

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA-25P16

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Análise da Sazonalidade da Natalidade no Brasil”

Sávio Campos de Souza

Airlane Pereira Alencar

Fábio Kon

São Paulo, junho de 2025

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Análise da Sazonalidade da Natalidade no Brasil

PESQUISADOR: Fábio Kon

INSTITUIÇÃO: IME-USP

RESPONSÁVEIS PELA ANÁLISE: Sávio Campos de Souza
Airlane Pereira Alencar
Fábio Kon

REFERÊNCIA DESTE TRABALHO: SOUZA, S. C.; ALENCAR, A. P.; KON, F. Relatório de análise estatística sobre o projeto: “Análise da sazonalidade da natalidade no Brasil”. São Paulo, IME-USP, 2025. (RAE–CEA-25P16)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

SITE DATASUS. **SINASC-Sistema de Informação de Nascidos Vivos.** Disponível em <<https://datasus.saude.gov.br/transferencia-de-arquivos/>> Acesso em: 2 maio de 2025

SITE TABNET DATASUS. **SINASC-Sistema de Informação de Nascidos Vivos.** Disponível em <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinasc/cnv/nvuf.def>> Acesso em: 2 maio de 2025

LUBBA, C.H.; SETHI, S.S.; KNAUTE, P. (2019). *catch22: CAnonical Time-series CHaracteristics*. **Data Min Knowl Disc**, 33, 1821–1852.
<<https://doi.org/10.1007/s10618-019-00647-x>

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Office LTSC Standart 2021

TÉCNICAS ESTATÍSTICAS UTILIZADAS

03:010 – Análise Descritiva Unidimensional

06:120 – Análise de Conglomerados

11:010 – Séries Temporais

ÁREA DE APLICAÇÃO

14:050 – Demografia

RESUMO

Este trabalho tem como objetivo analisar a sazonalidade da natalidade no Brasil entre os anos de 2000 e 2023, utilizando dados do SINASC (Sistema de Informação sobre Nascidos Vivos), disponibilizados pelo DATASUS. Inicialmente, foi realizada uma análise descritiva das variáveis mais relevantes associadas ao nascimento no Brasil, como tipo de parto, peso ao nascer, idade da mãe, número de consultas pré-natal, entre outras. As análises descritivas dos dados indicam mudanças no perfil etário das mães com diminuição da maternidade antes dos 25 anos e aumento após os 30 anos, crescimento significativo do número de partos cesáreos e melhoria nos indicadores de saúde neonatal, como o escore de Apgar. Além disso, observou-se uma distribuição não uniforme dos nascimentos ao longo do ano, com picos entre os meses de fevereiro e junho. Para aprofundar a análise de similaridades e diferenças dos estados, foram aplicados os métodos de clusterização k-means e hierárquico com base em características extraídas das séries temporais mensais das taxas de natalidade por estado, utilizando o algoritmo Catch22. Posteriormente, para a previsão da taxa de natalidade foram ajustados o modelo de regressão linear com erro autoregressivo e o modelo ARIMA(0,1,0), que apresentaram queda média anual entre 0,25 e 0,28 pontos percentuais.

Sumário

1. Introdução	7
2. Objetivo	7
3. Descrição do estudo	7
4. Descrição das variáveis	8
5.1 Análise descritiva	9
5.2 Análise de agrupamento	13
6. Análise Inferencial	16
7. Conclusão	17
APÊNDICE A	19
APÊNDICE B	45

1. Introdução

O estudo da natalidade de um país é importante para compreender a evolução da densidade demográfica, para o planejamento do sistema previdenciário e para identificar as necessidades dos serviços de saúde, além de envolver diferentes áreas do conhecimento. O Brasil possui uma excelente base de dados nesse campo de estudo, sendo que em cada nascimento são documentados dados da mãe, da saúde do bebê, do local do nascimento, entre outros aspectos. Monitorar e analisar esses dados é um trabalho necessário para embasar a tomada de decisões nas políticas públicas. Este estudo se propõe a facilitar a visualização e interpretação descritiva dos dados e em um segundo momento, realizar análise de agrupamento e previsão. Realizar essas análises permite entender a similaridade e as diferenças dos estados e regiões do Brasil na área da natalidade.

2. Objetivo

O objetivo é realizar uma análise da sazonalidade da natalidade no Brasil com os dados disponíveis publicamente na plataforma DATASUS SINASC - Sistema de Informação de Nascidos Vivos. Essa análise servirá para elaborar um dashboard interativo da natalidade análogo ao dashboard da sazonalidade da mortalidade por covid-19 disponível em <<http://seasonality.interscity.org/>>>.

3. Descrição do estudo

Os dados foram obtidos do SINASC - Sistema de Informação de Nascidos Vivos por meio da área de transferência de arquivos e pelo TabNet. Pela área de transferência de

arquivos é baixada a série temporal de um ano especificado com todas as variáveis documentadas em arquivo .dbc, sendo necessária a descompactação. Através do TabNet foi possível escolher as séries anuais por variável específica gerando um arquivo .csv. O período de interesse compreende os anos de 2000 a 2023 considerando as variáveis descritas a seguir.

4. Descrição das variáveis

Cada observação da base de dados apresenta uma variável de identificação com o número da declaração de nascimento, sequencial por UF informante e por ano.

- **Tipo de parto:** 1 – vaginal, 2 – cesáreo, 9 – ignorado
- **Unidade Federativa:** código da UF que informou o registro.
- **Duração da gestação:** 9 – Ignorado, 1 – Menos de 22 semanas, 2 – 22 a 27 semanas, 3 – 28 a 31 semanas, 4 – 32 a 36 semanas, 5 – 37 a 41 semanas, 6 – 42 semanas e mais.
- **Tipo de gravidez:** 9 – Ignorado, 1 – Única, 2 – Dupla, 3 – Tripla e mais.
- **Número de consultas de pré-natal:** 1 – Nenhuma, 2 – de 1 a 3, 3 – de 4 a 6, 4 – 7 e mais, 9 – Ignorado.
- **Sexo do bebê:** 0 – Ignorado, 1 – Masculino, 2 – Feminino.
- **APGAR 1:** Apgar no primeiro minuto. Avalia-se a frequência cardíaca, respiração, tônus muscular, reflexos e cor da pele, atribuindo uma pontuação de 0 a 2 para cada critério, num total máximo de 10.
- **Peso ao nascer:** em gramas.
- **Raça/Cor:** 1 – Branca, 2 – Preta, 3 – Amarela, 4 – Parda, 5 – Indígena.
- **Data do nascimento:** Data do nascimento, no formato ddmmaaaa
- **Hora:** Hora do nascimento

- **Idade da mãe (anos)**

A partir das variáveis citadas, foram criadas as seguintes:

- **Número de nascidos vivos**
- **Taxa de natalidade a cada 1000 habitantes**
- **Número de nascidos vivos por dia**

5.1 Análise descritiva

A Figura A.1 apresenta a taxa de natalidade a cada 1000 habitantes para todos os estados do Brasil. É perceptível que ao passar dos anos há uma tendência de diminuição da taxa de nascimentos para todos os estados. A Figura A.2 apresenta os estados com as maiores taxas de natalidade a cada 1000 habitantes (Amapá, Acre, Roraima e Amazonas) e os com as menores taxas (Rio Grande do Sul, Minas Gerais, São Paulo e Paraná). Nota-se que o estado de Roraima apresenta um crescimento da taxa de natalidade nos anos 2018 e 2019. Possivelmente isso ocorreu devido à entrada de Venezuelanos nesse estado.

Na Figura A.3 pode-se observar que de modo geral de 2000 a 2023 a quantidade de bebês que nasceram com mais de 3000 gramas é próxima de 70%. Do total de bebês nascidos vivos nesse período, a Figura A.4 mostra o percentual de bebês nascidos com menos de 3000g. É perceptível que há um aumento percentual de bebês que nascem com menos de 3000g, em especial os que nascem entre 1500g a 2499g e entre 2500g a 2999g. Provavelmente esse aumento é devido a um maior desenvolvimento das tecnologias e procedimentos de cuidado dos bebês.

A Figura A.5 mostra que ao longo dos anos o parto por cesárea passou a ser mais frequente que o parto vaginal. Observa-se que entre os anos 2000 a 2003, aproximadamente 62% dos partos eram do tipo vaginal e no ano 2023, aproximadamente 60% dos partos eram do tipo cesáreo.

Na Figura A.6 pode-se observar que ao longo dos anos houve um aumento considerável da porcentagem de nascidos vivos com Apgar 1 entre 8 a 10, ou seja, que nasceram com

boa saúde, sem necessidade de intervenção médica imediata. Na Figura A.7 se percebe uma diminuição considerável da porcentagem de nascidos com Apgar 1 entre 6 e 7, saindo de 13% no ano 2000 para 8,75% em 2023. Para os que nasceram com Apgar 1 entre 3 a 5, indicando uma maior gravidez, também houve diminuição, saindo de 3,8% em 2000 para 2,3% no ano 2023, que é praticamente o mesmo valor desde 2012. Finalmente para os nascidos com Apgar 1 entre 0 e 2, no ano 2000 havia 1,3% e em 2023, 0,9%, que é praticamente o mesmo valor desde 2012.

Na Figura A.8 nota-se um grande aumento da porcentagem de nascimentos em que a mãe realizou 7 ou mais consultas pré-natal. Interessante perceber que essa porcentagem aumenta a cada ano, saindo de 45% em 2001, com um acréscimo percentual aproximado de 1,45% a cada ano, chegando à porcentagem de 77,5% em 2023. Esse maior cuidado no pré-natal pode indicar o aumento do número de nascimentos com ótimo Apgar 1 ou mesmo o pequeno aumento do número de bebês nascidos vivos abaixo do peso.

Na Figura A.9, sobre a duração da gestação, vemos, no período de 2020 a 2023, que aproximadamente 85,5% dos bebês nascem com 37 a 41 semanas de gestação, a cada ano. Nos anos anteriores a 2011, havia aproximadamente 0,75% de gestações entre 42 ou mais semanas, a partir de 2011 esse valor passou para aproximadamente 3% estando a 2,3% em 2023. Na Figura A.10, sobre a duração da gestação de bebês prematuros (nascidos vivos com menos de 37 semanas de gestação), notamos que entre os anos 2010 e 2012 houve um aumento de 5% para 10% da porcentagem de nascimentos com gestação de 32 a 36 semanas. Até 2010 essa porcentagem era próxima de 5% e depois de 2012 a porcentagem permaneceu próxima a 10%.

A Figura A.11 mostra a idade das mães ao decorrer dos anos. Percebemos que a cada ano há uma redução na porcentagem de mães com menos de 19 anos e um aumento expressivo na porcentagem de mães com mais de 30 anos. Comparando os anos inicial e final houve uma redução da porcentagem praticamente pela metade: no ano 2000 havia 23,4% de mães com menos de 19 anos e em 2023 esse valor passou a ser 11,95%. Para mães entre 20 e 24 anos, a redução foi de 31,14% para 23,71%. Para as mães entre 25 e 29 anos, a porcentagem passou de 22,46% para 25,43%. A porcentagem de mães de 30 a 34 anos passou de 13,83% para 20,96%. As mães entre 35 e 39 passaram

de 6,7% em 2000 para 13,61% em 2023. As mães entre 40 e 44 passaram de 1,74% para 4,06%. A porcentagem de mães com 45 anos ou mais cresceu de 0,014 no ano 2000 para 0,026 em 2023. Os dados com idade ignorada diminuíram, passando de 0,58% para 0,0013, respectivamente em 2000 e 2023.

Na Figura A.12, sobre a porcentagem de nascimentos por sexo, fica evidente que ao longo dos anos há uma porcentagem constante de aproximadamente 48,8% de nascimentos do sexo feminino e 51,2% do sexo masculino. Segundo o censo demográfico de 2022, a porcentagem de homens e mulheres no total é de 48,5% e 51,5%, respectivamente. Isso se deve ao fato de haver uma alta mortalidade de homens jovens e das mulheres buscarem mais frequentemente os estabelecimentos de saúde.

Na Figura A.13 vemos que 2% dos nascimentos são de gestação dupla e 98% são de gestação única. Em 2023 houve 1063 gestações triplas.

O número de nascimentos por cor e raça, apresentado na Figura A.14 mostra que em 2000 havia uma grande quantidade (400000) de valores ignorados, que ao longo dos anos foi diminuindo chegando a 45982 em 2023. Se percebe um aumento da declaração de nascimentos de bebês de cor parda e desde 2006 houve mais declarações da cor parda do que da cor branca.

As Figuras A.15 e A.16 apresentam respectivamente o gráfico da média diária de nascimentos por mês e o gráfico da sua porcentagem. Para o cálculo da média diária de nascimentos por mês dividimos o total de nascimentos de cada mês pelo número de dias daquele mês. Percebemos que as curvas são similares ao longo dos anos, com maior media diária e porcentagem de nascimentos entre fevereiro e junho.

Nas Figuras A.17 e A.18 temos respectivamente os gráficos do número de nascimentos por dia da semana e horário, ambos por tipo de parto, no ano de 2023. A Figura A.17 mostra que os partos cesáreos ocorrem predominantemente nos dias úteis da semana e os partos de tipo vaginal são distribuídos de forma uniforme em cada dia da semana. Na Figura A.18 vemos que os partos cesáreos ocorrem na maior parte em horário comercial, enquanto o parto vaginal é distribuído praticamente de forma uniforme entre todos os horários do dia.

Na Figura A.19, sobre a taxa de natalidade a cada 1000 habitantes por UF, nos anos 2000 e 2023, percebemos em todos os estados uma redução nas taxas de natalidade do ano 2000 para o ano 2023.

Na Figura A.20, pelo mapa do Brasil observamos que os estados mais ao Norte e Oeste possuem maiores taxas de natalidade.

A Tabela B.1 apresenta a distribuição percentual do peso ao nascer por duração da gestação. Percebe-se que para bebês cuja duração da gestação foi menor do que 22 semanas, 50,64% nasceram com menos de 500g. Para gestações de 22 a 27 semanas, em 61,06% os bebês nasceram com peso entre 500 e 999g. Para 28 a 31 semanas de gestação, 69,61% dos bebês nasceram com peso entre 1500g e 2499g. Para gestações entre 32 a 36 semanas, 71,43% dos bebês nasceram com peso entre 1500 e 2999g. Para os bebês com 37 a 41 semanas de gestação, 67,67% nasceram com peso entre 3000 e 3999g. Para aqueles com 42 semanas de gestação ou mais, 71,40% nasceram com peso entre 3000g e 3999g. Assim vemos que, de modo geral, quando a duração da gestação é maior, o peso no recém nascido também é maior.

Na Figura A. 21, nota-se que houve uma redução da porcentagem de mães entre 15 a 19 anos em todas as regiões do Brasil, no norte, nordeste, sudeste, sul e centro oeste, a porcentagem passou de 29, 25, 20, 20, 25 no ano 2000 para 18, 14, 9, 9, 11, respectivamente. Houve uma redução da porcentagem também do número de mães entre 20 e 24 anos. Por outro lado, percebeu-se um aumento da porcentagem de mães com mais de 25 anos no ano de 2023.

Na Figura A.22, da porcentagem de consultas pré-natal por região, percebemos um grande aumento na porcentagem de mães que tiveram 7 ou mais consultas pré-natal, em todas as regiões, passando das porcentagens de 25, 31, 53, 52, 50 nas regiões N, NE, SE, S, CO, respectivamente, no ano 2000 para as porcentagem de 62, 75, 81, 84, 76 respectivamente no ano 2023.

Na Figura A.23 percebemos que em 2023 os estados com maior porcentagem de parto do tipo cesárea são RO e GO com aproximadamente 70% e os estados com menor

porcentagem são RR com aproximadamente 40% e AM e AP com aproximadamente 45%. Nas Figuras A.24 a A.28 temos a porcentagem do tipo de parto por região, do ano 2000 ao ano 2023, em que todas as regiões apresentam um aumento anual da porcentagem de parto do tipo cesárea do ano 2000 ao ano 2013. E para todas as regiões, do ano 2014 a 2023 a porcentagem de parto do tipo cesárea permaneceu aproximadamente a mesma. Na Figura A.29, sobre os partos por hora do dia e por região no ano de 2023, percebemos que a quantidade de partos vaginal é praticamente constante a cada horário do dia. No mesmo gráfico, para os partos do tipo cesárea, em todas as regiões percebe-se uma maior densidade no horário comercial e na região sudeste, em particular, há mais nascimentos no período da manhã.

5.2 Análise de agrupamento

O algoritmo Catch22 (Canonical Time Series Characteristics) vetoriza séries temporais considerando 22 principais características (variáveis) nas áreas de distribuição, estatísticas simples de séries temporais, autocorrelação linear e não linear, diferenças sucessivas e análise de flutuação e outras duas características da matriz de transição e periodicidade de Wang. A descrição de cada variável e a construção do algoritmo está descrita no artigo em LUBBA et al (2019).

Foi aplicado o algoritmo Catch22 para a série temporal das taxas (a cada 1000 habitantes) mensais de natalidade de cada estado do Brasil. Para obter as taxas mensais de natalidade de cada estado do Brasil foi preciso fazer a contagem da natalidade por mês e dividir pelo número estimado da população a cada mês e multiplicar por 1000. A população foi obtida por interpolação a partir da população divulgada nos censos.

