

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Perspectivas Pós-Pandemia de COVID-19: Previsão
do Desemprego no Brasil com o Método de
Holt-Winters**

Adriana Eva Fernandes da Silva

Trabalho de Conclusão de Curso

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Perspectivas Pós-Pandemia de COVID-19: Previsão do
Desemprego no Brasil com o Método de Holt-Winters

Adriana Eva Fernandes da Silva
Orientadora: Maria Sílvia de Assis Moura

Trabalho de Conclusão de Curso apresentado
como parte dos requisitos para obtenção do
título de Bacharel em Estatística.

São Carlos
Setembro de 2024

Adriana Eva Fernandes da Silva

Perspectivas Pós-Pandemia de COVID-19: Previsão do
Desemprego no Brasil com o Método de Holt-Winters

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Adriana Eva Fernandes da Silva e aprovado pela banca examinadora.

Aprovado em dia de mês de ano

Banca Examinadora:

- Maria Sílvia de Assis Moura (Orientadora)
- Francisco Antonio Rojas Rojas
- Luis Aparecido Milan

Agradecimentos

Agradeço primeiramente aos meus pais, Rosa e Ademir, por todo suporte, atenção e consideração para que tudo isso fosse possível; sem eles, esse sonho não teria se tornado realidade.

Agradeço às minhas colegas de curso, Carla Gabriela e Léia Timóteo; o caminho da graduação sempre foi árduo, mas se eu estivesse com elas, o percurso se tornava mais leve e significativo.

E por fim, expresso minha profunda gratidão à minha orientadora por toda paciência e consideração que teve comigo durante a escrita desse trabalho. Cada um de vocês desempenhou um papel fundamental nesta jornada e contribuiu para o sucesso deste trabalho. Muito obrigado por fazerem parte deste importante capítulo da minha vida.

Resumo

Este estudo visa analisar o impacto da pandemia de COVID-19 nas taxas de desemprego, com um foco especial na disparidade de gênero. Utilizando métodos de séries temporais, concentraremos nossa atenção na compreensão de como a pandemia afetou o desemprego. Empregaremos o método de suavização exponencial de Holt-Winters para realizar previsões sobre o comportamento das séries temporais relacionadas aos homens e mulheres. Posteriormente, compararemos essas previsões com os dados reais para avaliar a precisão e coerência das projeções realizadas. Este trabalho busca proporcionar insights valiosos sobre as nuances do impacto da pandemia no desemprego, especialmente no que diz respeito às diferenças de gênero.

Palavras-chave: *gênero, pandemia, desemprego, suavização exponencial, Holt-Winters, previsões, séries temporais, análise de intervenção .*

Lista de Figuras

4.1	Série Temporal da taxa de desemprego dos gêneros masculino e feminino. . .	31
4.2	Série Temporal do desemprego dos gêneros feminino e masculino antes da pandemia.	32
4.3	ACF da Série Temporal do desemprego do gênero masculino antes da pan- demia.	36
4.4	Aplicando a diferença no ACF da Série da Taxa de Desemprego do gênero Masculino antes da Pandemia	36
4.5	PACF da Série Temporal do desemprego do gênero feminino antes da pan- demia.	37
4.6	ACF da Série Temporal do desemprego do gênero feminino antes da pan- demia.	37
4.7	Aplicando a diferença no ACF da Série da Taxa de Desemprego do gênero Feminino antes da Pandemia	37
4.8	PACF da Série Temporal do desemprego do gênero feminino antes da pan- demia.	38
4.9	Gráfico da Série Temporal diferenciada do desemprego do gênero feminino antes da pandemia.	39
4.10	Gráfico da Série Temporal diferenciada do desemprego do gênero masculino antes da pandemia.	40
5.1	Gráfico da previsão da série de 2020 da taxa de desemprego feminino usando Holt-Winters	49
5.2	Gráfico da previsão da série de 2021 da taxa de desemprego feminino usando Holt-Winters	50
5.3	Gráfico da previsão da série de 2022 da taxa de desemprego feminino usando Holt-Winters	51

5.4	Gráfico da previsão da série de 2023 da taxa de desemprego feminino usando Holt-Winters	52
5.5	Gráfico da previsão da série de 2020 da taxa de desemprego masculino usando Holt-Winters	53
5.6	Gráfico da previsão da série de 2021 da taxa de desemprego masculino usando Holt-Winters	54
5.7	Gráfico da previsão da série de 2022 da taxa de desemprego masculino usando Holt-Winters	55
5.8	Gráfico da previsão da série de 2023 da taxa de desemprego masculino usando Holt-Winters	56

Lista de Tabelas

5.1	Métricas para os métodos de Holt-Winters da série de desemprego feminino	44
5.2	Métricas para os métodos de Holt-Winters da série de desemprego masculino	44
5.3	Valores reais e estimados da série de desemprego do gênero feminino	45
5.4	Valores reais e previsão da série de desemprego do gênero feminino	46
5.5	Valores reais e estimados da série de desemprego do gênero masculino	47
5.6	Valores reais e previsão da série de desemprego do gênero masculino	48
5.7	Comparação das métricas de métodos de previsão para dados masculinos e femininos no ano de 2020	57
5.8	Comparação das métricas de métodos de previsão para dados femininos e masculinos no ano de 2021	57
5.9	Comparação das métricas de métodos de previsão para dados masculinos e femininos no ano de 2022	57
5.10	Comparação das métricas de métodos de previsão para dados femininos e masculinos no ano de 2023	58

Sumário

1	Introdução	15
2	Objetivo	19
3	Materiais e Métodos	21
3.1	Dados sobre as séries de desemprego para os gêneros feminino e masculino	21
3.1.1	Considerações à respeito dos dados	21
3.2	Metodologia	22
3.3	Modelos de suavização exponencial	22
3.3.1	Suavização exponencial de Holt-Winters	23
3.3.2	Modelos para séries sazonais	24
3.3.3	Suavização exponencial sazonal de Holt-Winters	24
3.3.4	Análise de intervenção	29
4	Análise estatística dos dados	31
4.1	Conhecendo a série	31
4.1.1	Comportamento das séries antes da pandemia	32
4.1.2	Análise descritiva das séries	33
4.1.3	Estacionariedade	33
4.1.4	Tendência	34
4.1.5	Sazonalidade	38
4.1.6	Homocedasticidade	38
5	Aplicação	41
5.0.1	Método de Holt-Winters	41
5.0.2	Escolha do método	42
5.0.3	Previsão	44

5.0.4	Gráficos da previsão	47
5.0.5	Conclusão	56
6	Conclusão	61
	Referências Bibliográficas	63

Capítulo 1

Introdução

O desemprego, conforme destacado por [IBGE - Instituto Brasileiro de Geografia e Estatística \(2023a\)](#), de maneira simplificada, refere-se às pessoas com idade para trabalhar (acima de 14 anos) que não estão empregadas, mas estão disponíveis e buscando oportunidades de trabalho. Assim, para ser considerado desempregado, não é suficiente não possuir um emprego, é necessário estar à procura de emprego, pela definição do IBGE.

Diversos casos exemplificam situações em que, embora as pessoas não tenham emprego formal, não podem ser classificadas como desempregadas:

- Um estudante universitário que dedica seu tempo exclusivamente aos estudos.
- Uma dona de casa que não trabalha fora de casa.
- Uma empreendedora que administra seu próprio negócio.

Segundo a metodologia adotada pelo IBGE na Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua, o estudante e a dona de casa são considerados fora da força de trabalho, enquanto a empreendedora é classificada como ocupada.

A PNAD Contínua, responsável por mostrar a quantidade de desempregados no Brasil, utiliza o conceito de “desocupação” para aquilo que é popularmente conhecido como “desemprego”. Nessa pesquisa, a taxa de desocupação representa a porcentagem de pessoas na força de trabalho que estão desempregadas. A taxa de desemprego é calculada como a porcentagem de desocupados em relação à força de trabalho total, representada pela fórmula: $[\text{desocupados por gênero} / \text{total de pessoas por gênero}] \times 100$.

De acordo com [Organização Pan-Americana de Saúde \(Acessado em 2023\)](#), a pandemia de COVID-19, que surgiu em 2019 e se espalhou rapidamente pelo mundo, não apenas desencadeou uma crise global de saúde, mas também revelou e agravou disparidades sociais

profundamente enraizadas. Entre essas disparidades, as diferenças de gênero no mercado de trabalho emergiram como uma preocupação significativa. No contexto brasileiro, a pandemia teve repercussões no mercado de trabalho, com efeitos desiguais sobre homens e mulheres.

