percentual de 4,7% das crianças que deveriam nascer no estado do Rio de Janeiro não nasceram no período da pandemia da COVID-19. Esse percentual representa um total de 26.606 nascimentos que não

aconteceram em durante esse evento. Mostra-se necessária a elaboração de um estudo após a divulgação dos dados de nascimentos no estado relativo a 2023, para verificação da persistência ou não dessa discrepância.

Finalmente o trabalho apresentado visa contribuir para o desenvolvimento do tema tanto em futuras pesquisas, quanto no esclarecimento para a população. É necessário o acompanhamento do número de crianças nos anos seguintes à pandemia para identificação de possíveis sequelas da mesma deixada não só no estado do Rio de Janeiro, como também no resto do mundo.

Referências

- BOX, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509-1526.
- CERQUEIRA, C. A., & Givisiez, G. H. N. (2015). Conceitos básicos em demografia e dinâmica demográfica brasileira. *Livros*, 13-44.
- CHANDRA, S., Christensen, J., Mamelund, S. E., & Paneth, N. (2018). Short-term birth sequelae of the 1918–1920 influenza pandemic in the United States: state-level analysis. *American journal of epidemiology*, 187(12), 2585-2595.
- COUTINHO, R. Z., Lima, L. C. D., Leocádio, V. A., & Bernardes, T. (2020). Considerações sobre a pandemia de Covid-19 e seus efeitos sobre a fecundidade e a saúde sexual e reprodutiva das brasileiras. *Revista Brasileira de Estudos de População*, 37.
- DE FOZ, G. (2021). Métodos demográficos: uma visão desde os países de língua portuguesa. *Blucher Open Access*.
- DE FRANÇA, M. T. R. (2019). A Desigual Distribuição da Fecundidade. *Finanças Públicas: Complexidade Ainda que Gasto Primário em Contenção*, 40.
- DE SOUZA, F. V. S., Correio, A. D. S. S., & Correio, B. A. O. R. (2021). Envelhecimento populacional, políticas públicas voltadas para o idoso e a pandemia da COVID-19: Alguns apontamentos possíveis. *Revista Longevidade*.
- HYNDMAN, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- LJUNG, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- LIMA CMA de O. Information about the new coronavirus disease (COVID-19). *Radiol Bras [Internet]*. 2020 Mar;53(2):V–I.

LUPPI, F., Arpino, B., & Rosina, A. (2020). The impact of COVID-19 on fertility plans in Italy, Germany, France, Spain, and the United Kingdom. *Demographic Research*, 43, 1399-1412.

MORETTIN, Pedro A., TOLOI, Clélia M. C. *Análise de Séries Temporais*, São Paulo: Edgard Blucher. 1981.

RAMOS, L. R., Veras, R. P., & Kalache, A. (1987). Envelhecimento populacional: uma realidade brasileira. *Revista de Saúde Pública*, 21, 211-224

SHAPIRO, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.

VASCONCELOS, A. M. N., & Gomes, M. M. F. (2012). Transição demográfica: a experiência brasileira. *Epidemiologia e Serviços de Saúde*, 21(4), 539