A partir dos valores das 22 características obtidas pelo Catch22, para cada estado (Tabela B.2), foi realizada uma análise de componentes principais padronizando cada variável, já que nessa tabela percebe-se uma diferença de escala entre as variáveis, por exemplo CO_f1eaca possui valores que variam de 0 a 60, enquanto em DN_histogramMode_5, os valores estão entre -0,8 e 0,8. As componentes principais 1,

2, 3 e 4 explicam respectivamente 33,7%, 15,6%, 11,7% e 9,7% da variância total, somando ao total uma explicação de 70,7%. Na Tabela B.3 temos a contribuição de cada variável do catch22 para cada componente principal. Nela percebe-se que os maiores pesos (loadings) em módulo das variáveis do Catch22 que compõem a primeira componente principal são FC_LocalSimple_mean3_stderr com 0,35, CO_HistogramAMI_even_2_5 com -0,342, FC_LocalSimple_mean1_tauresrat com 0,309. Cada uma dessas variáveis significam respectivamente o erro médio de uma média móvel de 3 elementos, auto informação mútua e mudança no comprimento da correlação após a diferenciação iterativa. Na Tabela B.4 temos os escores dos componentes principais valorados para cada unidade federativa.

Em seguida, a partir das quatro componentes principais foi realizada uma análise de agrupamento por k-means e por método hierárquico. Para a definição do número k de grupos foi utilizado o “método do cotovelo” cujo gráfico é apresentado na Figura A.30. A partir dessa figura, escolhemos k igual a 5.

Na Figura A.31, temos as taxas mensais de natalidade (por 1000 hab) das UF agrupadas por k-means. O estado do Amazonas foi agrupado sozinho no grupo 4. RR e RO formam o grupo 3; ES, MS, MT, SC o grupo 1; AC, AL, AP, DF, MA, PA, PI, TO foram reunidos no grupo 5 e o restante dos estados no grupo 2. É possível visualizar esses clusters no mapa do Brasil na Figura A.32. No gráfico *biplot* da Figura A.33, que considera as duas primeiras componentes principais, vemos uma boa discriminação dos clusters 1 e 4, sendo que a CP1 discrimina bem o grupo 1 e a CP2 discrimina bem o estado do AM. Considerando apenas as CP1 e CP2, que somam 49,3% da variância explicada, não há uma tão boa discriminação entre os clusters 2 e 5, porque há estados como TO e AL do cluster 5 que estão muito próximos do cluster 2. O cluster 3 dos estados de RR e RO estão discriminados pela CP1, mas ainda não muito distantes do cluster 2. Assim vemos que é preciso utilizar os CP3 e CP4 para explicar com maior qualidade a discriminação dos clusters 2 e 5.

Para a análise de agrupamento utilizamos também o método hierárquico ward, que nos permite visualizar pelo gráfico heatmap quais componentes principais são decisivas para a separação de cada grupo. Na Figura A.34 vemos o dendograma obtido pelo método

hierárquico dividindo pela linha pontilhada vermelha 5 grupos. O método hierárquico classificou dois clusters de modo diverso do que o k-means: o cluster que tinha os estados AC, AL, AP, DF, MA, PA, PI, TO no k-means passou a ter AC, DF, MA, PA, PI apenas no método hierárquico, sendo que os estados que saíram daquele grupo foram para o cluster que reunia maior quantidade de estados, como podemos ver na Figura A.35.

Analizando a Figura A.36 do gráfico heatmap dos scores das componentes principais para o método hierárquico percebemos que o PC1 discrimina bem o cluster 4 com os estados do ES, SC, MS e MT, que se apresenta com tons vermelhos, ou seja, escores entre 4 e 5. O cluster 3, do estado do Amazonas, é explicado pelo PC2, pois apresenta um escore 6, bem vermelho, diferente de todos os outros estados. O cluster 5 é discriminado com maior eficiência pelo PC3 apresentando escores próximos de 4, diferentemente de todos os outros estados. O cluster 1 é discriminado parte pelo PC1 com escores negativos de cor azul (sendo que o DF possui um escore de -6, o PI de -4 e o AC, MA e PA de -2) e parte pelo PC2 que apresenta valores um pouco mais avermelhados comparado com os outros estados, exceção do DF que é um pouco azul, indicando que essa discriminação não é tão explícita. O cluster 2, com todos os outros estados, apresenta valores muito próximos do escore neutro em cada componente principal.

6. Análise inferencial

Para realizarmos a previsão da taxa de natalidade do Brasil inicialmente ajustamos um modelo de regressão linear como apresenta a Figura A.37, porque a série da taxa de natalidade aparenta possuir um decréscimo linear a cada ano. O ajuste do modelo de regressão linear é dado por $Taxa_t = \beta_0 + \beta_1 t + e_t$, em que “t” representa o ano, com $t = 1, \dots, 23$. No entanto, o ajuste do modelo linear apresentou resíduos correlacionados, conforme a Figura A.38. Na Figura A.39, percebemos que a primeira correlação para o lag 1 é significativa e maior que 0,5. Esse lag representa a autocorrelação entre o resíduo do ano t e do ano $t-1$. Por isso, modelamos o erro como sendo autorregressivo em que $e_t = \phi e_{t-1} + a_t$, $a_t \sim N(0, \sigma^2)$, ou seja, o erro do ano t é obtido pelo erro do ano anterior vezes um fator ϕ mais um erro aleatório a_t com distribuição normal de média 0 e variância σ^2 . Fazendo o ajuste desse modelo pelo método de máxima verossimilhança usando o comando Arima(1,0,0) da library(forecast) do pacote R, que utiliza apenas a parte autoregressiva de defasagem (lag) 1, o modelo passa a ter resíduos aleatórios com distribuição normal, como mostram as Figuras A.40 e A.41, respectivamente. O modelo deixou de possuir autocorrelação significativa, como mostra a Figura A.42. Na Figura A.43 vemos que o modelo está melhor ajustado comparado à regressão linear sem erro autoregressivo. Na Tabela B.5 temos os resultados do ajuste do modelo: o valor do coeficiente AR1 igual a 0,74 indica que o erro do ano $t-1$ possui forte influência no erro do ano t , ou seja, o ajuste com erro autorregressivo foi adequado; o coeficiente x reg indica que a cada ano que passa espera-se uma redução média de 0,2498 na taxa de natalidade; o intercepto de 18,54 indica a taxa de natalidade inicial (2000) considerada pelo modelo. Conforme vemos na previsão da taxa de natalidade para 2024, 2025 e 2026 na Figura A.44, espera-se que a cada 4 anos haja uma queda de 1 ponto percentual na taxa de natalidade.

Outra forma de realizar a previsão é utilizando o modelo ARIMA. É perceptível que há uma tendência negativa na série das taxas, então, conforme a Figura A.45, notamos que o cálculo de uma diferença na série já a tornou estacionária. Realizar uma diferença

significa escrever o modelo na forma $Taxa_t - Taxa_{t-1} = c + e_t$ em que $e_t \sim N(0, \sigma^2)$. A Figura A.45 mostra também que, no ano de 2016, houve uma grande queda da diferença da taxa, indicada provavelmente pela epidemia de zika vírus e pela crise econômica do País.

Ajustando o modelo ARIMA(0, 1, 0) vemos na Figura A.46, que não presença autocorrelação significativa. Além disso, os resíduos são aleatórios e com distribuição normal, ver Figura A.47, indicando um bom ajuste. Os resultados do ajuste do modelo são apresentados na Tabela B.6. O coeficiente drift com valor -0,2846 indica que há uma redução média de 0,2846 da taxa de natalidade a cada ano. Dado o erro padrão (s.e) de 0,0712 podemos fazer o teste t: vamos definir a hipótese nula do coeficiente drift ser igual a 0, então $t = (-0,2846 - 0)/0,0712 = -3,9971$. Como temos 23 anos observados e uma variável no modelo, temos $23 - 1 = 22$ graus de Liberdade (df), obtemos então da tabela t-student o valor-p < 0,001, ou seja, há fortes evidências que o coeficiente drift é diferente de 0. A série ajustada é apresentada na Figura A.48.

Comparando o modelo ARIMA(0,1,0) com o modelo de regressão linear com erro ARIMA(1,0,0), o ARIMA(0,1,0) apresentou valores AIC, AICc e BIC um pouco menores, sendo então preferível em relação ao outro modelo, mas ambos apresentaram bons ajustes. Na Figura A.49 vemos a previsão do modelo ARIMA(0,1,0) que possui um decrescimento levemente maior.

7. Conclusão

O presente estudo permitiu uma análise abrangente da natalidade no Brasil ao longo do período de 2000 a 2023, utilizando dados do SINASC/DATASUS. Por meio de análises descritivas, observou-se uma tendência de queda contínua na taxa de natalidade em praticamente todas as unidades federativas, acompanhada de mudanças importantes nos perfis maternos, como o aumento da idade das mães e o crescimento no número de consultas pré-natal.

A análise descritiva da série temporal revelou sazonalidades, com maiores médias diárias de nascimento entre os meses de fevereiro e junho. Foram destacados padrões relevantes como o aumento da prevalência de partos cesáreos, especialmente

concentrados em dias úteis e horários comerciais, e melhorias nos indicadores de saúde neonatal, como a elevação da proporção de recém-nascidos com Apgar 1 entre 8 e 10.

A aplicação de métodos de agrupamento k-means e hierárquico com base nas características das séries temporais mensais de natalidade por estado permitiu identificar cinco grupos distintos. Três desses clusters foram bem discriminados por ambos os métodos, sendo eles: AM; RR e RO; ES, MS, MT e SC. Para os outros estados houve uma divergência entre os dois métodos alocando AL, AP e TO em clusters diferentes. De modo geral, os clusters ajustaram estados conectados geograficamente. No entanto, para extrair mais informações sobre as similaridades e diferenças entre os estados é necessário comparar variáveis descritivas dos estados. Possivelmente um caminho a ser seguido seja comparar a descrição de diferentes variáveis dos estados de cada cluster. Desse modo, as informações serão mais consistentes para subsidiar políticas públicas regionais ajustadas às realidades locais.

Por fim, a modelagem estatística para previsão da taxa de natalidade, tanto via regressão linear com erro autoregressivo quanto por meio do modelo ARIMA(0,1,0), indicou uma queda média anual entre 0,25 e 0,28 pontos percentuais.

APÊNDICE A

Figuras

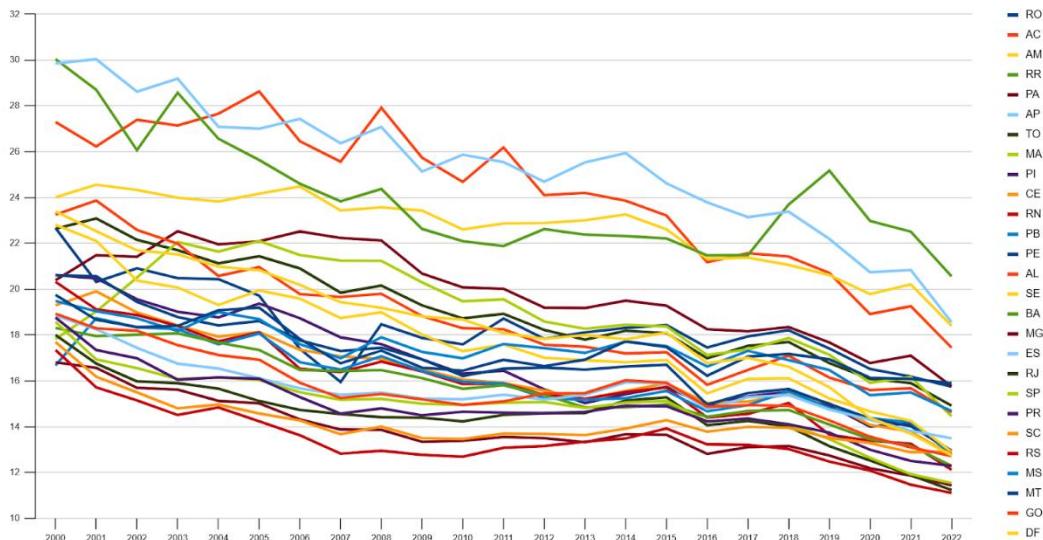


Figura A.1 Taxa de natalidade a cada 1000 habitantes segundo unidade federativa e ano

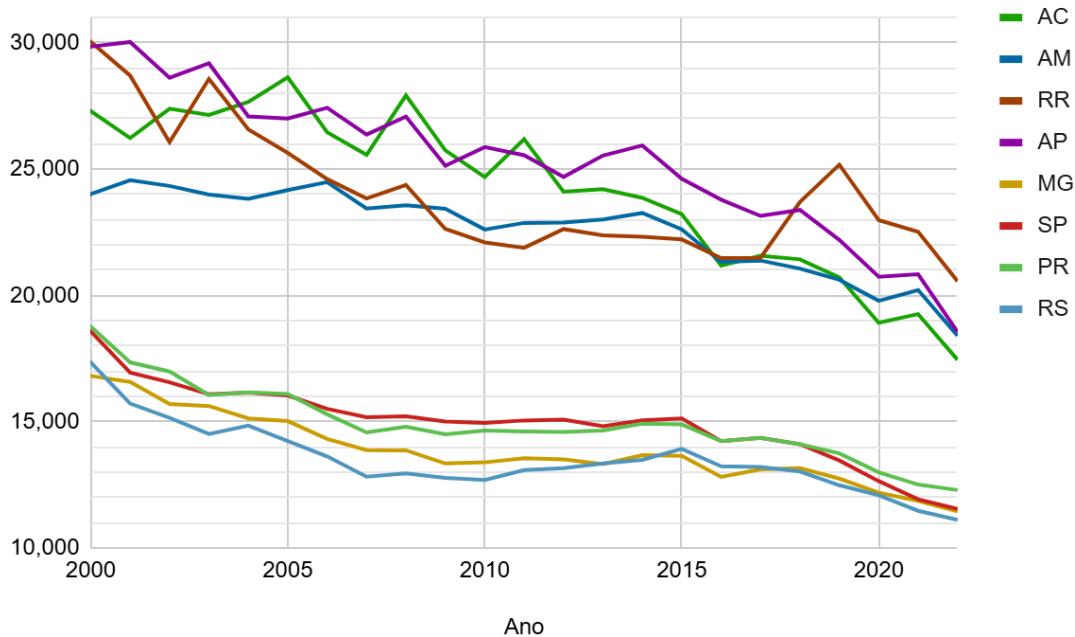
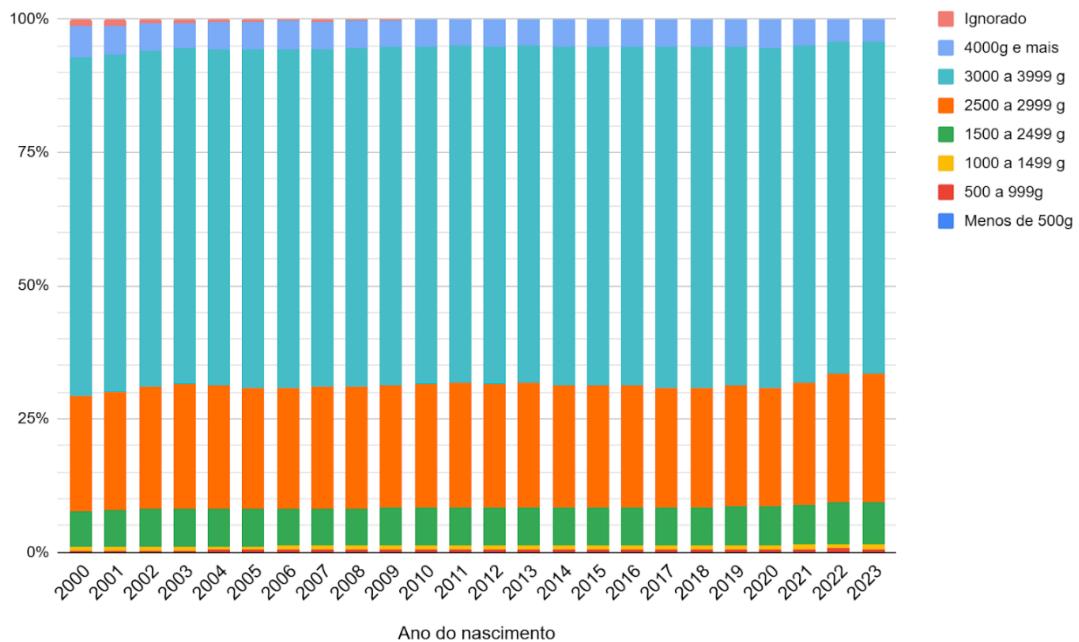
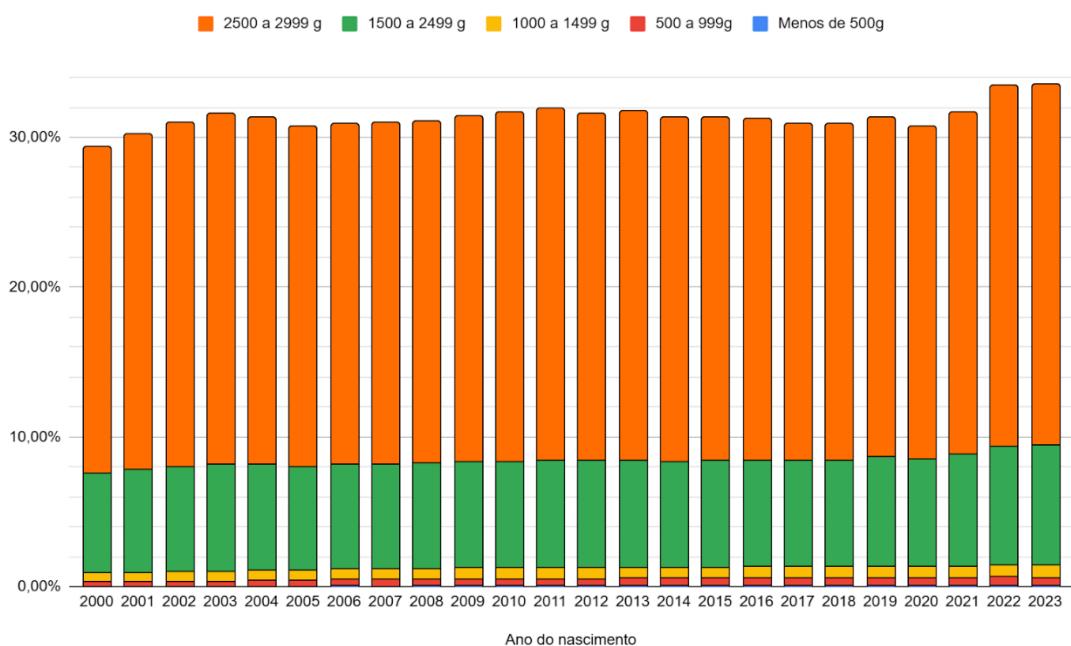