De acordo com [IBGE - Instituto Brasileiro de Geografia e Estatística \(2023b\)](#), a PNAD Contínua anual de 2019 já indicava a desigualdade de gênero, uma vez que 67,3 milhões de indivíduos fora da força de trabalho, dos quais 64,5% eram mulheres. Estar fora da força de trabalho significa que esses indivíduos não estavam empregados e não estavam procurando ativamente por emprego, por motivos como estudos, aposentadoria, afazeres domésticos, problemas de saúde, ou a crença de que não há oportunidades de emprego disponíveis.”

As restrições de *lockdown*, a retração econômica e a transformação dos padrões de trabalho tiveram implicações diretas nas taxas de emprego, especialmente quando observadas através da lente de gênero.

Mulheres enfrentaram desafios únicos, desde a sobrecarga de trabalho não remunerado até a maior vulnerabilidade ao desemprego em setores fortemente impactados pela pandemia, como turismo e serviços.

Conforme apontado por [Almeida de Castro \(2022\)](#), ao abranger a América Latina e considerar mulheres de todas as raças, observou-se que, embora por uma diferença percentual reduzida, elas enfrentaram efetivamente menos oportunidades em comparação com os trabalhadores do sexo masculino.

Direcionando o foco para o Brasil, especialmente na região Nordeste, um estudo conduzido por [Fernandes Silva \(2023\)](#) destaca uma disparidade mais pronunciada entre as taxas de desemprego de mulheres e homens.

Nessas análises, podemos observar que a disparidade no desemprego entre mulheres e homens pode ser mais pronunciada, especialmente ao considerarmos o contexto brasileiro. Este estudo se concentrará em dados abrangentes coletados em todo o Brasil. Utilizaremos técnicas de séries temporais para examinar a disparidade de desemprego entre homens e mulheres, utilizando a pandemia como variável de intervenção.

Este documento está estruturado da seguinte maneira: no Capítulo 2, são apresentados os objetivos deste estudo. No Capítulo 3, discutimos a exposição dos dados, incluindo informações sobre onde podem ser encontrados e oferecendo insights sobre como estão dispostos para facilitar o manuseio. Posteriormente, no Capítulo 4, vamos conhecer um

poucos dos dados, usando a análise descritiva. Finalmente, no Capítulo 5, vamos aplicar o método descrito e por fim, no Capítulo 6, faremos a conclusão com base no objetivo.

Capítulo 2

Objetivo

O objetivo deste estudo é analisar as tendências de desemprego entre homens e mulheres no Brasil, utilizando técnicas de séries temporais, tanto no período anterior quanto no período posterior à pandemia de COVID-19.

Essas técnicas são baseadas em aplicar análise descritiva das séries estatísticas de desemprego para os gêneros, aplicar suavização exponencial de Holt-Winters e fazer previsões, considerando uma análise de intervenção que vai ser a pandemia de COVID-19.

Pretendemos identificar padrões, flutuações e possíveis mudanças nas taxas de desemprego, permitindo uma compreensão aprofundada das disparidades de emprego de gênero ao longo do tempo, especialmente em resposta às influências da pandemia.

Capítulo 3

Materiais e Métodos

3.1 Dados sobre as séries de desemprego para os gêneros feminino e masculino

Os dados do trabalho serão retirados do portal [IBGE - Instituto Brasileiro de Geografia e Estatística \(2023b\)](#). Os dados estão disponíveis por taxa de desocupação trimestral e vamos utilizar o período do primeiro trimestre de 2012 até o quarto trimestre de 2019. Para realizar comparações para a variação, empregaremos os dados do primeiro trimestre de 2012 até o último trimestre de 2023 como referência. Para obter os dados de interesse, aplicamos filtros nos dados de desemprego segmentados por gênero, abrangendo o período do primeiro trimestre de 2012 até o quarto trimestre de 2023. Além disso, utilizamos filtros para a contagem de pessoas por gênero, cobrindo o período do primeiro trimestre de 2012 até o último trimestre de 2023.

3.1.1 Considerações à respeito dos dados

Os dados utilizados neste estudo são de acesso público. Optamos por empregar as taxas de desemprego, pois são os dados disponibilizados pelo Instituto Brasileiro de Geografia. Vale ressaltar que não temos acesso aos dados individuais que compõem essas taxas; elas são calculadas com base na proporção de pessoas desempregadas em relação ao total de pessoas de cada gênero em cada trimestre. Consequentemente, enfrentamos uma limitação na quantidade de observações, com apenas quatro observações anuais. Considerando o período que antecede a pandemia, de 2012 a 2019, temos um total de 32 observações, enquanto no período pós-pandemia, de 2020 a 2023, contamos com 16 observações.

3.2 Metodologia

Para atingir o objetivo, utilizaremos a linguagem de programação [R Core Team \(2019\)](#). Primeiramente, será feita a análise descritiva dos dados da Pesquisa Nacional por Amostra de Domicílio Contínua (PNAD Contínua) de 2012 até 2019. Após considerarmos o método de alisamento exponencial e fazer previsões e por fim, discutiremos se houve alteração ou não dos dados de 2020 até 2023 com base nas previsões feitas.

3.3 Modelos de suavização exponencial

Segundo [Morettin e Tolo \(2006\)](#) a maioria dos métodos de previsão baseia-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série.

Uma categoria abrangente de métodos de previsão, destinada a lidar com as diversas causas de flutuações em séries temporais, é conhecida como suavizações. Essas técnicas particulares dessa abordagem partem do pressuposto de que os valores extremos na série temporal refletem aleatoriedade. Portanto, ao suavizar esses extremos, busca-se identificar o padrão subjacente.

A popularidade expressiva dos métodos de suavização é atribuída à sua simplicidade, eficiência computacional e precisão. [Morettin e Tolo \(2006\)](#) destacam diversas abordagens de suavização, incluindo modelos apropriados para séries com constância local, nos quais podemos empregar médias móveis simples e suavização exponencial simples. No entanto, devido à limitada quantidade de observações neste trabalho, optamos por utilizar os modelos de suavização de Holt-Winters, reconhecidos por sua robustez mesmo em cenários com um número reduzido de observações, proporcionando resultados confiáveis nessas condições.

As informações sobre as taxas de desocupação são registradas trimestralmente, resultando em apenas quatro observações por ano. Desde o início do levantamento em 2012, acumulamos somente 32 observações trimestrais sobre as taxas de desocupação. Diante desse conjunto de dados restrito, destaca-se a relevância de métodos de previsão, como a suavização exponencial de Holt-Winters, que se mostram particularmente eficazes diante de um conjunto de dados tão limitado.

3.3.1 Suavização exponencial de Holt-Winters

Conforme mencionado anteriormente, a suavização exponencial de Holt-Winters demonstra robustez em séries temporais com poucos dados, sendo também amplamente aplicada em séries que apresentam tendência.

De acordo com [Morettin e Tolo \(2006\)](#), o método não apenas suaviza o nível da série, mas também utiliza uma nova constante de suavização para modelar a tendência da série. Os valores do nível e da tendência da série no instante t serão estimados por:

$$\bar{Z}_t = AZ_t + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad \text{e } t = 2, \dots, N, \quad (3.1)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad \text{e } t = 2, \dots, N, \quad (3.2)$$

respectivamente.

Temos,

- \bar{Z}_t : É o nível da série suavizada no tempo t ,
- \bar{Z}_{t-1} : É o nível da série suavizada no tempo $t - 1$,
- \hat{T}_t : Componente tendência do alisamento exponencial,
- \hat{F}_t : Componente sazonal do alisamento exponencial,

De maneira específica, A e C são identificadas como constantes de suavização.

Previsão

A previsão para o valor Z_{t+h} , com origem em t é:

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t, \quad \forall h > 0. \quad (3.3)$$

Em outras palavras, a previsão é calculada ao somar ao valor base (\bar{Z}_t) a tendência multiplicada pelo número de passos à frente desejados na previsão (h).

A fim de estabelecer as constantes de suavização, o processo envolve a seleção do valor do par de vetores (A, C) de modo a obter a mínima soma dos erros quadrados de previsão.

As equações (3.1) e (3.2) podem ser empregadas para atualizar a previsão com base em uma nova observação Z_{t+1} . Assim,

$$\begin{aligned}\bar{Z}_{t+1} &= AZ_{t+1} + (1 - A)(\bar{Z}_t + \hat{T}_t), \\ \hat{T}_{t+1} &= C(\bar{Z}_{t+1} - \bar{Z}_t) + (1 + C)\hat{T}_t\end{aligned}$$

Portanto a nova previsão para Z_{t+h} será:

$$\hat{Z}_{t+1}(h - 1) = \bar{Z}_{t+1} + (h - 1)\hat{T}_{t+1}. \quad (3.4)$$

Para que as equações possam ser aplicadas, é necessário estabelecer hipóteses para os seus valores iniciais.

3.3.2 Modelos para séries sazonais

No livro de [Ferreira *et al.* \(2017\)](#), é apresentado um exemplo em que a oferta de empregos tende a aumentar no final do ano, devido às festividades natalinas. Isso ocorre devido à maior demanda por bens e serviços, resultando em um aumento nas contratações. No entanto, dado que a maioria dessas vagas é temporária, é comum observar uma redução no nível de pessoal ocupado no período subsequente.

Dessa forma, é razoável antecipar variações sazonais na série em análise. Em termos mais simples, podemos inferir a presença de uma componente sazonal na série que está sendo examinada. Isso sugere a possibilidade de existirem padrões e variações regulares nos dados ao longo do tempo, associados a essa sazonalidade. Em face dessa realidade, não apenas estamos empregando modelos de previsão baseados em suavização exponencial para capturar as tendências subjacentes, mas também é crucial adotar metodologias específicas para lidar com a sazonalidade presente nos dados.

Ao contemplar tanto os aspectos de suavização exponencial quanto as técnicas apropriadas para a análise de séries temporais sazonais, podemos aprimorar significativamente nossa capacidade de prever com a maior precisão possível as futuras taxas de desocupação e obter uma compreensão mais aprofundada dos padrões sazonais que influenciam essas taxas.

3.3.3 Suavização exponencial sazonal de Holt-Winters

O método de suavização exponencial de Holt-Winters, aborda tanto a tendência quanto a sazonalidade na série temporal. É importante ressaltar que enfrentamos a limitação de contar com um número reduzido de observações, o que pode dificultar a identificação

clara da tendência e da sazonalidade nos dados, deixando-nos com predominância da componente aleatória. Para superar essa limitação, procuraremos discernir a tendência e a sazonalidade por meio de uma análise descritiva, e posteriormente modelaremos o Holt-Winters com base nas conclusões extraídas. Independentemente disso, forneceremos explicações abrangendo tanto o Holt-Winters sazonal quanto aquele que considera apenas a tendência.

De acordo com [Morettin e Tolo \(2006\)](#), há dois procedimentos cujo uso está condicionado às características da série em questão: seja a Série Sazonal Multiplicativa ou a Série Sazonal Aditiva.

(a) Série Sazonal Multiplicativa

Levando em consideração uma série sazonal com período s , a variante mais comum do método de Holt-Winters trata o fator sazonal F_t como multiplicativo, ao passo que a tendência permanece aditiva, isto é,

$$Z_t = \mu_t F_t + T_t + a_t, \quad t = 1, \dots, N, \quad (3.5)$$

em que,

- Z_t : Esta é a série temporal original, que representa as séries de gênero,
- μ_t : É a componente média, multiplicativa da série. Representa a variação multiplicativa da série temporal em relação à sazonalidade,
- F_t : É a componente sazonal multiplicativa, representando os padrões sazonais que se repetem em intervalos regulares,
- T_t : É a componente tendência da série,
- a_t : É o termo de ruído, que captura as flutuações aleatórias na série temporal.

Levando em consideração as três equações de suavização, temos:

$$\hat{F}_t = D \left(\frac{Z_t}{\bar{Z}_t} \right) + (1 - D) \hat{F}_{t-s}, \quad 0 < D < 1, \quad t = s + 1, \dots, N, \quad (3.6)$$

$$\bar{Z}_t = A \left(\frac{Z_t}{\hat{F}_{t-s}} \right) + (1 - A)(\hat{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad t = s + 1, \dots, N, \quad (3.7)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad t = s + 1, \dots, N, \quad (3.8)$$

e representam estimativas do fator sazonal, do nível e da tendência, respectivamente.

As constantes A , C e D são utilizadas para suavização.

Previsão

As previsões, segundo [Morettin e Tolo \(2006\)](#), são dadas:

Considerando \hat{Z}_t a previsão da série para valores futuros, temos:

$$\begin{aligned} \hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-s}, \quad h = 1, \dots, s, \\ \hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-2s}, \quad h = s + 1, \dots, 2s, \\ &\vdots \end{aligned} \quad (3.9)$$

em que,

- $Z_t(h)$: Esta é a série temporal original, que representa as séries de gênero, e é dada por [3.7](#)
- \hat{Z}_t : É a previsão da série para h períodos no futuro. No caso $h = 1, \dots, 4$, são os trimestres após 2020,
- \hat{T}_t : Componente tendência da previsão e é dada por [3.8](#),
- \hat{F}_t : Componente sazonal da previsão e é dada por [3.6](#),
- s : É a periodicidade sazonal, que representa o número de períodos em um ciclo sazonal completo,
- h : Representa o número de períodos no futuro para os quais estamos fazendo a previsão. No caso $h = 1, \dots, 4$. São os trimestres após 2020.

Para realizar atualizações nas previsões ao recebermos uma nova observação Z_{t+1} , empregamos as equações (3.6), (3.7) e (3.8). Assim,

$$\begin{aligned}\hat{F}_{t+1} &= D \left(\frac{Z_{t+1}}{\bar{Z}_{t+1}} \right) + (1 - D) \hat{F}_{t+1-s}, \\ \bar{Z}_{t+1} &= A \left(\frac{Z_{t+1}}{\hat{F}_{t+1-s}} \right) + (1 - A) (\hat{Z}_t + \hat{T}_t), \\ \hat{T}_{t+1} &= C (\bar{Z}_{t+1} - \bar{Z}_t) + (1 - C) \hat{T}_t,\end{aligned}$$

e a nova previsão para a observação Z_{t+h} será:

$$\begin{aligned}\hat{Z}_{t+1}(h-1) &= (\bar{Z}_{t+1} + (h-1)\hat{T}_{t+1})\hat{F}_{t+1+h-s}, \quad h = 1, \dots, s+1, \\ \hat{Z}_{t+1}(h-1) &= (\bar{Z}_{t+1} + (h-1)\hat{T}_{t+1})\hat{F}_{t+1+h-2s}, \quad h = 1, \dots, 2s+1, \\ &\vdots\end{aligned}\tag{3.10}$$

Os valores iniciais podem ser obtidos:

$$\hat{F}_j = \frac{Z_j}{\binom{s}{j} \sum_{k=1}^s X_k}, \quad j = 1, \dots, s; \quad \bar{Z}_s = \frac{1}{s} \sum_{k=1}^s Z_k; \quad \hat{T}_s = 0.$$

(b) Série Sazonal Aditiva

Nesse caso, temos que o fator sazonal será aditivo,

$$Z_t = \mu_t + F_t + T_t + a_t,\tag{3.11}$$

em que,

- Z_t : Esta é a série temporal original, que representa as séries de gênero,
- μ_t : É a componente nível da série. Representa o nível médio da série temporal ao longo do tempo,
- F_t : É a componente sazonal aditiva, representando os padrões sazonais que se repetem em intervalos regulares,
- T_t : É a componente tendência da série,
- a_t : É o termo de ruído, que captura as flutuações aleatórias na série temporal.

As estimativas do fator sazonal, nível e tendência da série são dadas por,

$$\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}, \quad 0 < D < 1, \quad (3.12)$$

$$\bar{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1 - A)(\hat{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad (3.13)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad (3.14)$$

respectivamente; A , C e D são constantes de suavização.

Previsão

Considerando \hat{Z}_t a previsão da série para valores futuros, temos:

$$\begin{aligned} \hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-s}, \quad h = 1, \dots, s \\ \hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-2s}, \quad h = 1, \dots, 2s \\ &\vdots \end{aligned} \quad (3.15)$$

Em que,

- \bar{Z}_t : É o nível da série suavizado e é dado em 3.13
- $\hat{Z}_t(h)$: É a previsão da série para h períodos no futuro. No caso $h = 1, \dots, 4$, são os trimestres após 2020.
- \hat{T}_t : Componente tendência e é dado em 3.14.
- \hat{F}_t : Componente sazonal e é dado em 3.12.
- s : É a periodicidade sazonal, que representa o número de períodos em um ciclo sazonal completo.
- h : Representa o número de períodos no futuro para os quais estamos fazendo a previsão. No caso $h = 1, \dots, 4$. São os trimestres após 2020.