Figura A.2 Taxa de natalidade a cada 1000 habitantes dos estados com as maiores e menores taxas por ano

**Figura A.3** Peso ao nascer no Brasil**Figura A.4** Peso ao nascer de bebês de baixo peso no Brasil

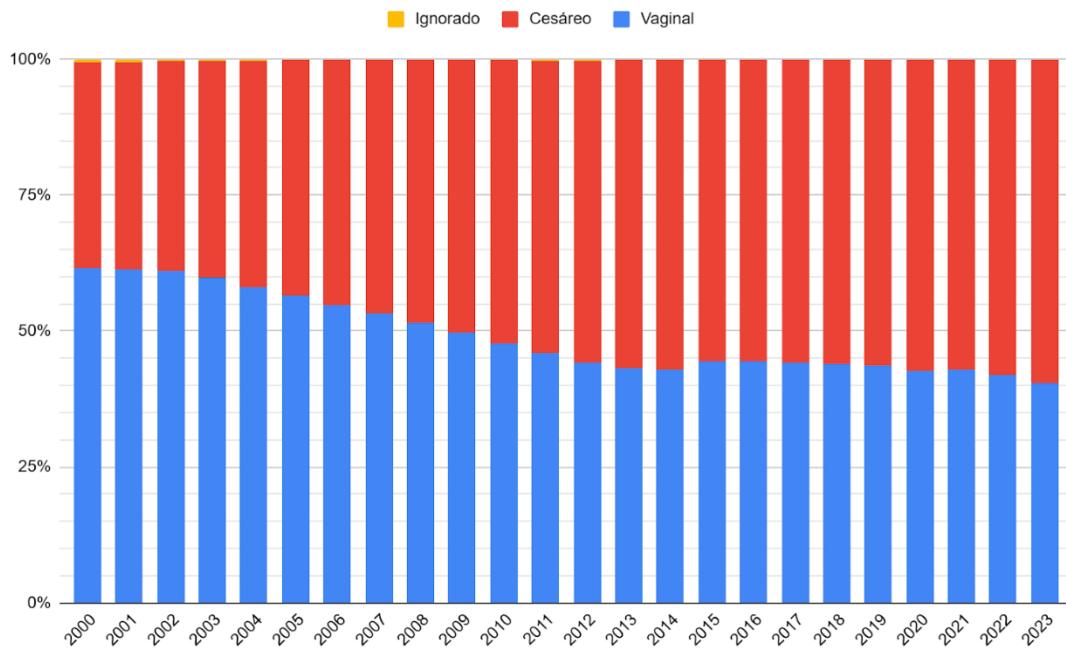


Figura A.5 Tipo de parto no Brasil

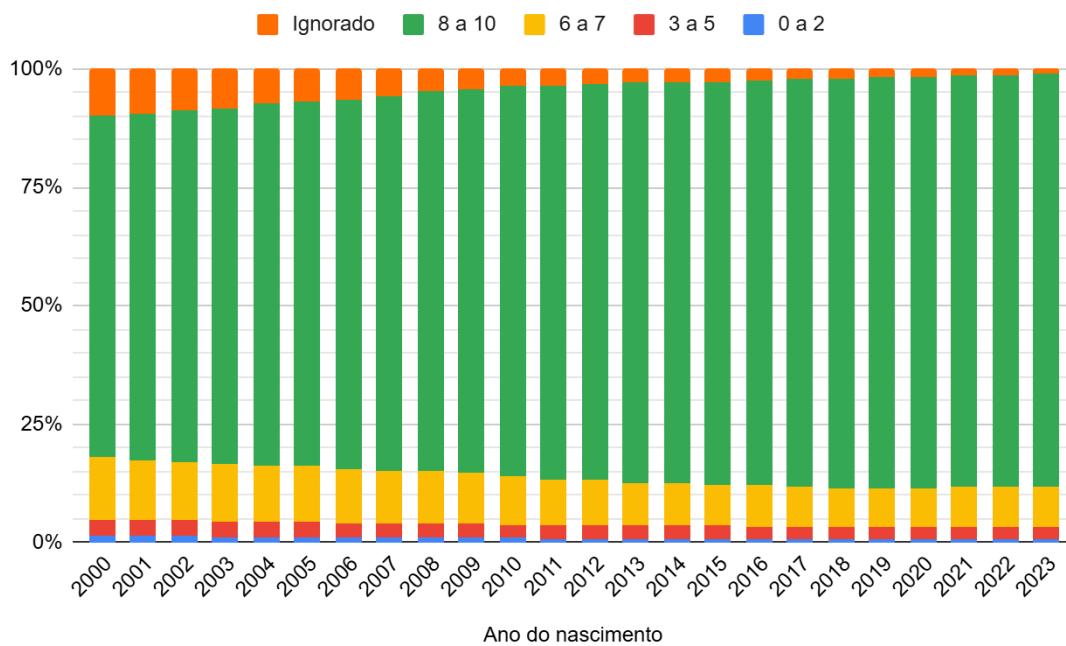
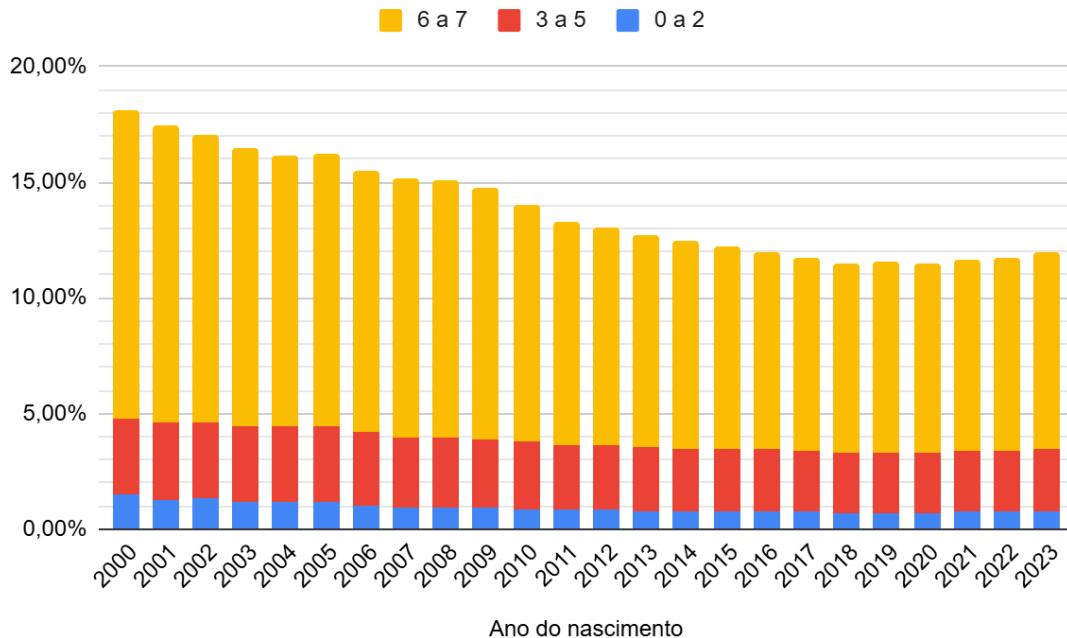
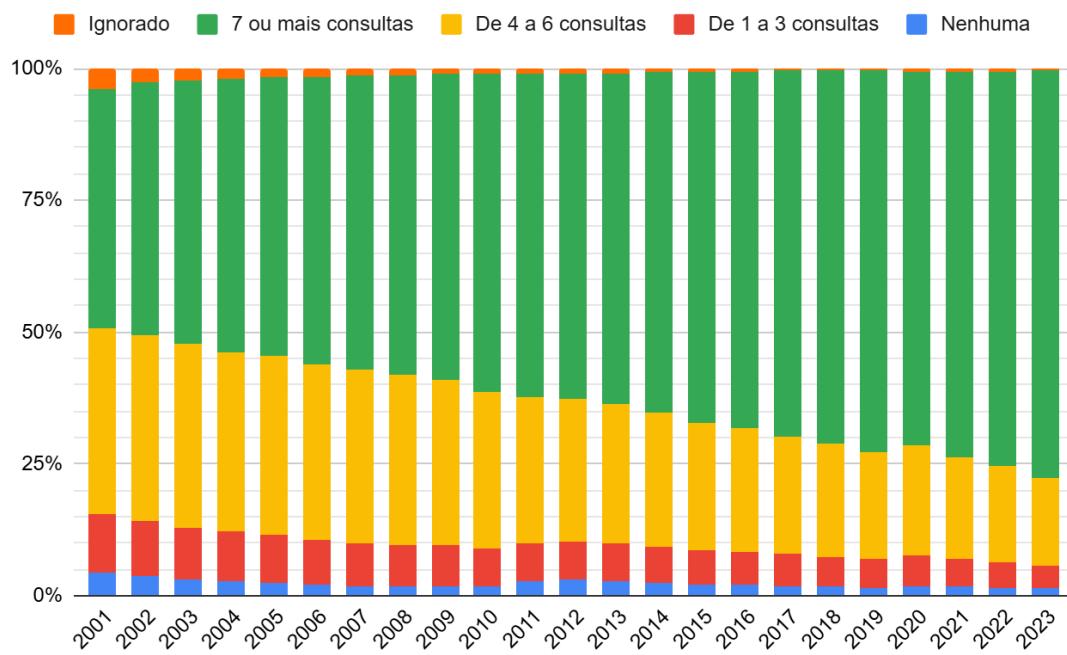