Para fazermos atualizações das previsões, quando temos uma nova observação Z_{t+1} , utilizamos as equações 3.6, 3.7 e 3.8:

$$\hat{F}_{t+1} = D(Z_{t+1} - \bar{Z}_{t+1}) + (1 - D)\hat{F}_{t+1-s},$$

$$\begin{aligned}\bar{Z}_{t+1} &= A(Z_{t+1} - \hat{F}_{t+1-s}) + (1 - A)(\hat{Z}_t + \hat{T}_t), \\ \hat{T}_{t+1} &= C(Z_{t+1} - \bar{Z}_t) + (1 - C)\hat{T}_t,\end{aligned}$$

e a nova previsão para o valor Z_{t+h} ou seja, o valor da variável Z no tempo t , h passos futuros será:

$$\begin{aligned}\hat{Z}_{t+1}(h-1) &= \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-s}, & h = 1, \dots, s+1 \\ \hat{Z}_{t+1}(h-1) &= \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-2s}, & h = 1, \dots, 2s+1 \\ &\vdots\end{aligned}\tag{3.16}$$

Em que,

- $\bar{Z}_{t+1}(h-1)$: É o nível da série suavizado a h passos futuros.
- $\hat{Z}_t(h)$: É a previsão da série para h períodos no futuro. No caso $h = 1, \dots, 16$, são os trimestres após 2020.
- \hat{T}_t : Componente tendência da previsão.
- \hat{F}_t : Componente sazonal da previsão.
- s : É a periodicidade sazonal, que representa o número de períodos em um ciclo sazonal completo.
- h : Representa o número de períodos no futuro para os quais estamos fazendo a previsão. No caso $h = 1, \dots, 16$. São os trimestres após 2020.

Para avançar com este procedimento, é fundamental determinar se a série é aditiva, multiplicativa ou se exibe sazonalidade. Iniciaremos com uma análise descritiva para identificar indícios de sazonalidade nos dados. Se detectarmos sazonalidade, investigaremos se esta é de natureza multiplicativa ou aditiva.

3.3.4 Análise de intervenção

Segundo [Morettin e Tolo \(2006\)](#) por uma intervenção entendemos a ocorrência de algum tipo de evento em dado instante de tempo T , conhecido a priori e $1 < T < N$. Tal ocorrência pode manifestar-se por um intervalo de tempo subsequente e que afeta temporariamente, ou permanentemente, a série em estudo. A análise de intervenção tem

por objetivo avaliar o impacto de tal evento no comportamento da série.

Para o trabalho, iremos utilizar a variável binária chamada função impulso:

$$S_t^{(I)} = \begin{cases} 0, & \text{se } t \neq I, \\ 1, & \text{se } t = I. \end{cases} \quad (3.17)$$

Em que,

- $t, t=1, \dots, 32$, trimestres.
- I = período da pandemia que será considerado do primeiro trimestre de 2020 até o quarto trimestre de 2023.

No contexto da função 3.17, observamos que o efeito da intervenção é temporária a partir do instante I .

Ao aplicar a análise de intervenção, estamos interessados em entender como a pandemia de COVID-19 influenciou as taxas de desemprego, fazendo uma distinção entre homens e mulheres para compreender possíveis disparidades de impacto de gênero.

Capítulo 4

Análise estatística dos dados

4.1 Conhecendo a série

Com o objetivo de compreender o comportamento da série temporal, iremos criar um gráfico que representa as taxas de desemprego por gênero, destacando a intervenção ocorrida durante a pandemia. Dessa forma, buscamos analisar as variações no comportamento tanto antes quanto depois desse evento significativo.

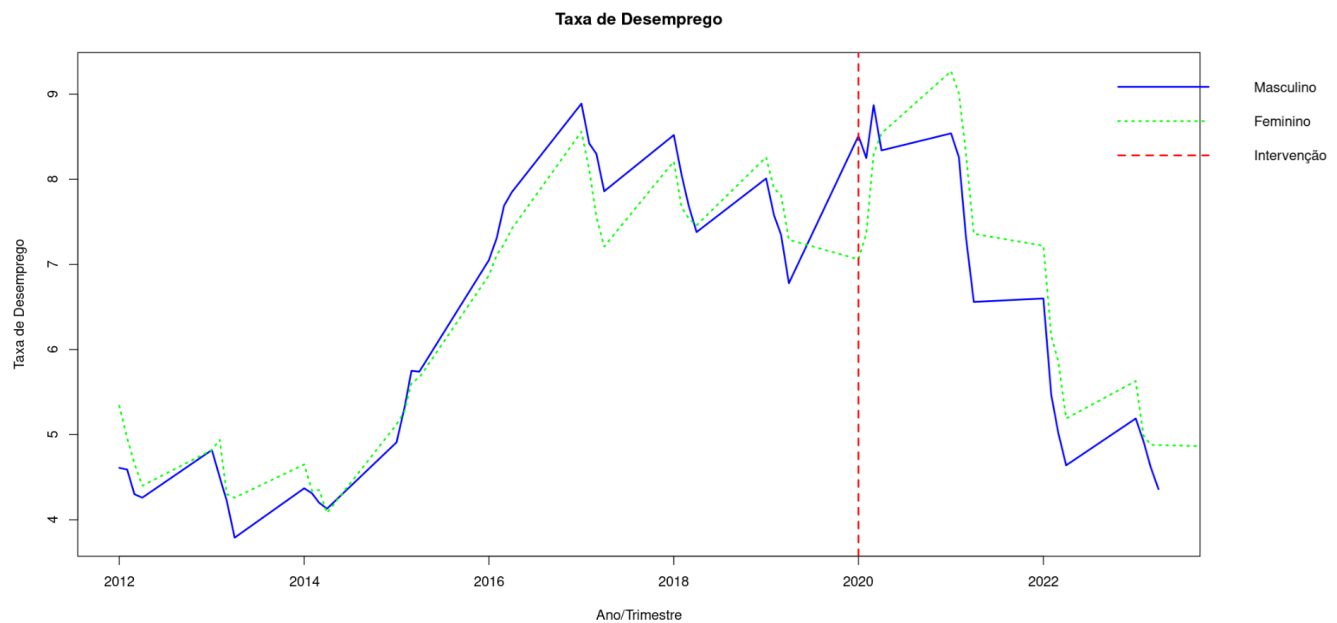


Figura 4.1: Série Temporal da taxa de desemprego dos gêneros masculino e feminino.

Pelos gráficos em 4.1, nota-se que as duas séries apresentam comportamentos semelhantes. Observa-se que, até 2014, ambas as séries mantêm um patamar mais baixo de desemprego, com o gênero feminino apresentando uma taxa de desemprego maior, embora ambas sigam a mesma tendência geral. De 2014 a 2017, há uma subida considerável nas taxas de desemprego. Posteriormente, entre 2016 e 2019, as taxas seguem uma ligeira tendência de queda. De 2019 a 2020, houve um aumento na taxa de desemprego, seguido por um decaimento considerável. Durante a maior parte das séries, é evidente que o público feminino apresentou taxas de desemprego mais altas do que os homens, exceção feita ao período entre os anos de 2015 e 2018.

4.1.1 Comportamento das séries antes da pandemia

Para uma compreensão abrangente das séries temporais de desemprego por gênero, examinaremos o seu comportamento ao longo do tempo. Especificamente, investigaremos o período que antecede a pandemia, definido neste trabalho como o intervalo do primeiro trimestre de 2012 até o quarto trimestre de 2019.

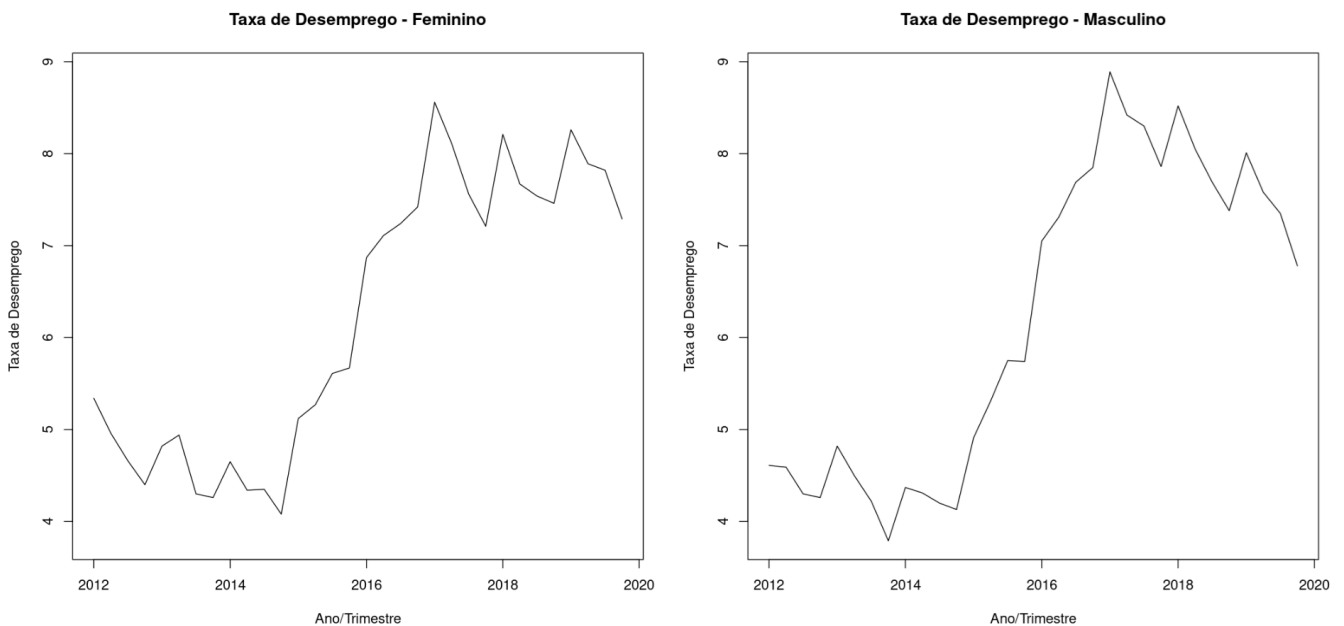


Figura 4.2: Série Temporal do desemprego dos gêneros feminino e masculino antes da pandemia.

Analisando os gráficos da Figura 4.2, que apresentam os dados antes da pandemia, é evidente a semelhança no formato das duas séries temporais. Dessa forma, ao aplicar o método de Holt-Winters, espera-se que ambas as séries tenham resultados semelhantes, isto é, que as previsões sejam similares para ambas as séries.

4.1.2 Análise descritiva das séries

Conforme previamente indicado, é necessário realizar uma análise mais detalhada da série para termos o indicativo se o método de Holt-Winters deve ser aplicado de forma multiplicativa ou aditiva. Para atingir esse objetivo, procederemos com uma análise descritiva das séries estatísticas de desemprego para os gêneros feminino e masculino anteriores à pandemia. É importante destacar que, por meio da análise descritiva, obteremos uma compreensão inicial do comportamento da série temporal. No entanto, para uma investigação mais abrangente, é necessário empregar técnicas além da análise descritiva. Neste contexto, a análise descritiva das séries nos proporcionará insights para determinar se devemos aplicar o método de Holt-Winters de forma multiplicativa ou aditiva.

4.1.3 Estacionariedade

Antes de apresentar a estacionariedade, vamos aprender brevemente sobre o que é um processo estocástico.

Processo estocástico

Considerando elementos arbitrários t_1, t_2, \dots, t_n pertencentes a um conjunto T , e definindo

$$F(z_1, z_2, \dots, z_n; t_1, t_2, \dots, t_n) = P\{Z(t_1) \leq z_1, Z(t_2) \leq z_2, \dots, Z(t_n) \leq z_n\}. \quad (4.1)$$

O processo estocástico $Z = \{Z(t), t \in T\}$ será completamente especificado quando soubermos as distribuições finito-dimensionais (4.1) para todos $n \geq 1$. Isso implica que, para $n = 1$, conhecemos as distribuições unidimensionais da variável aleatória $Z(t_1)$, onde $t_1 \in T$. Para $n = 2$, conhecemos as distribuições bidimensionais da variável aleatória $(Z(t_1), Z(t_2))$, onde $t_1, t_2 \in T$, e assim por diante.

Estacionariedade

Um processo estocástico Z é considerado estacionário se sua evolução ao longo do tempo não é influenciada pela escolha da origem dos tempos. Em outras palavras, as características de $Z(t + \tau)$ para qualquer τ são equivalentes às de $Z(t)$.

A importância da estacionariedade é evidente por várias razões. Em séries temporais estacionárias, os comportamentos estatísticos tornam-se mais previsíveis e uniformes ao longo do tempo, facilitando a aplicação de modelos estatísticos. A regularidade nesse contexto simplifica a análise estatística, permitindo o uso de métodos que dependem da

estacionariedade para gerar resultados confiáveis.

A estacionariedade indica que as características, como média e variância, da série temporal permanecem relativamente constantes ao longo do tempo, o que é benéfico para a realização de previsões. A estabilidade proporcionada pela estacionariedade torna desafiadora a previsão de padrões futuros quando as propriedades da série estão em constante alteração.

Além disso, a estacionariedade sugere que tanto a média quanto a variância da série temporal se mantêm constantes, ao longo do tempo simplificando a análise estatística. Esse aspecto viabiliza a aplicação de métodos mais diretos, como médias móveis e suavização exponencial.

A presença de estacionariedade simplifica a interpretação dos resultados, pois os parâmetros do modelo podem ser mais facilmente correlacionados com os aspectos do comportamento temporal da série. Em contrapartida, séries temporais sem estacionariedade podem resultar em estimativas de parâmetros enviesadas, levando a previsões imprecisas e interpretações incorretas.

Para aproveitar os benefícios da estacionariedade, será realizada uma análise gráfica para verificar se as séries de desemprego por gênero são estacionárias.

Para este trabalho vamos usar a estacionariedade fraca. É definido em [Ferreira et al. \(2017\)](#) que uma série fracamente estacionária apresenta propriedades estatísticas homogêneas ao longo do tempo, com expectativa média constante e covariância dependente apenas da defasagem temporal.

4.1.4 Tendência

Para que a estacionariedade ocorra, é necessário que a série não apresente tendência.

Em [Morettin e Tolo \(2006\)](#) o método somente com tendência é:

$$Z_t = T_t + a_t, \quad (4.2)$$

em que a_t é ruído branco com variância σ_a^2 .

Conforme definido por [Shumway e Stoffer \(2011\)](#), o ruído branco é um exemplo simples de série gerada, composta por sequências não correlacionadas representadas por variáveis aleatórias, Z_t . Essas variáveis possuem média zero e uma variância finita σ_Z^2 . Ampla e frequentemente utilizada como modelo para o ruído em várias aplicações de engenharia.

A denominação “branco” é uma metáfora baseada na analogia com a luz branca, sugerindo que todas as possíveis oscilações estão presentes com igual intensidade.

Estimando-se a tendência através de \hat{T}_t , podemos obter a série ajustada para a tendência ou livre dela.

$$Y_t = Z_t - Z_{t-1} \quad (4.3)$$

Um procedimento que é também utilizado para eliminar a tendência de uma série é aquele de tomar diferenças.

Podemos fazer diferenças em sequência até eliminar a tendência, porém ao fazer várias, podemos adicionar correlação parcial.

Se com uma diferença já obtivermos a estacionariedade, temos a equação:

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (4.4)$$

Para utilizar o procedimento de identificação da tendência, vamos usar as taxas de desemprego antes da pandemia, ou seja, do primeiro trimestre de 2012 até o quarto trimestre de 2019.

Graficamente, podemos usar a função de autocorrelação.

Segundo [Hyndman e Athanasopoulos \(2021\)](#) assim como a correlação mensura a extensão de uma relação linear entre duas variáveis, a autocorrelação avalia a relação linear entre os valores defasados de uma série temporal.

Diversos coeficientes de autocorrelação estão associados a cada painel no gráfico de defasagem. Por exemplo, R_1 representa a relação entre Z_t e Z_{t-1} , R_2 representa a relação entre Z_t e Z_{t-2} , e assim por diante.

O valor de R_k pode ser expresso como

$$R_k = \frac{\sum_{t=k+1}^T (Z_t - \bar{Z}_t)(Z_{t-k} - \bar{Z}_t)}{\sum_{t=1}^T (Z_t - \bar{Z}_t)^2},$$

em que T é o comprimento da série temporal e k denota a defasagem. Estes coeficientes de autocorrelação compõem a função de autocorrelação, também conhecida como ACF (*Autocorrelation Function*).

Quando uma série temporal apresenta uma tendência, é comum observar autocorrelações significativas e positivas para pequenas defasagens. Isso ocorre porque as ob-

servações próximas no tempo tendem a ter valores semelhantes, contribuindo para uma correlação positiva. Dessa forma, a Função de Autocorrelação (ACF) de uma série temporal com tendência geralmente exibe valores positivos que diminuem gradualmente à medida que as defasagens aumentam.