Figura A.6 Apgar 1 no Brasil

**Figura A.7** Apgar 1 – valores de 0 a 7**Figura A.8** Consultas pré-natal no Brasil

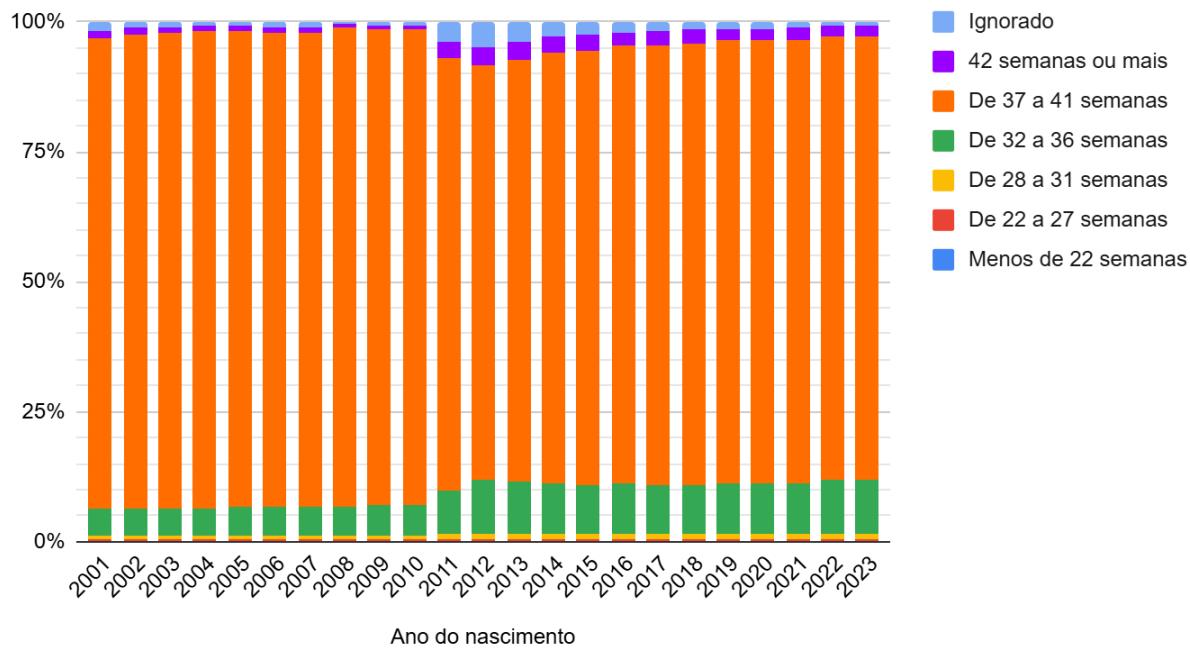


Figura A.9 Duração da gestação no Brasil

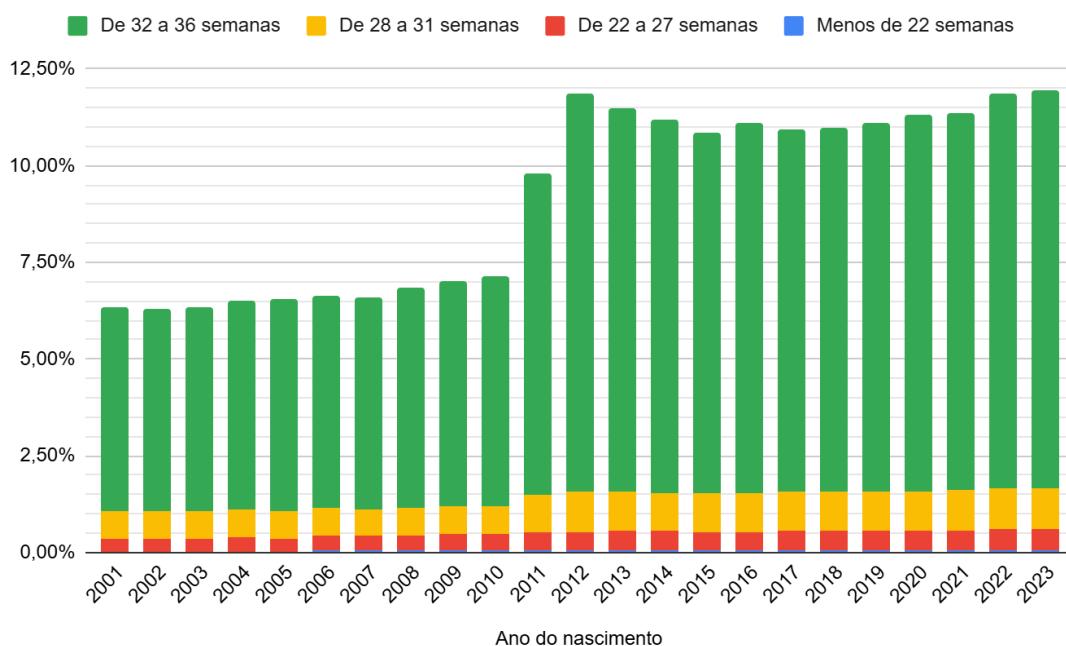


Figura A.10 Duração da gestação de bebês prematuros

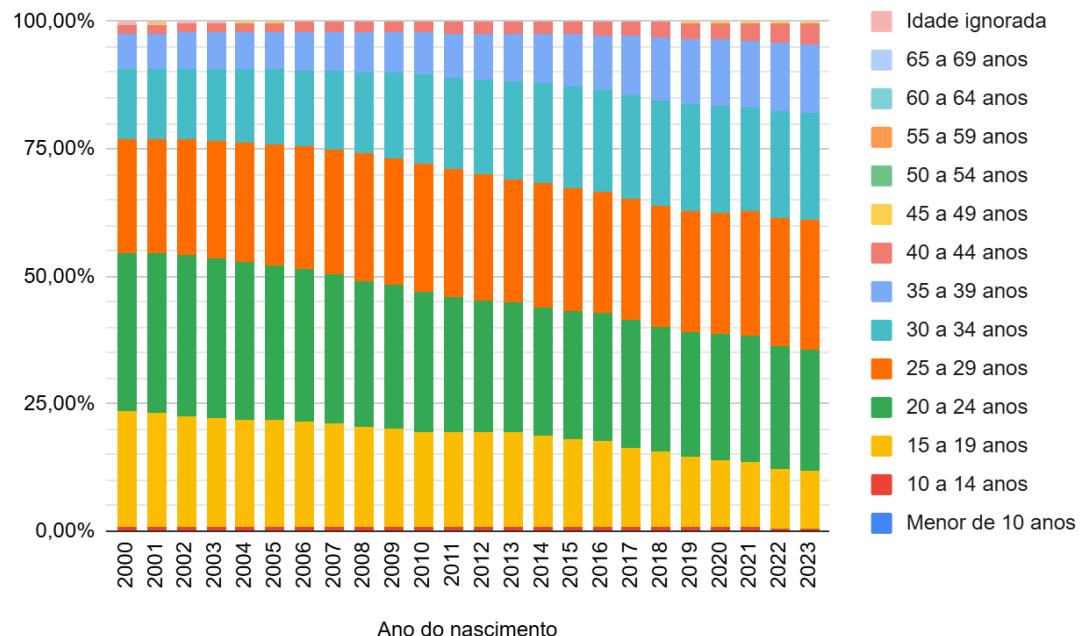


Figura A.11 Idade da mãe no Brasil

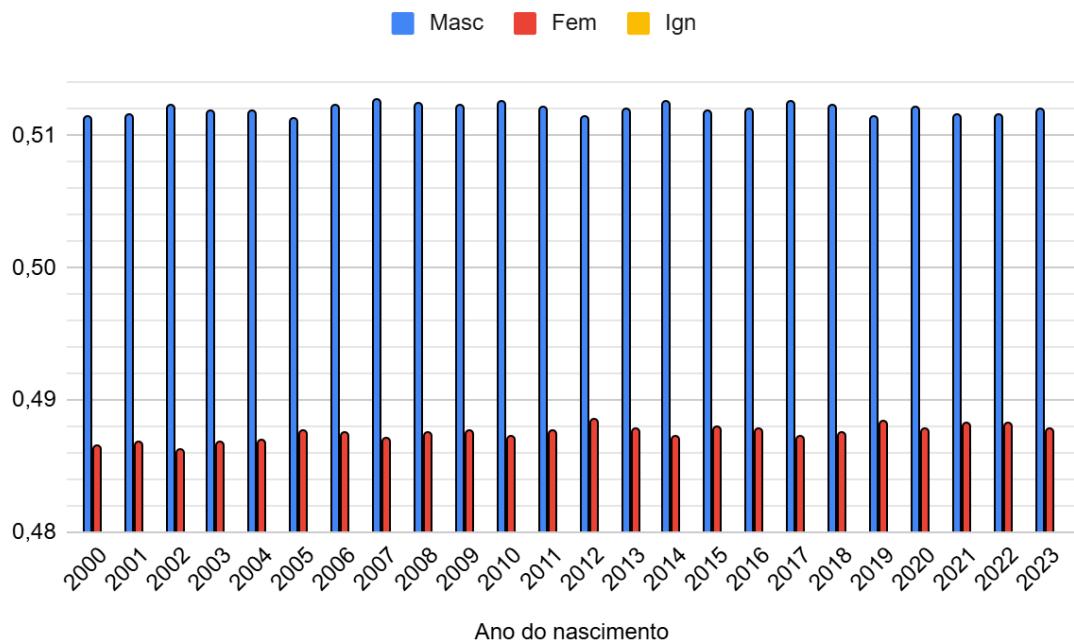


Figura A.12 Sexo do bebê no Brasil

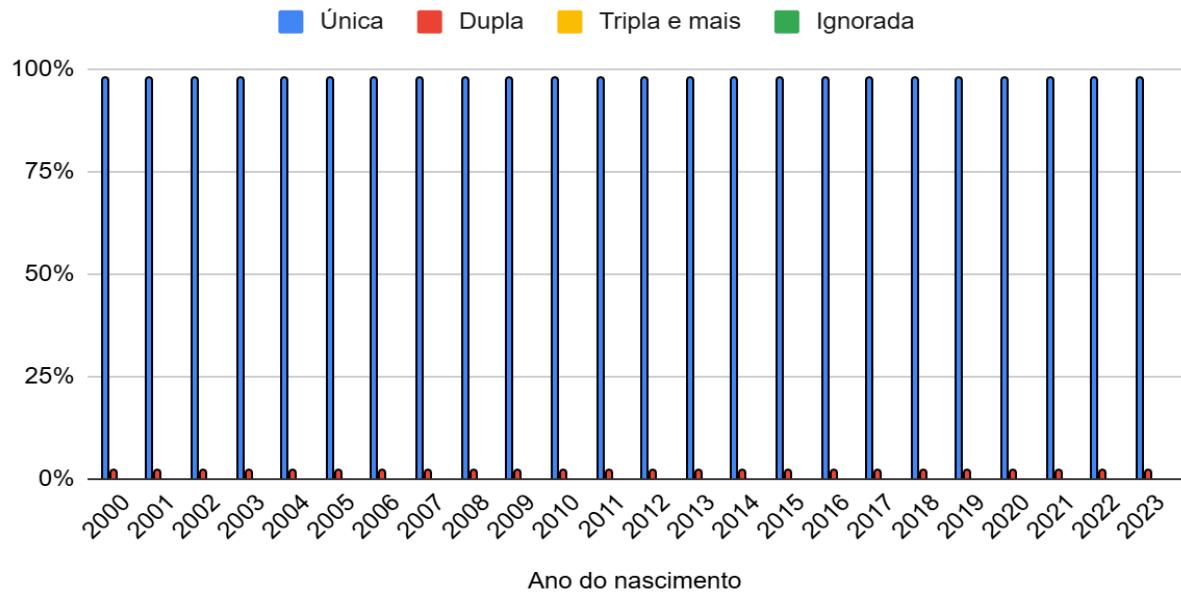


Figura A.13 Tipo de gravidez no Brasil

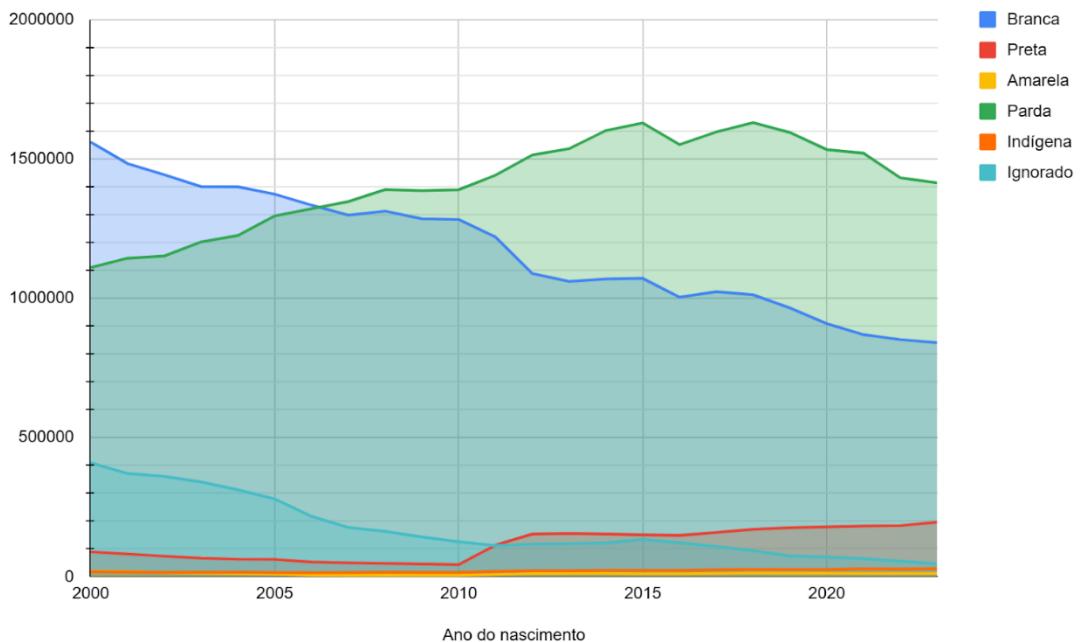


Figura A.14 Cor/raça dos recém-nascidos no Brasil

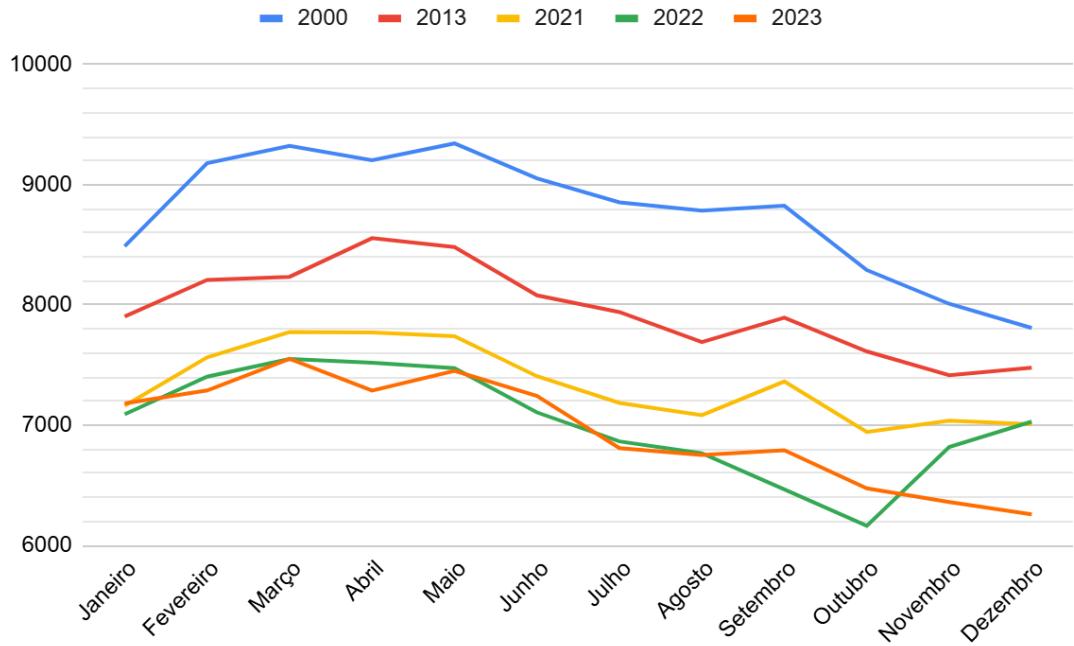


Figura A.15 Média diária de nascimentos por mês no Brasil

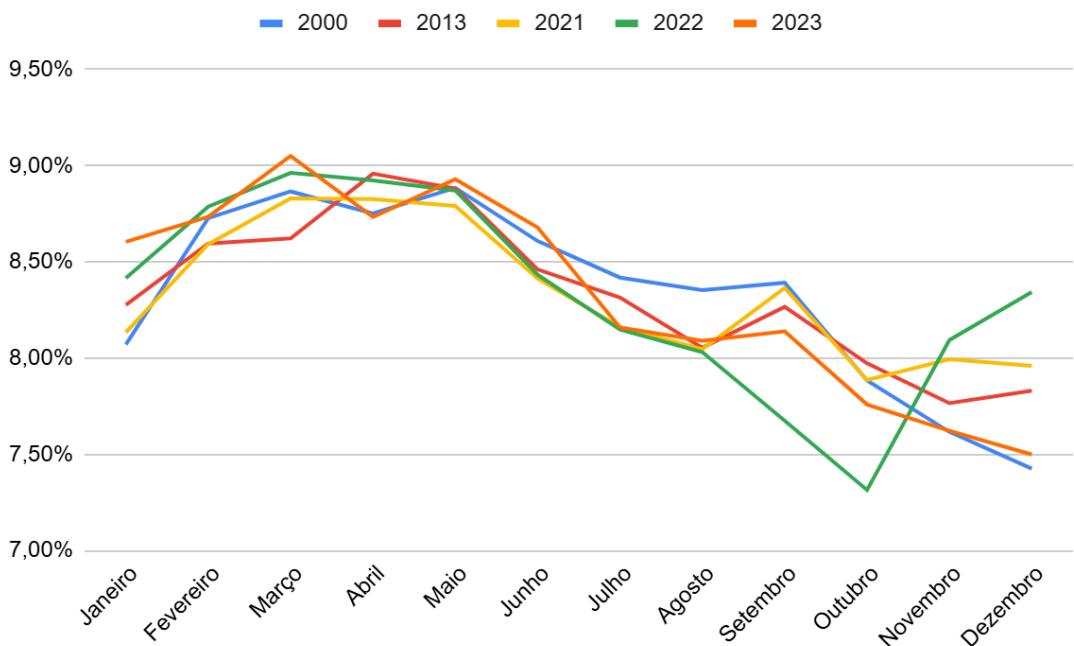


Figura A.16 Média diária de nascimentos por mês no Brasil

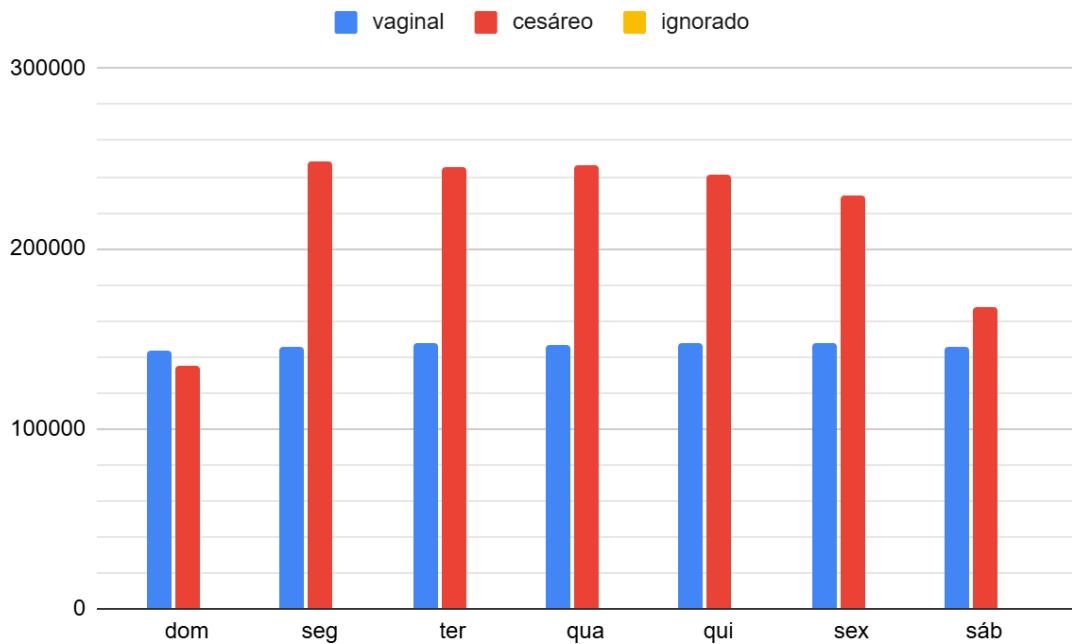


Figura A.17 Nascimentos por dia da semana e Tipo de parto, no ano 2023 no Brasil

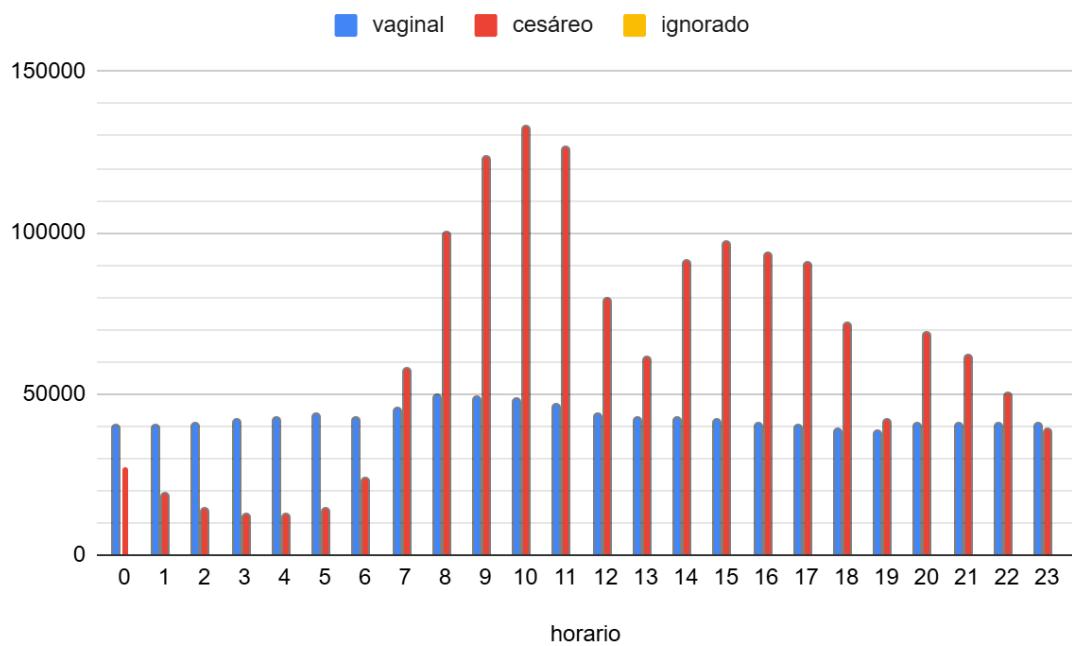


Figura A.18 Nascimentos por horário do parto e Tipo de parto, no ano de 2023 no Brasil

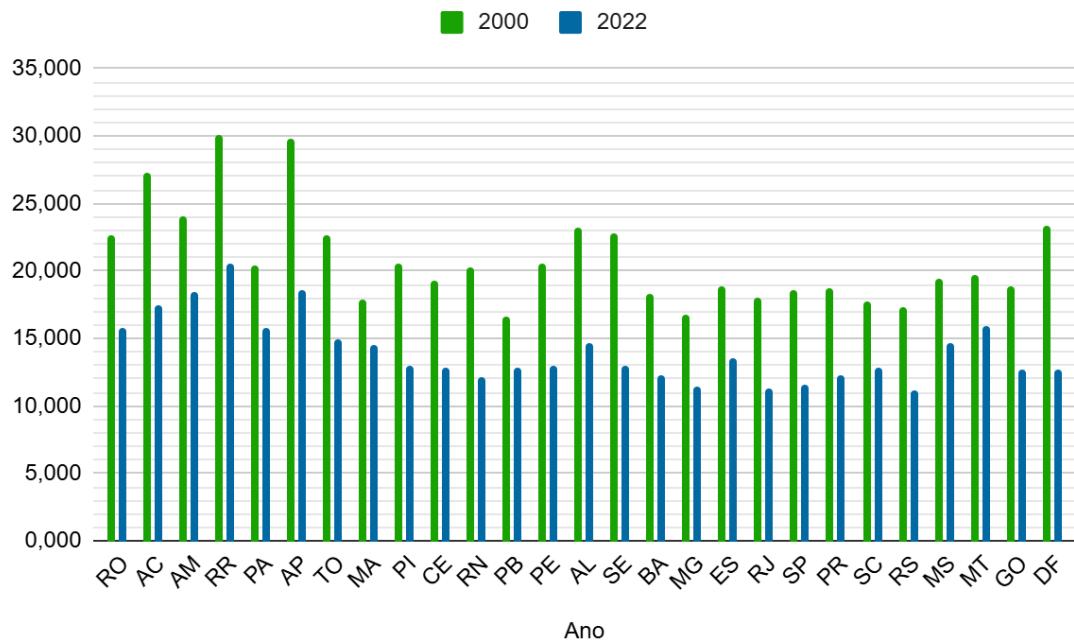


Figura A.19 Taxa de natalidade a cada 1000 habitantes por UF nos anos 2000 e 2022