Para as nossas séries de desemprego por gênero, obtemos os seguintes gráficos:

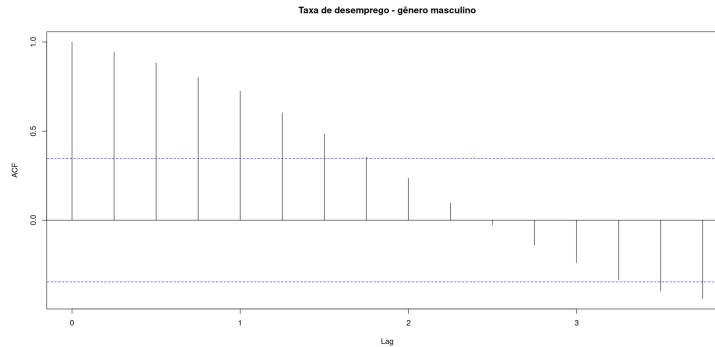


Figura 4.3: ACF da Série Temporal do desemprego do gênero masculino antes da pandemia.

Na Figura 4.3, observamos que o gráfico de autocorrelação parcial apresenta uma tendência, evidenciada pela significância de vários lags. Para mitigar essa tendência, aplicamos uma diferenciação e obtemos o gráfico subsequente.

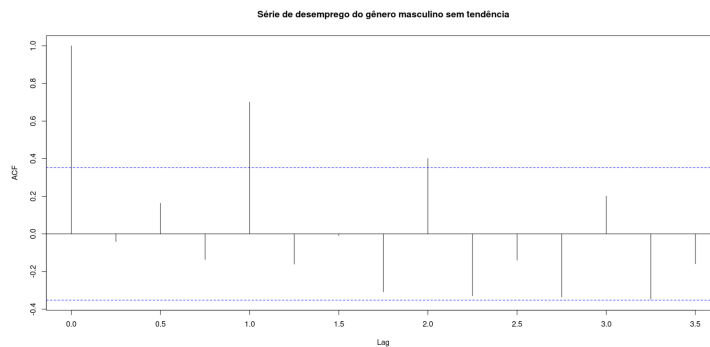


Figura 4.4: Aplicando a diferença no ACF da Série da Taxa de Desemprego do gênero Masculino antes da Pandemia

Pelo gráfico em 4.4, observamos que a tendência foi removida, uma vez que não há muitos lags subsequentes significativos. No entanto, após a remoção da tendência, tornou-se evidente a presença de sazonalidade na série. Isso é percebido pelo fato de que a cada quatro lags, temos um lag significativo. Essa observação era esperada, uma vez que nossos dados são trimestrais. No próximo tópico, iremos explorar mais a fundo o conceito de sazonalidade.

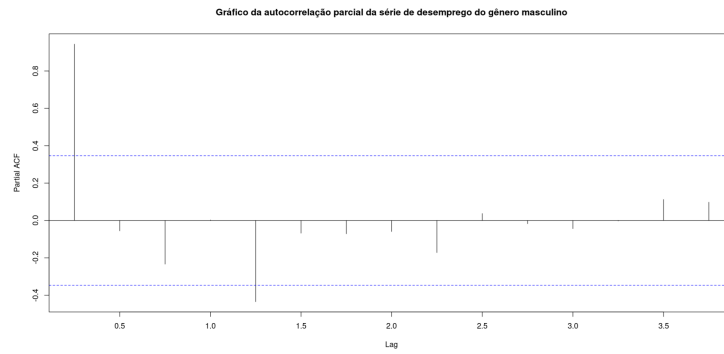


Figura 4.5: PACF da Série Temporal do desemprego do gênero feminino antes da pandemia.

No gráfico de autocorrelação parcial na Figura 4.5, observamos uma sazonalidade evidente, caracterizada pela significância do primeiro lag e, posteriormente, a cada quatro lags, temos outro lag maior que os outros.

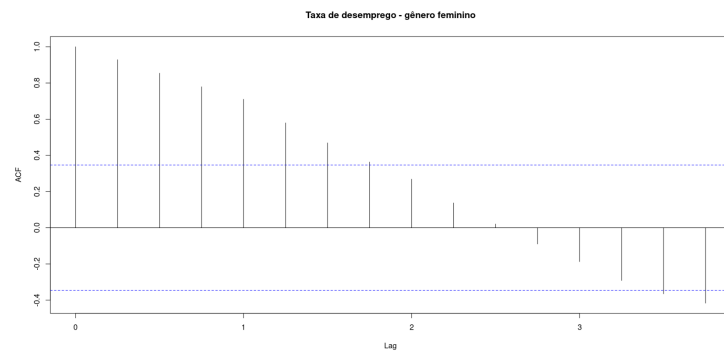


Figura 4.6: ACF da Série Temporal do desemprego do gênero feminino antes da pandemia.

A interpretação para o gráfico em 4.6 é semelhante ao gráfico de autocorrelação parcial do gênero masculino. Ou seja, a série também apresenta tendência. Portanto, vamos aplicar uma diferenciação para remover essa tendência.

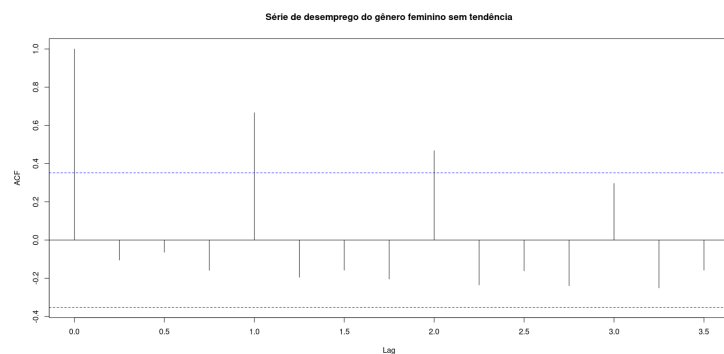


Figura 4.7: Aplicando a diferença no ACF da Série da Taxa de Desemprego do gênero Feminino antes da Pandemia

Na Figura 4.7, a interpretação é semelhante à do gênero masculino. Após removermos a

tendência, conseguimos observar a sazonalidade presente na série. A cada 4 lags, podemos notar um crescimento significativo.

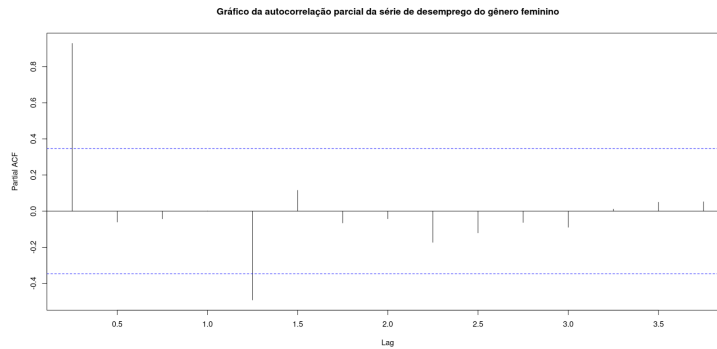


Figura 4.8: PACF da Série Temporal do desemprego do gênero feminino antes da pandemia.

Por fim, ao analisar o gráfico de autocorrelação parcial disposto na Figura 4.8 notamos a sazonalidade, onde a cada quatro lags, nota-se aumentos significativos.

4.1.5 Sazonalidade

Desejamos que as séries de desemprego por gênero se tornem estacionárias, o que implica a ausência de padrões sazonais.

Segundo Hyndman e Athanasopoulos (2021) quando lidamos com dados sazonais, é esperado que as autocorrelações sejam mais expressivas para os desfasamentos sazonais (múltiplos do período sazonal) do que para outros desfasamentos.

Ao examinar os gráficos apresentados nas Figuras 4.3,4.4,4.5,4.6,4.7 ,4.8, notamos que as séries que abordam tanto o gênero feminino quando o masculino possuem lags significativos, portanto, temos o indicativo de que as duas séries são sazonais.

4.1.6 Homocedasticidade

Para alcançarmos a estacionariedade em uma série, é necessário que ela seja homocedástica. Uma série temporal é considerada homocedástica quando a variabilidade dos dados permanece constante ao longo do tempo. Em outras palavras, a variação nos erros ou resíduos da série é uniforme em todas as observações. Essa característica implica que a amplitude das oscilações na série exibe variação sistemática à medida que a série progride temporalmente.

Graficamente, para a série de desemprego do gênero feminino temos a Figura 4.9 e

para a série de desemprego do gênero masculino a Figura 4.10. Em ambos os gráficos, o procedimento de verificar a homocedasticidade foi feita usando as séries diferenciadas.