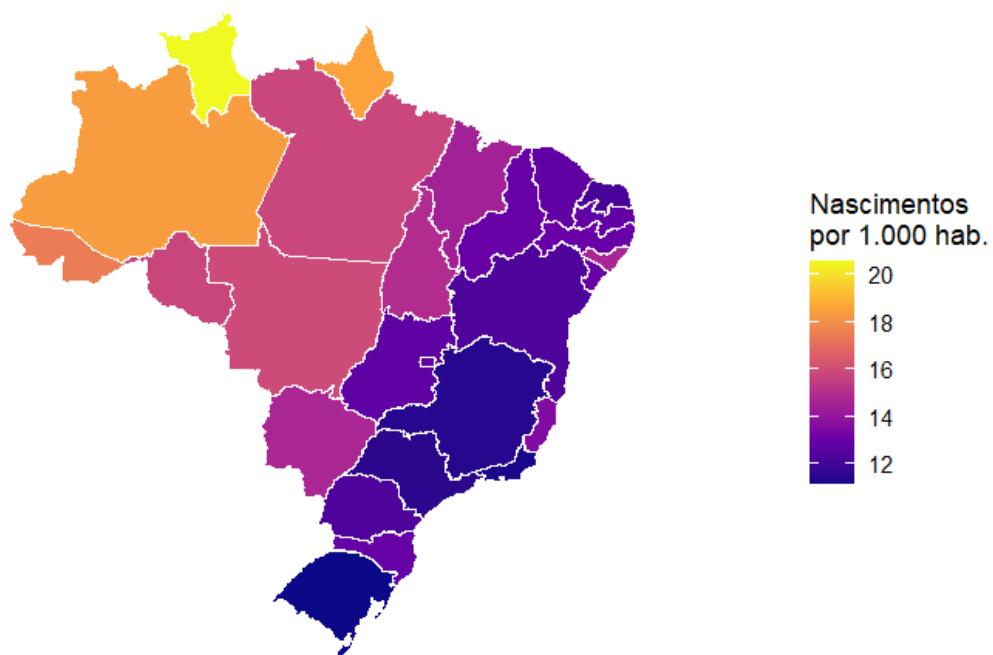


Figura A.20 Taxa de natalidade a cada 1000 habitantes por UF em 2022

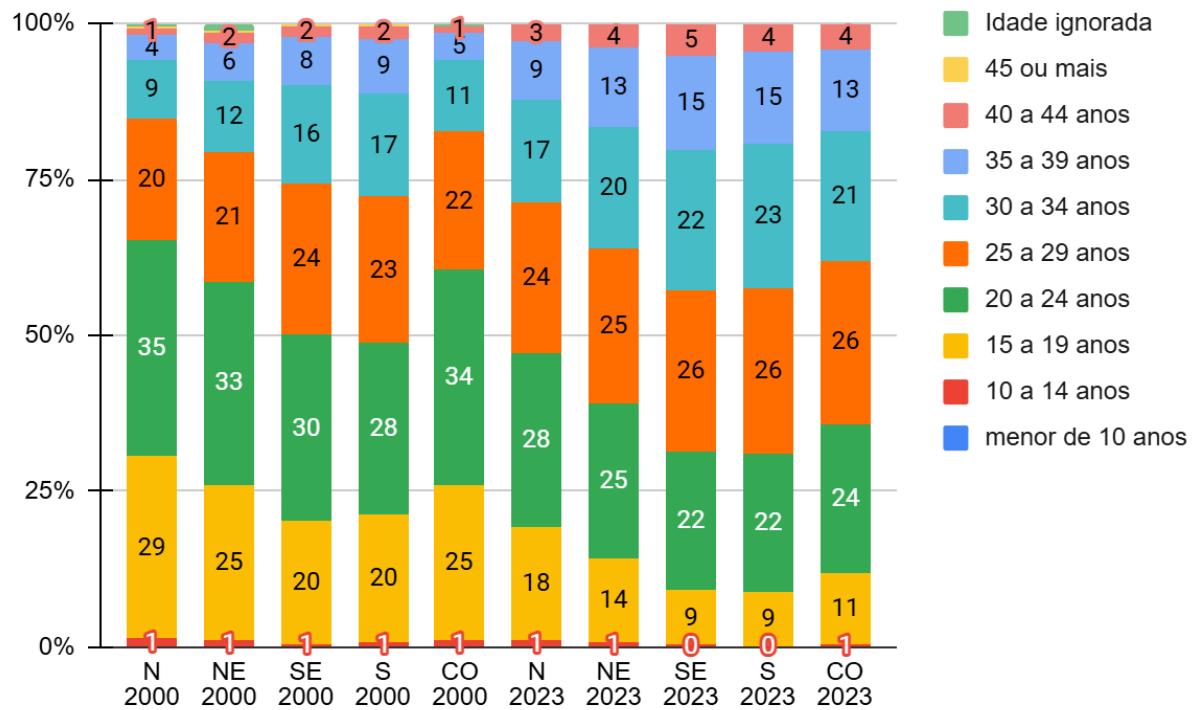


Figura A.21 Idade da mãe por região nos anos 2000 e 2023

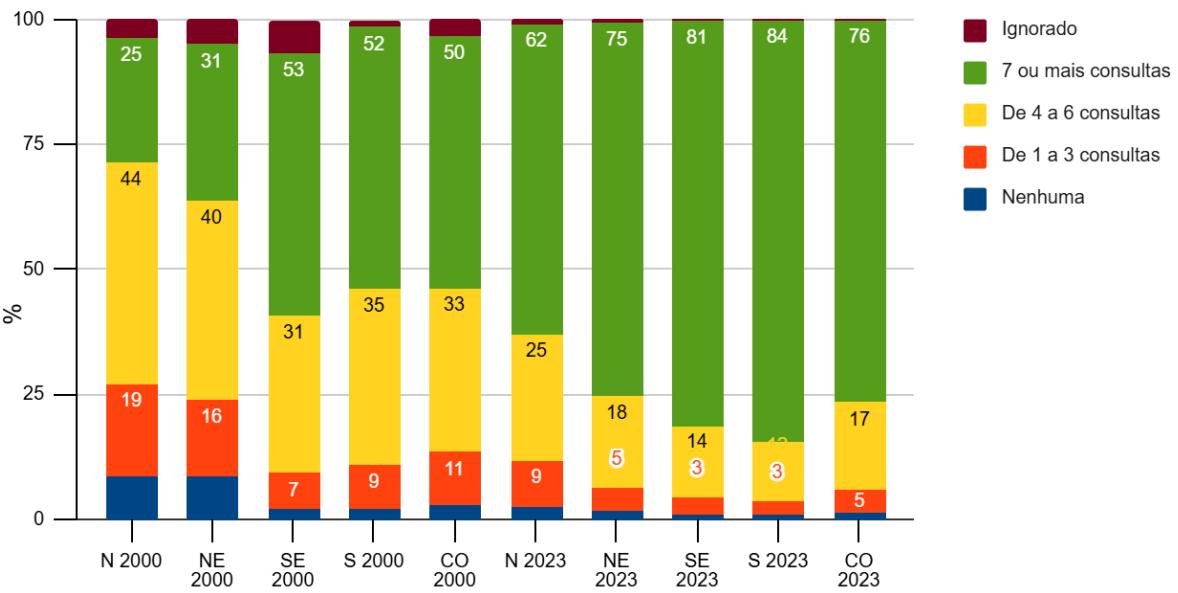


Figura A.22 Consultas pré-natal por região nos anos 2000 e 2023

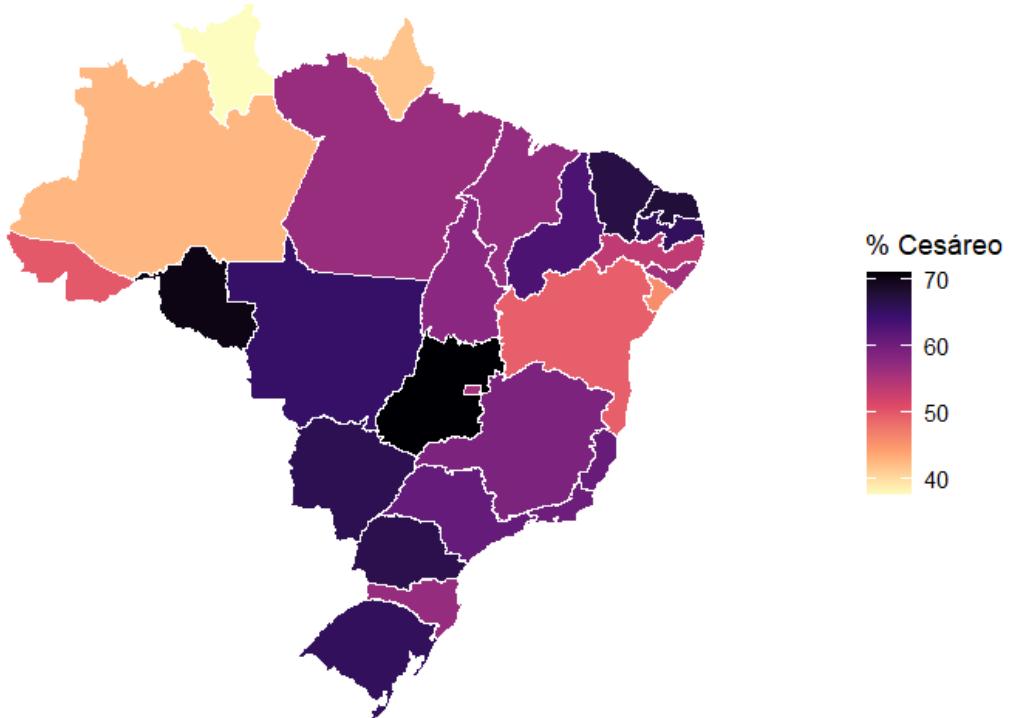


Figura A.23 Porcentagem de parto do tipo cesárea por UF em 2023

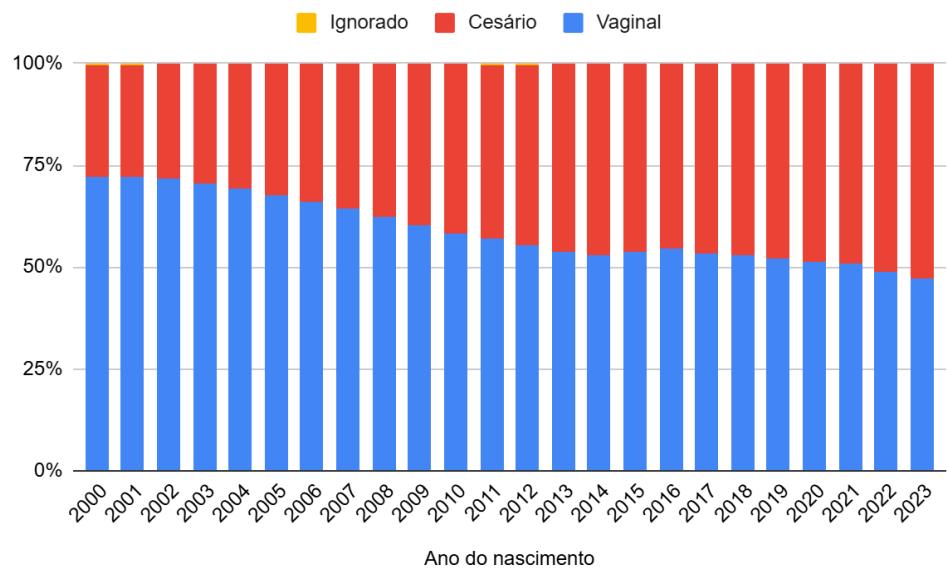


Figura A.24 Tipo de parto na região Norte do ano 2000 ao ano 2023

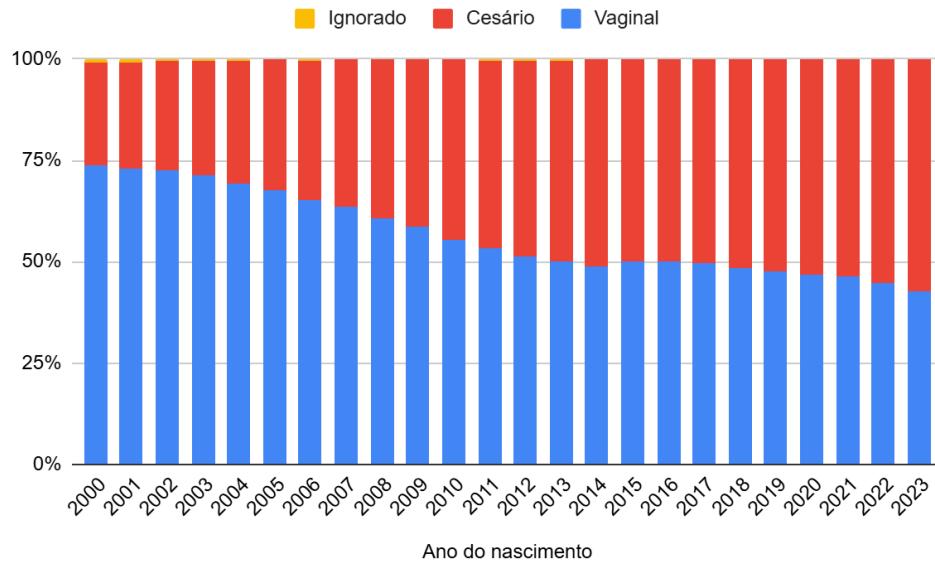


Figura A.25 Tipo de parto na região Nordeste do ano 2000 ao ano 2023

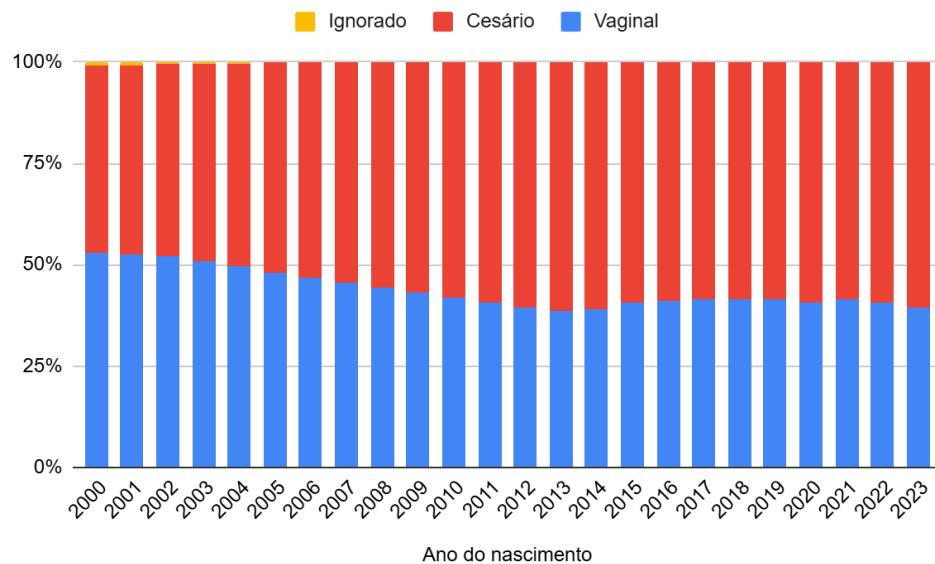


Figura A.26 Tipo de parto na região Sudeste do ano 2000 ao ano 2023

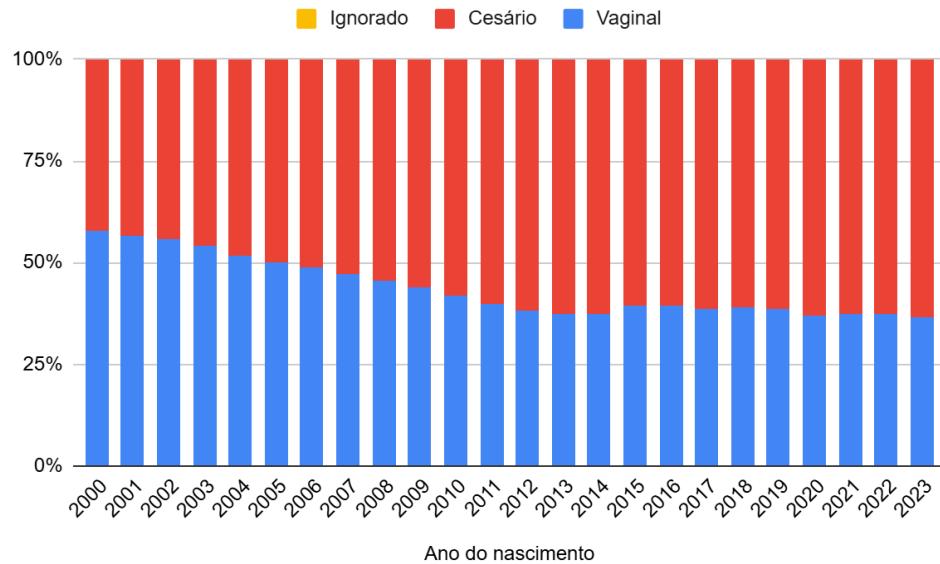


Figura A.27 Tipo de parto na região Sul do ano 2000 ao ano 2023

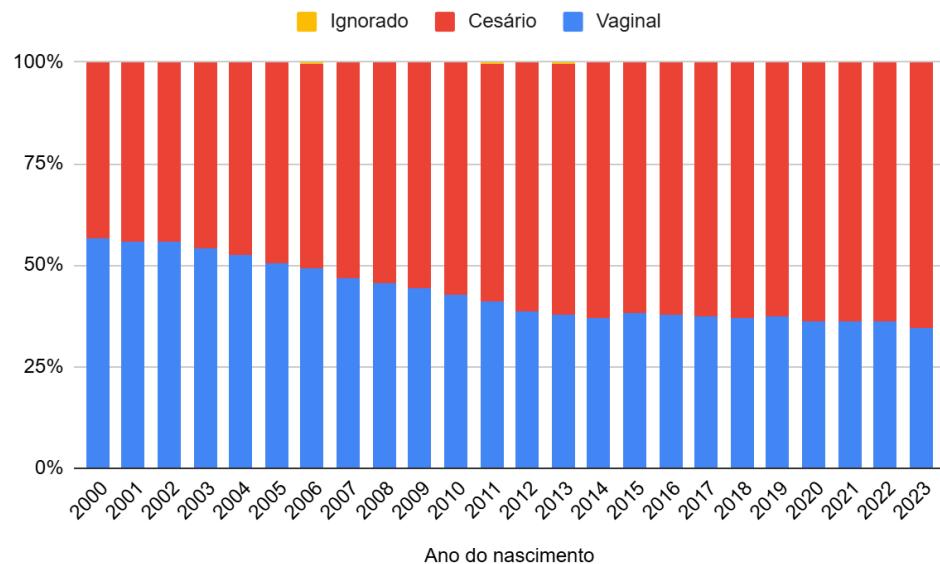