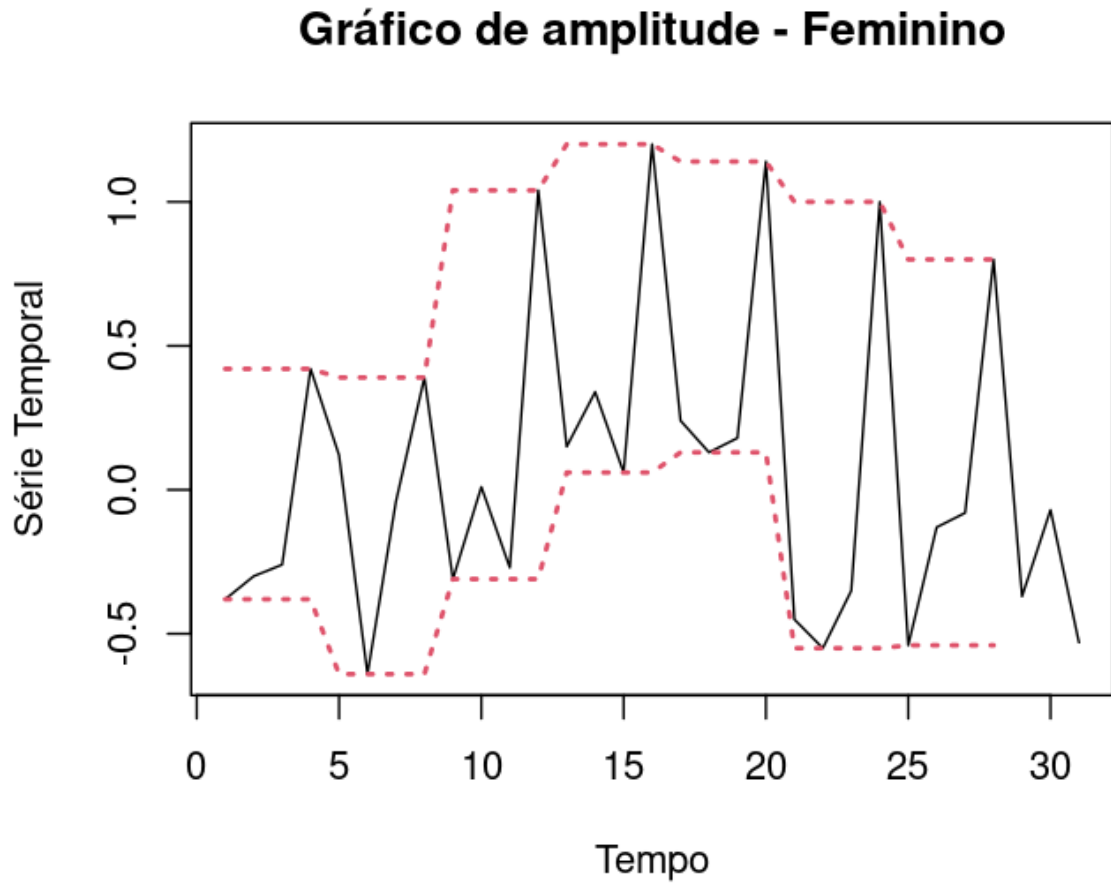


Figura 4.9: Gráfico da Série Temporal diferenciada do desemprego do gênero feminino antes da pandemia.

Gráfico de amplitude - Masculino

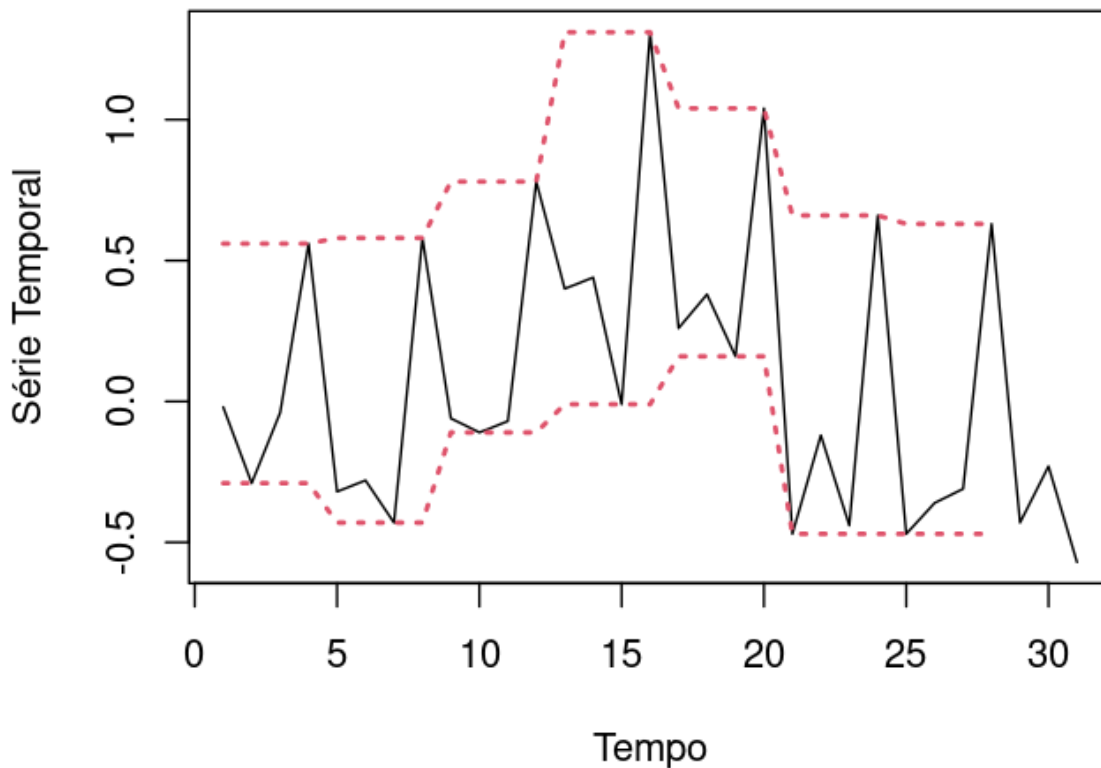


Figura 4.10: Gráfico da Série Temporal diferenciada do desemprego do gênero masculino antes da pandemia.

Ao analisarmos os gráficos referentes as Figuras 4.9 e 4.10, percebemos que, com o passar do tempo, a amplitude não sofre grandes alterações. As linhas pontilhadas em vermelho representam os maiores e menores valores das séries em cada ano, e a diferença entre esses valores define a amplitude. Como essa diferença não varia significativamente ao longo das séries, concluímos que as séries são homocedásticas.

Dado que ambas as séries de desemprego por gênero não apresentaram heterocedasticidade, conforme observado nas Figuras 4.9 e 4.10, podemos concluir que se tratam de séries temporais apenas com tendência e sazonalidade. Nesse contexto, aplicaremos o método de Holt-Winters com tendência e sazonalidade.

Capítulo 5

Aplicação

5.0.1 Método de Holt-Winters

Vamos aplicar o método exponencial de Holt-Winters apresentado na Subseção 3.3.3, em nossos dados. Para encontrar as constantes de suavização (A , C e D) usamos o mínimo da soma de quadrados da diferença entre valores observados e valores preditos.

Soma de quadrados: A equação para calcular a soma de quadrados total (SQT) em um conjunto de dados é:

$$SQT = \sum_{i=1}^n (x_i - \hat{x}_i), \quad (5.1)$$

em que:

- n é o número total de observações no conjunto de dados,
- x_i representa cada valor observado,
- \hat{x}_i são os valores preditos.

Essa equação calcula a diferença ao quadrado entre cada valor observado e a média dos valores, e então soma esses quadrados para todos os valores no conjunto de dados. Isso resulta na soma total de quadrados, que é uma medida da variabilidade total dos dados.

Buscamos o vetor (A , C , D) que minimize a soma dos quadrados. O vetor com os valores (0.87, 0.60, 1) resultou em uma soma de quadrados de 0.21 para a série da taxa de desemprego masculino, enquanto (0.66, 0.71, 0.91) resultou em 0.23 para a série da taxa de desemprego feminino.

Para a série da taxa de desemprego do gênero masculino, temos:

O coeficiente de suavização estimado para o nível, $\hat{A} = 0.87$, indica um valor alto que demonstra rápida adaptação para capturar séries com alta tendência. Quanto ao coeficiente de suavização estimado para a inclinação, $\hat{C} = 0.60$, observamos um valor elevado também, sugerindo mudanças frequentes na tendência, com os valores mais recentes exercendo significativa influência. O valor $\hat{D} = 1$ indica que a estimativa da sazonalidade exerce forte influência nas observações mais recentes.

Para a série da taxa de desemprego feminino, os valores são $\hat{A} = 0.66$, $\hat{C} = 0.71$ e $\hat{D} = 0.91$, onde podemos fazer uma interpretação semelhante à série da taxa de desemprego do gênero masculino.

Utilizamos o software *R*, previamente citado neste estudo, para obter esses valores.

5.0.2 Escolha do método

Para a seleção do método mais adequado, sazonal aditivo ou multiplicativo, será primeiramente descrito o conjunto de métricas que serão utilizadas para esta escolha.