Figura A.28 Tipo de parto na região Centro Oeste do ano 2000 ao ano 2023

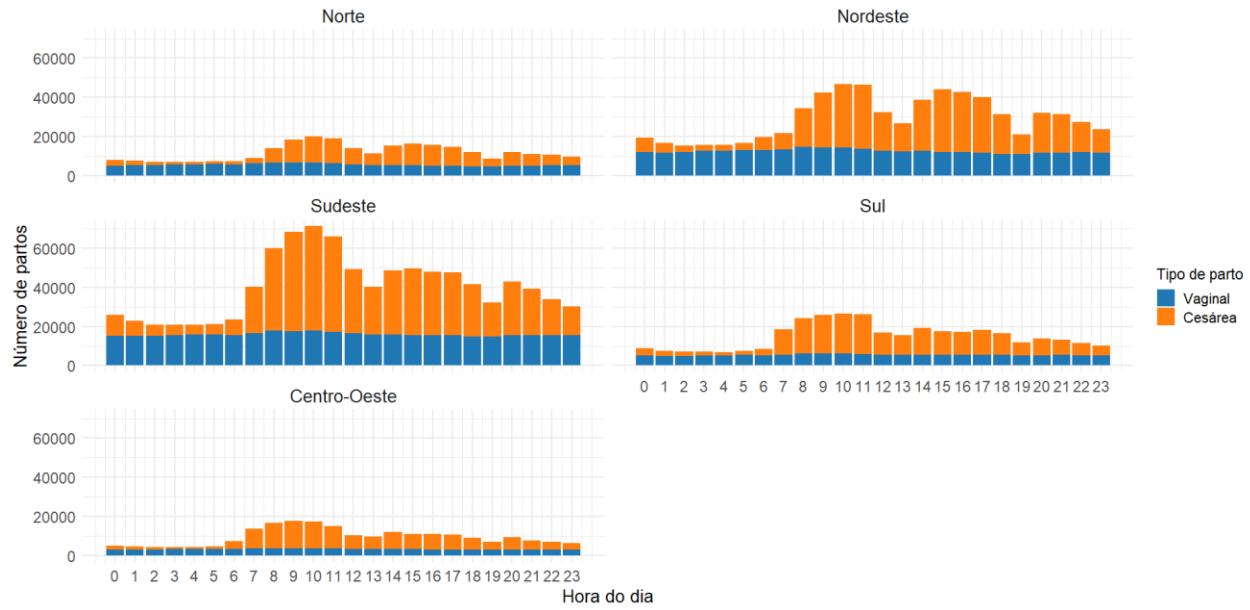


Figura A.29 Número de partos por hora do dia e por região no ano de 2023

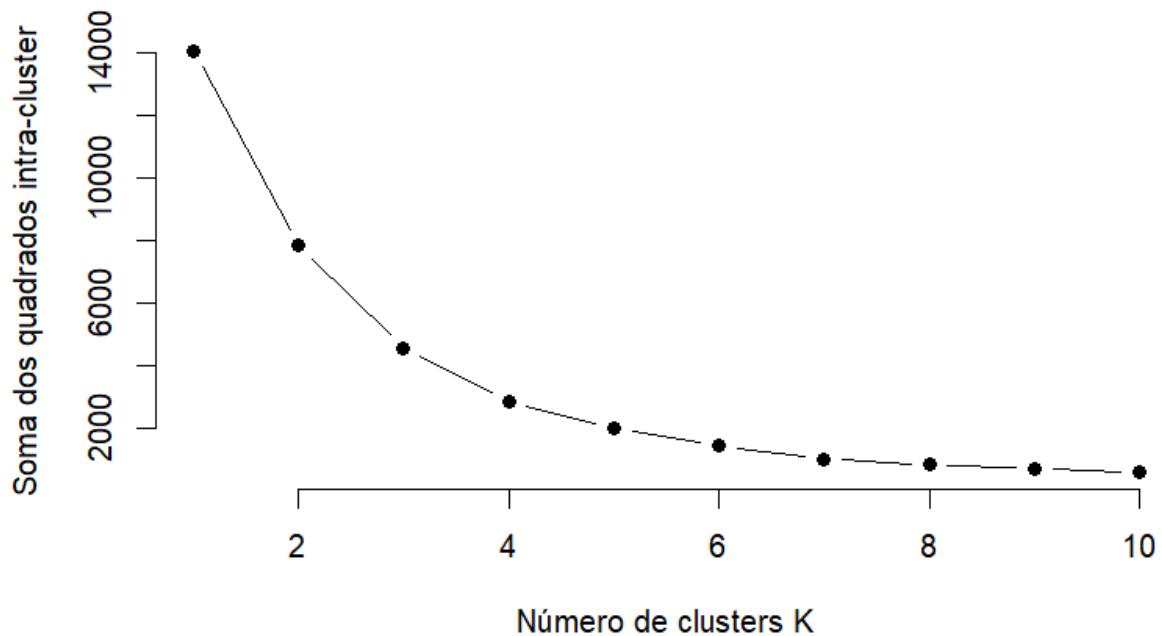


Figura A.30 Método do cotovelo para determinar o número k de grupos ideal

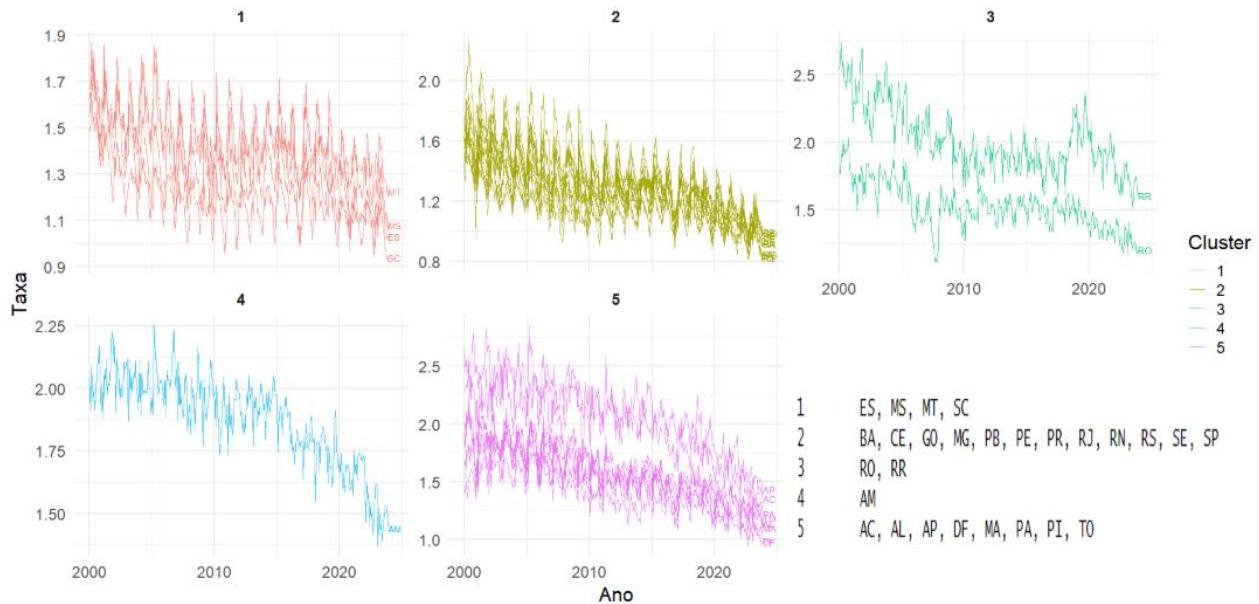


Figura A.31 Taxas mensais de natalidade (por 1000 hab) das UF por cluster - k-means

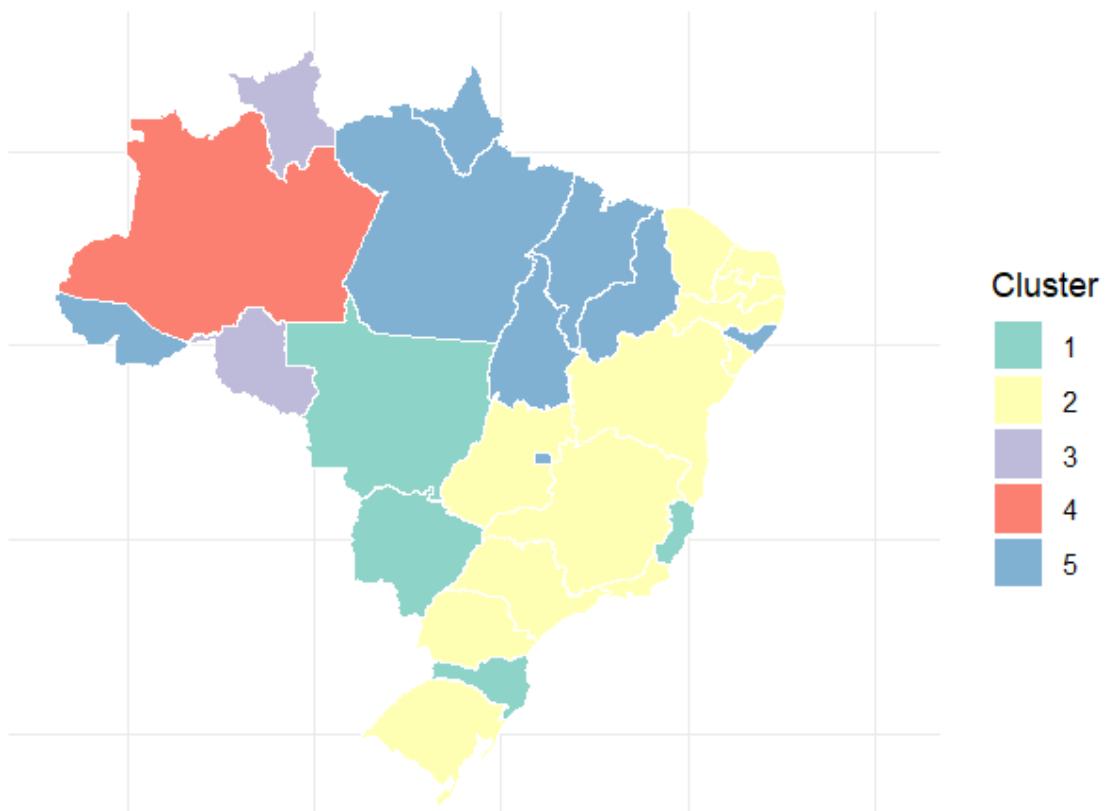


Figura A.32 Mapa dos agrupamentos por k-means

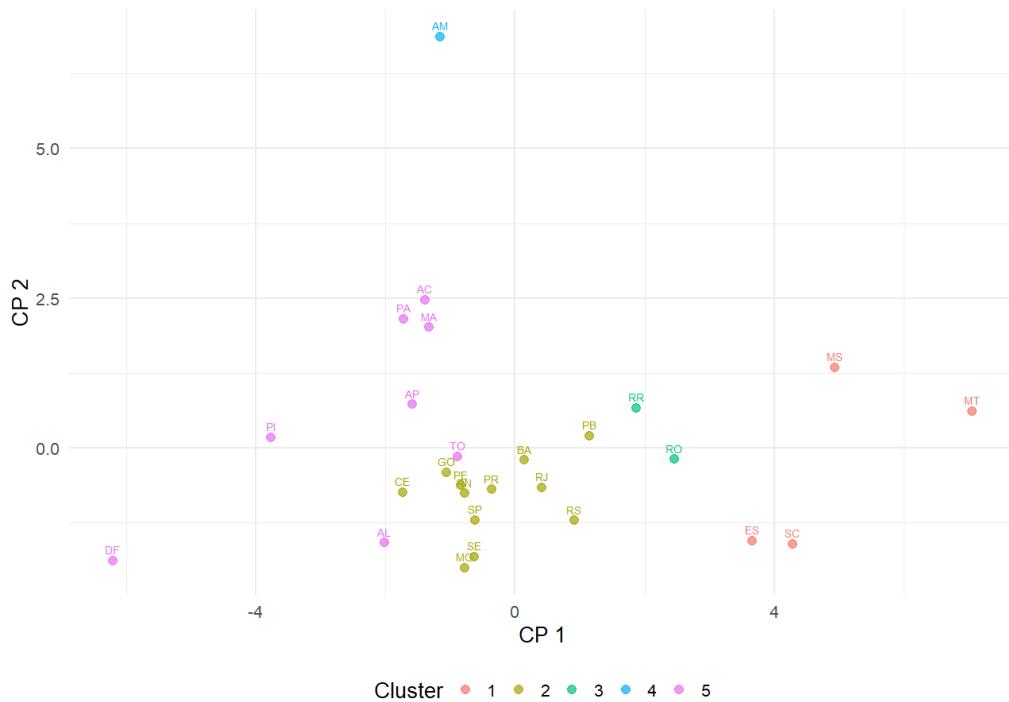


Figura A.33 Gráfico *biplot* das componentes principais 1 e 2 com clusters das UF agrupadas por k-means

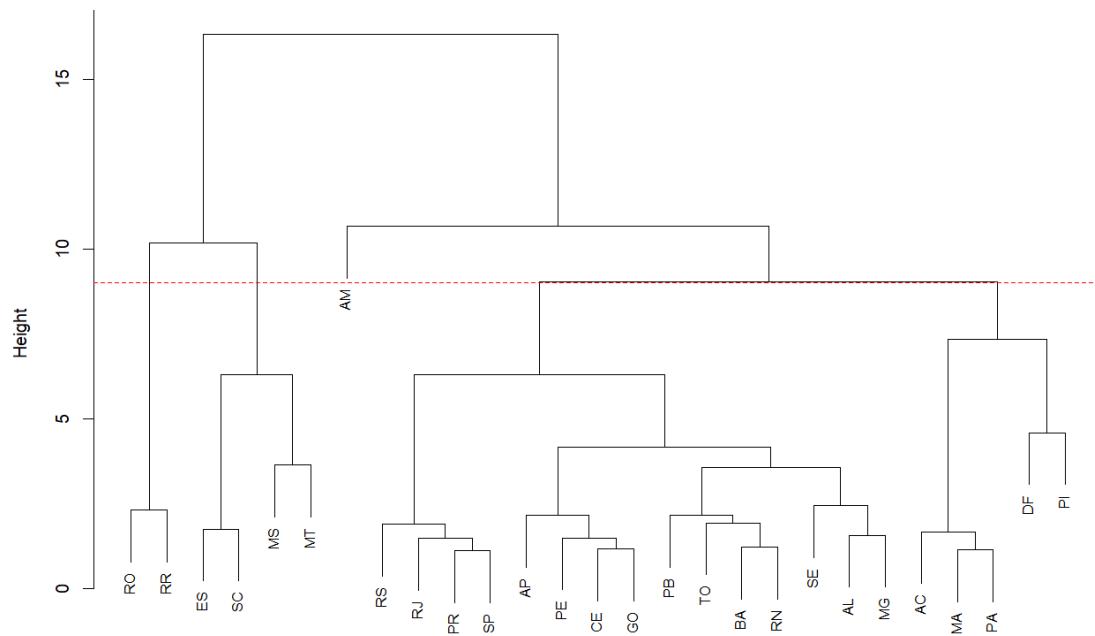


Figura A.34 Dendrograma das UF obtido pela método ward

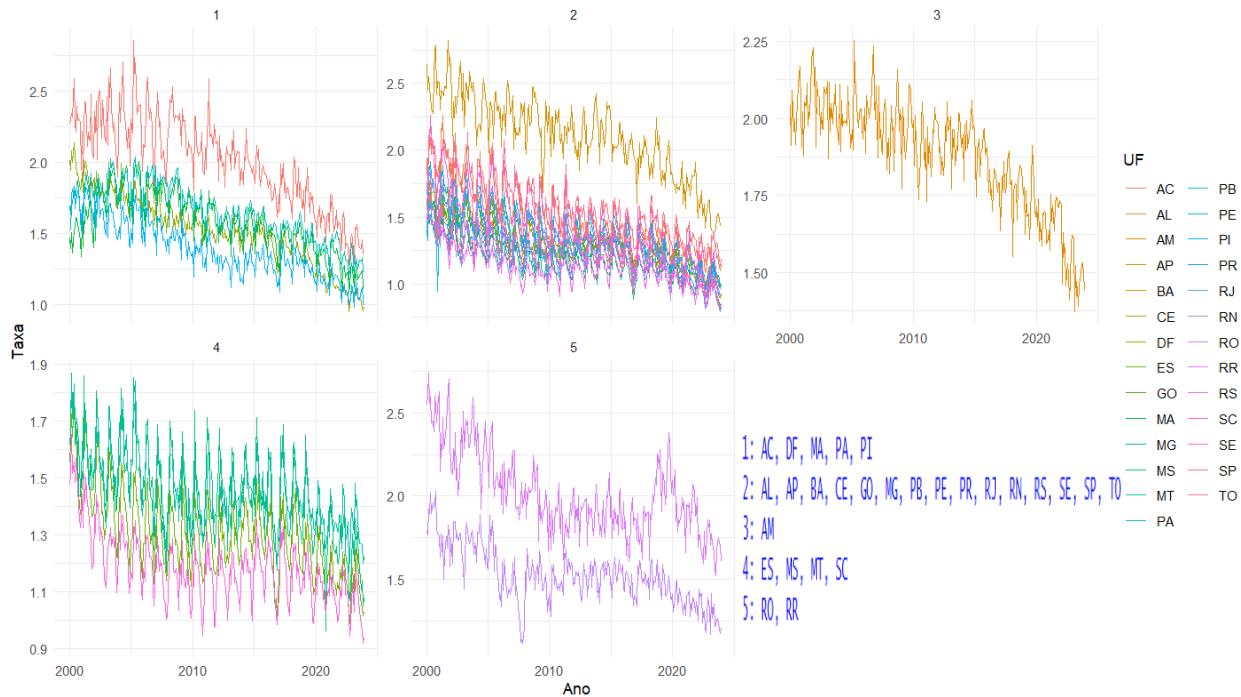


Figura A.35 Taxas mensais de natalidade (por 1000 hab) das UF por cluster hierárquico

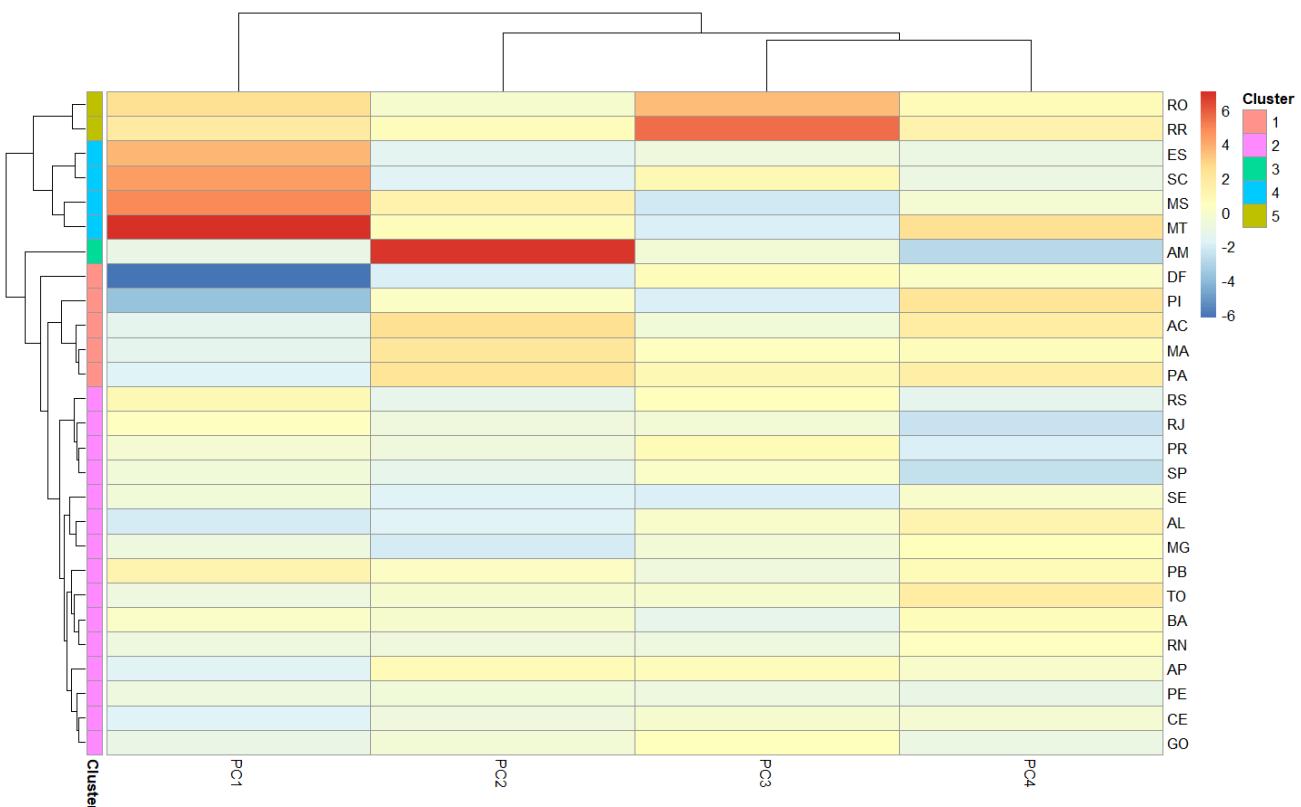


Figura A.36 Gráfico heatmap dos escores das componentes principais

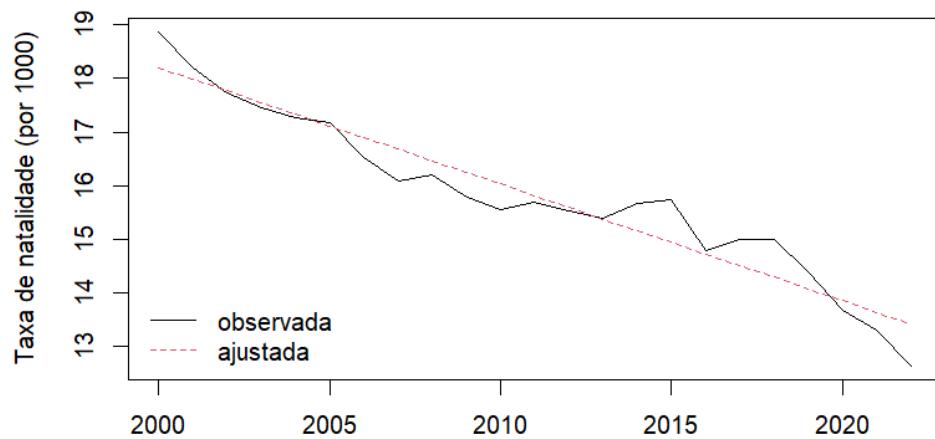


Figura A.37 Taxa de natalidade (por 1000 hab) do Brasil com ajuste de regressão linear

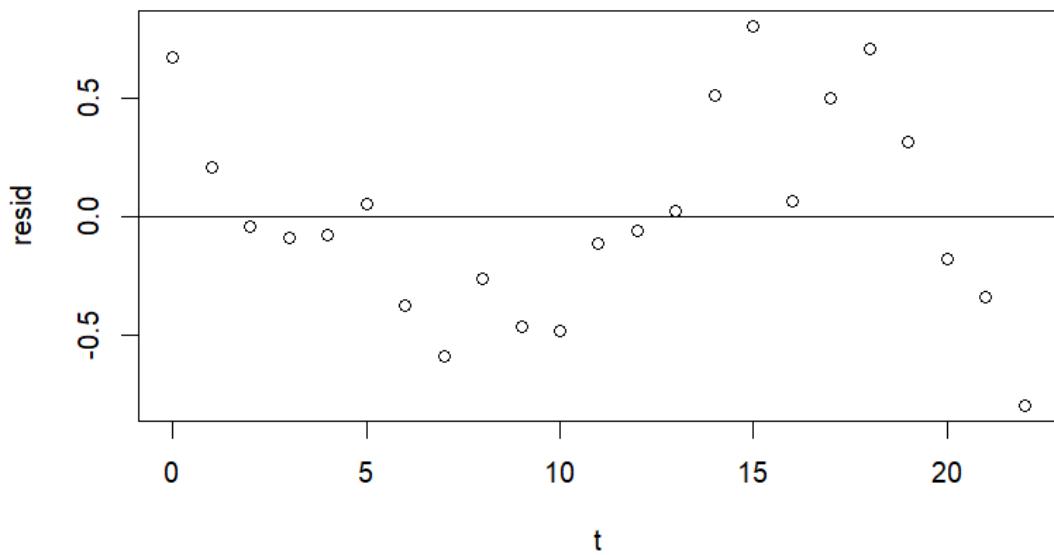


Figura A.38 Gráfico dos resíduos do ajuste por regressão linear

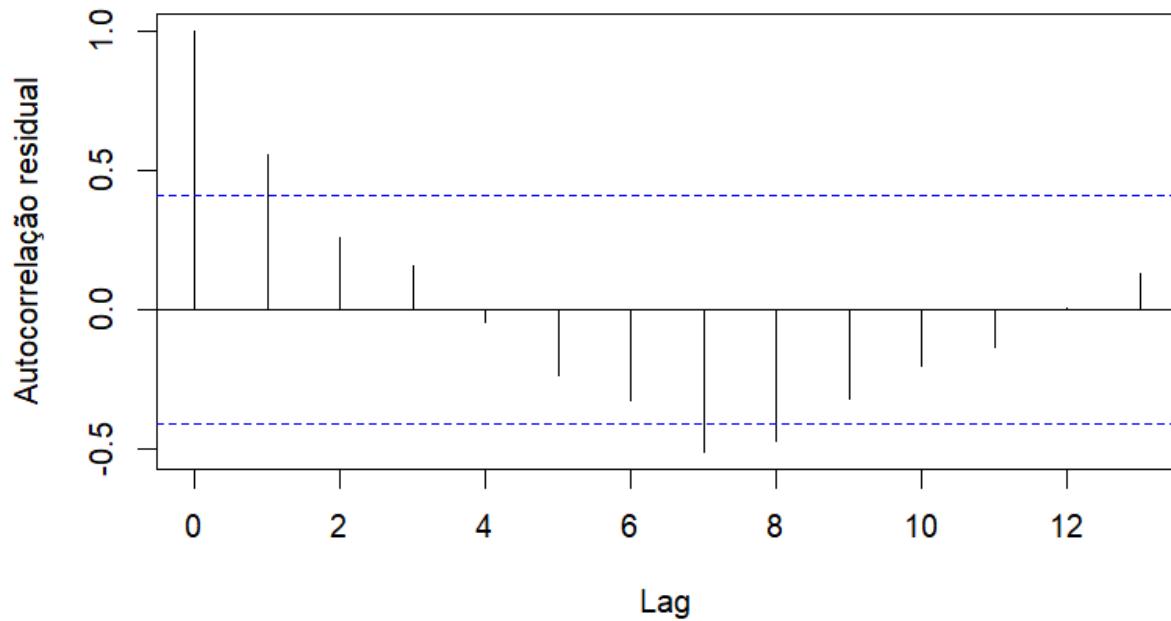


Figura A.39 Gráfico da função de autocorrelação (ACF) do modelo de regressão linear

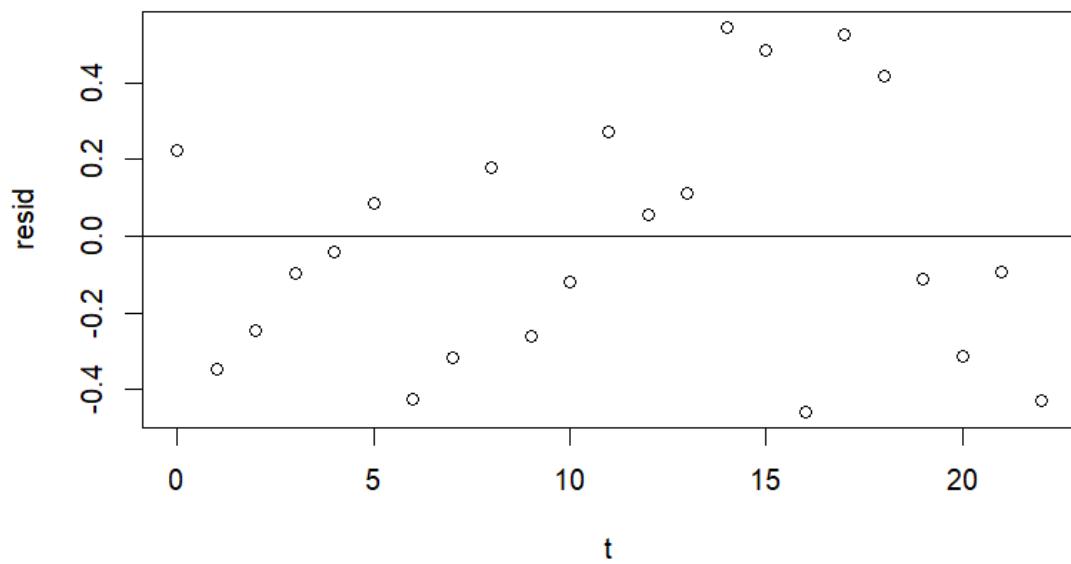


Figura A.40 Gráfico da função de autocorrelação (ACF) do modelo de regressão linear com erro autoregressivo

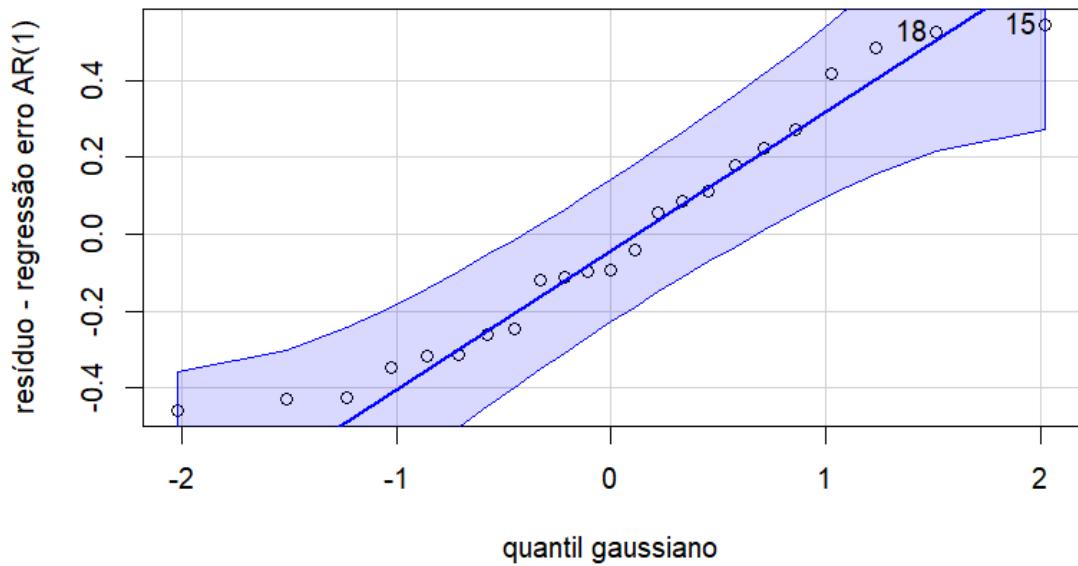


Figura A.41 Gráfico quantil dos resíduos para modelo de regressão linear com erro autoregressivo

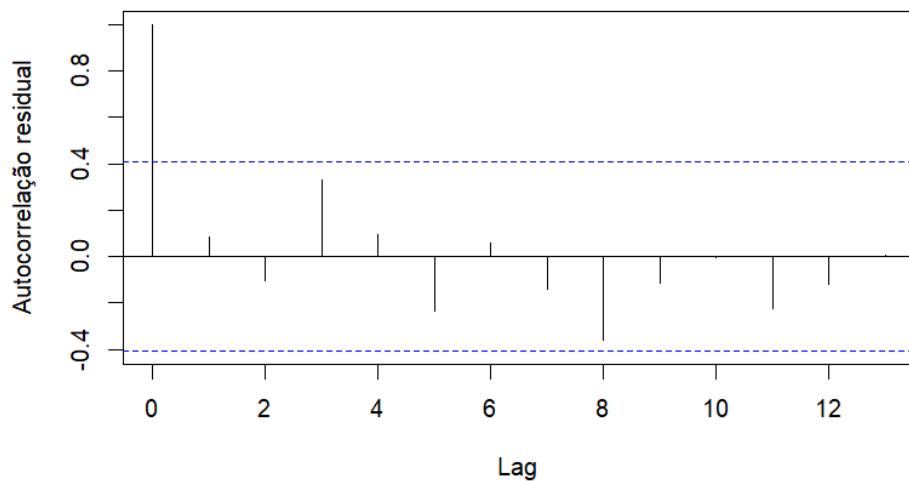


Figura A.42 Gráfico da função de autocorrelação (ACF) do modelo de regressão linear com erro autoregressivo

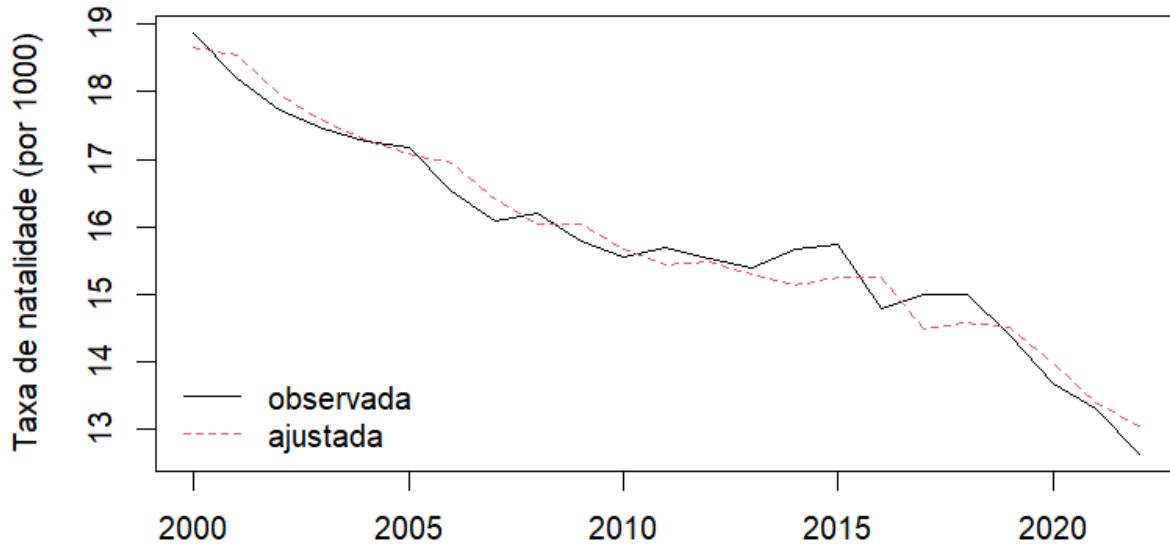


Figura A.43 Taxa de natalidade (por 1000 hab) do Brasil com ajuste do modelo de regressão linear com erro autoregressivo

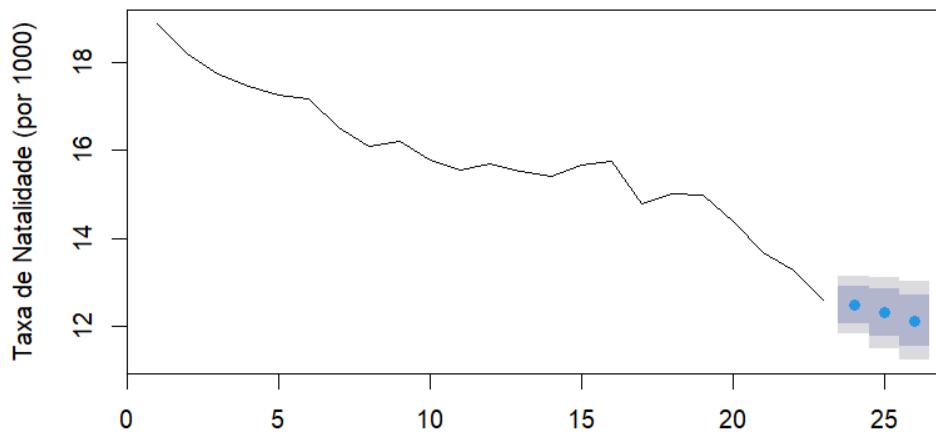


Figura A.44 Previsão da taxa de natalidade para 2024, 2025 e 2026 a partir do modelo de regressão linear com erro autorregressivo

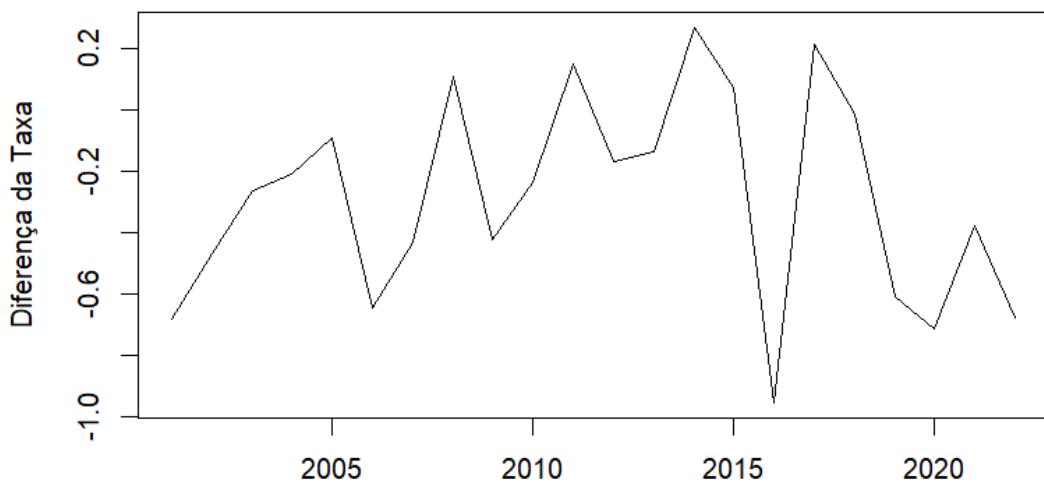


Figura A.45 Diferença da série de taxa da natalidade do Brasil

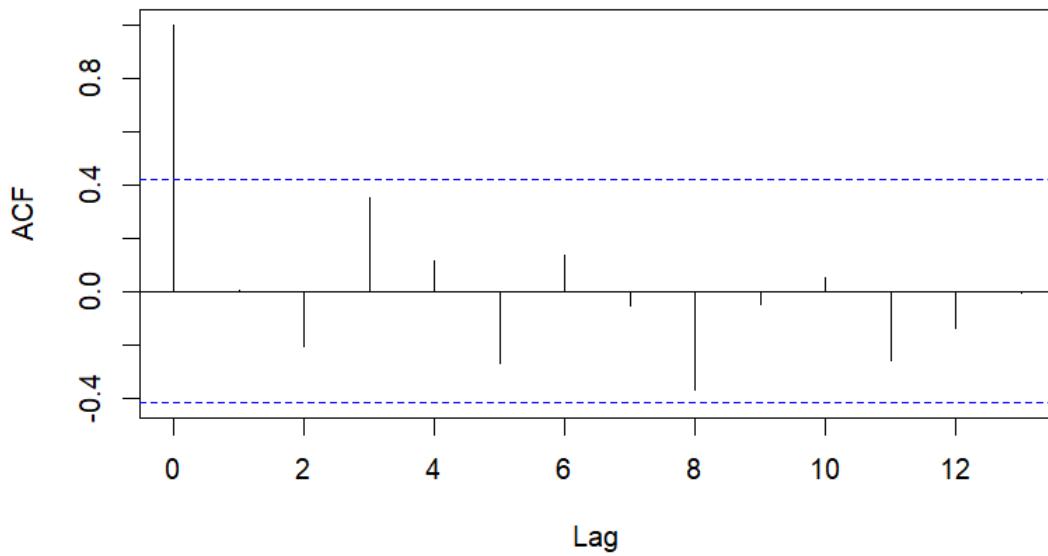


Figura A.46 Gráfico da função de autocorrelação (ACF) do modelo ARIMA(0,1,0)

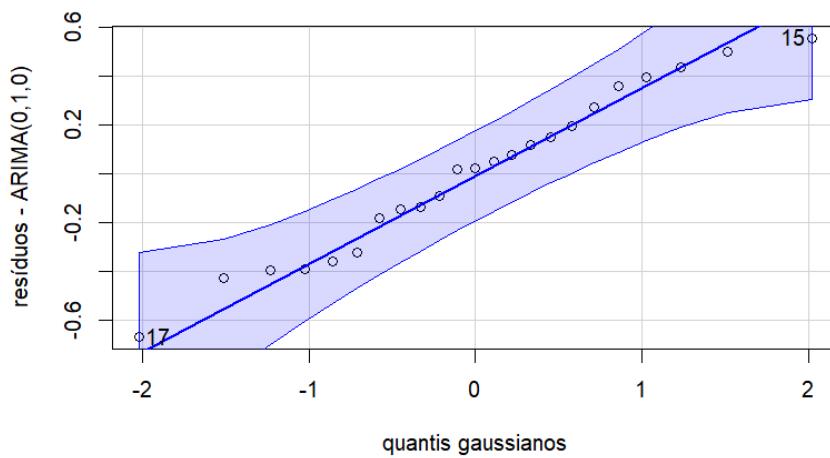


Figura A.47 Gráfico quantil dos resíduos para o modelo ARIMA(0,1,0)

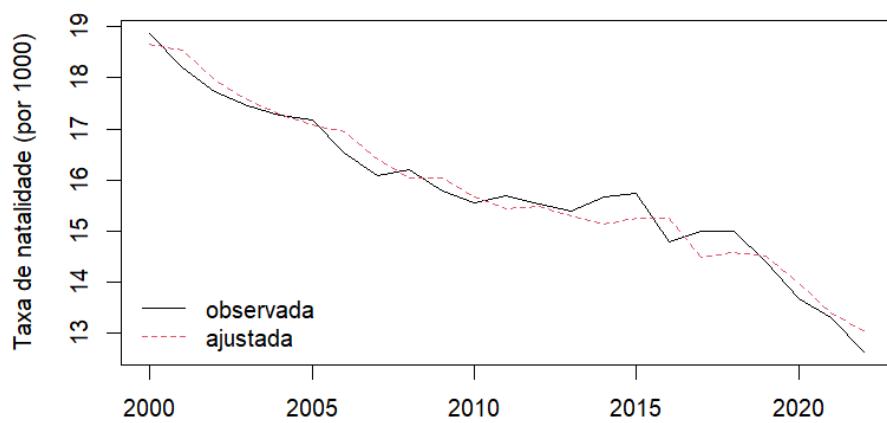


Figura A.48 Taxa de natalidade (por 1000 hab) do Brasil com ajuste do modelo ARIMA(0,1,0)

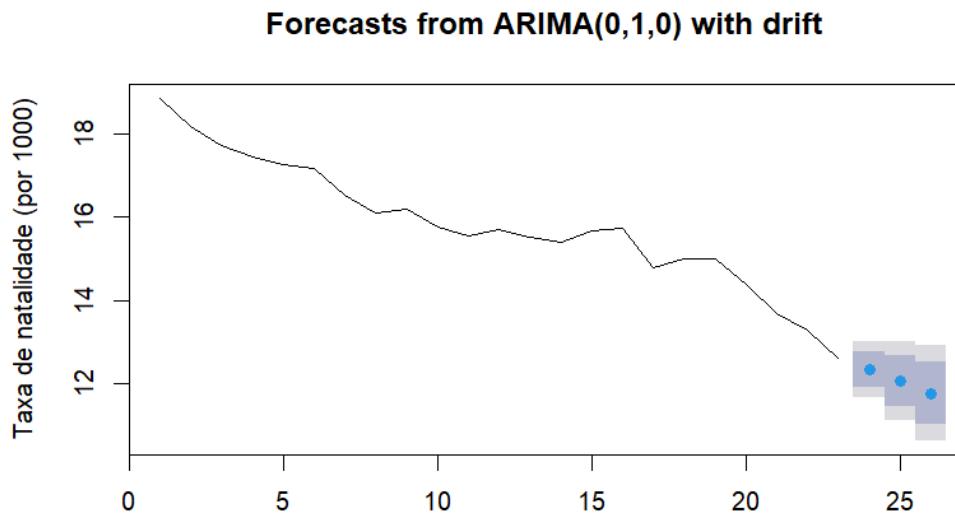


Figura A.49 Previsão da taxa de natalidade para 2024, 2025 e 2026 a partir do modelo ARIMA(0,1,0)

APÊNDICE B

Tabelas

Tabela B.1 Distribuição do Peso ao nascer em porcentagem pela Duração da gestação

Duração da gestação	Peso ao nascer								Total
	<499	500 a 999	1000 a 1499	1500 a 2499	2500 a 2999	3000 a 3999	≥ 4000	Ignorado	
≤ 21	50,64	15,07	3,79	5,57	7,64	16,50	0,79	0,00	100,00
De 22 a 27	9,31	61,06	14,34	4,49	3,84	6,48	0,45	0,04	100,00
De 28 a 31	0,40	12,41	39,41	30,20	7,18	9,90	0,48	0,01	100,00
De 32 a 36	0,09	0,22	2,52	39,75	31,78	24,41	1,21	0,01	100,00
De 37 a 41	0,07	0,02	0,05	4,01	23,58	67,67	4,58	0,01	100,00
≥ 42	0,10	0,05	0,08	2,80	18,86	71,40	6,70	0,02	100,00

Tabela B.2 Recorte dos valores das 22 características selecionadas pelo Catch22 para a série mensal da natalidade de cada estado do Brasil

UF	DN_HistogramMode_5	DN_HistogramMode_10	CO_f1ecac	CO_FirstMin_ac	CO_HistogramAMI_even_2_5
AC	0.18708827	-0.059509317	51.140964	7	0.4241175
AL	-0.41891985	-0.656231369	40.749323	7	0.5252703
AM	0.61759630	0.368977779	42.159298	3	0.4216889
AP	-0.01582841	-0.254541958	39.697263	7	0.4887636
BA	0.06937659	-0.629452740	18.009591	6	0.4272406
CE	0.19791137	-0.043059405	30.608550	7	0.5278324
DF	0.07307889	-0.161004928	49.675949	9	0.7962308
ES	-0.54741618	-0.282275921	4.884785	7	0.3075165
GO	0.23957207	-0.479749232	29.360070	7	0.4975478
MA	-0.09293581	-0.316268564	41.085945	6	0.5079159

Tabela B.3 Contribuição de cada variável no cálculo das componentes principais (Loadings)

	PC1	PC2	PC3	PC4
DN_HistogramMode_5	-0.085412653	0.19614383	0.056689564	-0.351832358
DN_HistogramMode_10	0.030545893	0.22638555	-0.046808752	-0.406394117
CO_f1ecac	-0.306610390	0.20093544	0.076641809	0.197213695
CO_FirstMin_ac	0.017314422	-0.28459047	0.375489686	0.210502921
CO_HistogramAMI_even_2_5	-0.342186639	-0.07438083	0.100483478	0.007488934
CO_trev_1_num	0.170133575	-0.10166180	-0.335198654	0.180596616
MD_hrv_classic_pnn40	0.154415210	-0.05055158	0.349058146	-0.253522848
SB_BinaryStats_mean_longstretch1	-0.251218574	-0.08979088	-0.118261866	0.152283473
SB_TransitionMatrix_3ac_sumdiagcov	-0.254075767	0.04927581	-0.008883018	0.323502432
PD_PeriodicityWang_th0_01	0.077318855	-0.37724972	-0.131479676	0.159284268
CO_EMBED2_Dist_tau_d_expfit_meandiff	-0.252554993	-0.21497668	-0.012454951	-0.187376073
IN_AutoMutualInfoStats_40_gaussian_fmmi	0.069788555	-0.23974558	0.386287825	0.229123768
FC_LocalSimple_mean1_tauresrat	0.308810048	-0.08525240	-0.102782857	-0.111566994
DN_OutlierInclude_p_001_mdrmd	-0.010959239	0.40832402	-0.157313224	0.238029802
DN_OutlierInclude_n_001_mdrmd	-0.218307686	0.07437967	-0.255818088	-0.228439999
SP_Summaries_welch_rect_area_5_1	-0.265369895	-0.25847178	-0.144583722	-0.206973402
SB_BinaryStats_diff_longstretch0	0.107839905	-0.23966773	-0.398913594	0.056522795
SB_MotifThree_quantile_hh	0.290234868	0.07506443	0.263858842	-0.106581112
SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1	0.103572349	0.16768994	0.105629439	0.012964711
SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1	-0.004256776	0.41656510	0.067182640	0.311641404
SP_Summaries_welch_rect_centroid	0.285942084	0.03380442	-0.209493012	0.095733831
FC_LocalSimple_mean3_stderr	0.351205998	0.04225670	-0.112853667	0.070349182

Tabela B.4 Escores dos estados em ordem numérica por componente principal

PC1	PC2	PC3	PC4	UF
-1.3937337	2.4747216	-0.59355078	1.669180120	AC
-2.0112108	-1.5810957	0.01952246	1.200403839	AL
-1.1600103	6.8697082	-0.46066786	-2.829651693	AM
-1.5896928	0.7287277	0.59639217	-0.029597634	AP
0.1477522	-0.2011346	-1.28697776	0.603706168	BA
-1.7303298	-0.7359087	-0.12832056	-0.287327863	CE
-6.2107624	-1.8813148	0.61131712	0.088157166	DF
3.6640032	-1.5471392	-0.67825318	-0.997841786	ES
-1.0560921	-0.4025370	0.43912756	-0.959494260	GO
-1.3238276	2.0147574	0.40670980	0.580591242	MA
-0.7796630	-2.0037841	-0.37629393	0.448945948	MG
4.9264577	1.3494023	-2.18470433	-0.294986893	MS
7.0527520	0.6180371	-1.83958872	2.538453849	MT
-1.7217441	2.1490252	0.88305895	1.511163393	PA
1.1431333	0.2081890	-0.71747721	0.783768270	PB
-0.8432230	-0.6267934	-0.86957406	-1.150213352	PE
-3.7606227	0.1792288	-1.84997819	2.263484215	PI
-0.3567544	-0.6909916	0.79666169	-1.922973149	PR
0.4058417	-0.6618829	-0.38819176	-2.445777738	RJ
-0.7833154	-0.7585845	-0.80308707	0.395357648	RN
2.4596861	-0.1885536	3.52503510	0.736293994	RO
1.8687848	0.6633138	5.51470952	1.249444457	RR
0.9114202	-1.2103274	0.46443799	-1.327279114	RS
4.2855331	-1.5988648	0.94708815	-1.027344225	SC
-0.6333845	-1.8138265	-1.96156337	0.003421251	SE
-0.6184551	-1.2066489	0.09087184	-2.543218030	SP
-0.8925425	-0.1457236	-0.15670357	1.743334177	TO

Tabela B.5 Resultados do ajuste do modelo de regressão linear com erro ARIMA(1,0,0)

Coefficients:

	ar1	intercept	xreg
s.e.	0.7401	18.5443	-0.2498
	0.1536	0.4094	0.0315

$\sigma^2 = 0.1118$: log likelihood = -6.23
AIC=20.46 AICc=22.68 BIC=25

Tabela B.6 Resultados do ajuste do modelo ARIMA(0,1,0)

Coefficients:

	drift
s.e.	-0.2846
	0.0712

$\sigma^2 = 0.1168$: log likelihood = -7.08
AIC=18.16 AICc=18.79 BIC=20.34