Erro Médio Absoluto (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

em que,

- n é o número de observações.
- y_i é o valor real da observação i .
- \hat{y}_i é o valor previsto para a observação i .
- $|\cdot|$ representa o valor absoluto.

Erro Quadrático Médio (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

em que,

- n é o número de observações.
- y_i é o valor real da observação i .

- \hat{y}_i é o valor previsto para a observação i .

Erro Percentual Médio (MPE)

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i},$$

em que,

- n é o número de observações.
- y_i é o valor real da observação i .
- \hat{y}_i é o valor previsto para a observação i .

Erro Médio Absoluto de Escala (MASE)

$$MASE = \frac{\text{MAE}}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|},$$

em que,

- MAE é o Erro Médio Absoluto entre as previsões e os valores reais.
- n é o número de observações na série temporal.
- y_i é o valor real da observação i .
- y_{i-1} é o valor real da observação anterior à observação i .

Erro Percentual Absoluto Médio (MAPE)

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

em que,

- n é o número de observações.
- y_i é o valor real da observação i .
- \hat{y}_i é o valor previsto para a observação i .

Para a escolha do melhor método temos:

5.0.3 Previsão

Como já mencionado no trabalho, temos duas equações para seguirmos quando tomamos a série sazonal de Holt-Winters. Temos a **Série Sazonal Multiplicativa** e **Série Sazonal Aditiva**. A diferença fundamental entre esses dois métodos é a forma como eles lidam com a sazonalidade em relação à tendência. Para escolher o melhor método, vamos ajustar as séries para os dois casos e escolheremos o método conforme as métricas mencionadas.

Para a série de desemprego abordando o público feminino temos:

Tabela 5.1: Métricas para os métodos de Holt-Winters da série de desemprego feminino

Método	MSE	MAE	MPE	MAPE	MASE
Multiplicativo	0.23	0.18	0.45	3.14	0.28
Aditivo	0.26	0.20	0.22	3.48	0.31

Podemos observar nas métricas apresentadas na Tabela 5.1 que os métodos de Holt-Winters sazonal multiplicativo e aditivo indicam que, dentre esses métodos, o que apresenta o menor erro é o método sazonal multiplicativo. Portanto, adotaremos este método para analisar o restante da série temporal do desemprego feminino.

Agora analisando para a série de desemprego do gênero masculino:

Tabela 5.2: Métricas para os métodos de Holt-Winters da série de desemprego masculino

Método	MSE	MAE	MPE	MAPE	MASE
Aditivo	0.20	0.15	0.26	2.76	0.20
Multiplicativo	0.22	0.17	0.48	3.15	0.21

Percebemos através da Tabela 5.2 que as métricas dos métodos de Holt-Winters sazonal aditivo e multiplicativo mostram que, entre essas opções, o método de série sazonal aditivo é o que possui o menor erro. Assim, optaremos por utilizar este procedimento para examinar o restante da série temporal relacionada ao desemprego do grupo masculino.

Para realizar as previsões, começaremos coletando os dados e faremos a projeção para um ano à frente. Ou seja, vamos prever considerando $h = 4$, já que, como mencionado anteriormente, existem quatro trimestres por ano.

Tabela 5.3: Valores reais e estimados da série de desemprego do gênero feminino

Período t	Valor real Z_t	Valor estimado
2012.1	5.34	
2012.2	4.96	
2012.3	4.66	
2012.4	4.40	
2013.1	4.82	4.91
2013.2	4.94	4.93
2013.3	4.30	4.39
2013.4	4.26	4.01
2014.1	4.65	4.57
2014.2	4.34	4.85
2014.3	4.35	3.86
2014.4	4.08	4.13
2015.1	5.12	4.44
2015.2	5.27	5.23
2015.3	5.61	5.41
2015.4	5.67	5.62
2016.1	6.87	6.89
2016.2	7.11	7.08
2016.3	7.24	7.37
2016.4	7.42	7.17
2017.1	8.56	8.84
2017.2	8.11	8.76
2017.3	7.56	8.15
2017.4	7.21	7.18
2018.1	8.21	7.78
2018.2	7.67	7.69
2018.3	7.54	7.43
2018.4	7.46	7.35
2019.1	8.26	8.45
2019.2	7.89	7.81
2019.3	7.82	7.70
2019.4	7.29	7.68

A Tabela 5.3 fornece os números referentes à previsão antes da ocorrência da variável de intervenção, que é a pandemia. Podemos observar que não há uma diferença significativa em relação aos valores reais, o que era esperado, uma vez que o método de Holt-Winters tem uma boa capacidade de previsão para séries com um número limitado de dados.

Ao analisar a Tabela 5.4, observamos que os valores estimados para a série que representa o desemprego feminino foram bastante próximos dos valores reais. No entanto, há discrepância maior entre os valores estimados e reais quando comparados com os dados antes da pandemia.

Tabela 5.4: Valores reais e previsão da série de desemprego do gênero feminino

Período t	Valor Real Z_t	$\hat{Z}_t(h)$ $h=1,\dots,4$
2020.1	7.06	8.25
2020.2	7.36	7.80
2020.3	8.30	7.46
2020.4	8.54	7.11
2021.1	9.27	7.94
2021.2	9.01	7.52
2021.3	8.30	7.19
2021.4	7.36	6.85
2022.1	7.22	7.66
2022.2	6.16	7.24
2022.3	5.86	6.92
2022.4	5.19	6.59
2023.1	5.63	7.19
2023.2	4.99	6.89
2023.3	4.88	6.64
2023.4	4.85	6.38

Para as Tabelas 5.5 e 5.6, que apresentam os dados estimados e previsões para as séries de desemprego do gênero masculino, podemos fazer uma interpretação similar à do gênero feminino. Observamos valores próximos entre os estimados e os reais, assim como uma similaridade entre os valores previstos e os reais. Mas também há uma diferença maior dos valores previstos e reais depois da pandemia se comparados com os valores após a variável de intervenção.

Tabela 5.5: Valores reais e estimados da série de desemprego do gênero masculino

Período t	Valor real Z_t	Valor estimado
2012.1	4.61	
2012.2	4.59	
2012.3	4.30	
2012.4	4.26	
2013.1	4.82	4.82
2013.2	4.50	4.54
2013.3	4.22	4.17
2013.4	3.79	4.13
2014.1	4.37	4.20
2014.2	4.31	3.97
2014.3	4.20	4.06
2014.4	4.13	4.21
2015.1	4.91	4.87
2015.2	5.31	4.77
2015.3	5.75	5.34
2015.4	5.74	6.17
2016.1	7.05	6.82
2016.2	7.31	7.34
2016.3	7.69	7.48
2016.4	7.85	8.01
2017.1	8.89	9.10
2017.2	8.42	9.09
2017.3	8.30	8.24
2017.4	7.86	8.06
2018.1	8.52	8.55
2018.2	8.05	8.18
2018.3	7.69	7.72
2018.4	7.38	7.21
2019.1	8.01	8.03
2019.2	7.58	7.64
2019.3	7.35	7.28
2019.4	6.78	6.96

5.0.4 Gráficos da previsão

A seguir, apresentamos o gráfico das taxas de desemprego para ambos os gêneros, antes e durante a pandemia, bem como a previsão utilizando o método de Holt-Winters, incluindo o intervalo de confiança de 95%.

Foram realizados ajustes para os anos de 2020, 2021, 2022 e 2023 da seguinte forma: consideramos os dados até 2019 para prever os valores de 2020. Em seguida, incorporamos

Tabela 5.6: Valores reais e previsão da série de desemprego do gênero masculino

Período t	Valor Real Z_t	$\hat{Z}_t(h)$ $h=1,\dots,4$
2020.1	8.51	6.87
2020.2	8.25	6.56
2020.3	8.87	6.25
2020.4	8.87	5.95
2021.1	8.54	5.64
2021.2	8.26	5.34
2021.3	7.32	5.04
2021.4	6.56	4.74
2022.1	6.60	4.43
2022.2	5.47	4.13
2022.3	5.02	3.83
2022.4	4.64	3.53
2023.1	5.19	3.23
2023.2	4.92	2.93
2023.3	4.62	2.63
2023.4	4.36	2.33

os dados de 2020 e fizemos previsões para 2021. O mesmo processo foi repetido para 2022, utilizando os dados até 2021. Finalmente, incluímos os dados até 2022 e fizemos previsões para 2023.

Nos gráficos vamos representar a série real em azul, a previsão de Holt-Winters é mostrada em pontilhado vermelho e os valores reais da série são destacados em verde. O sombreado indica o intervalo com 95% de confiança

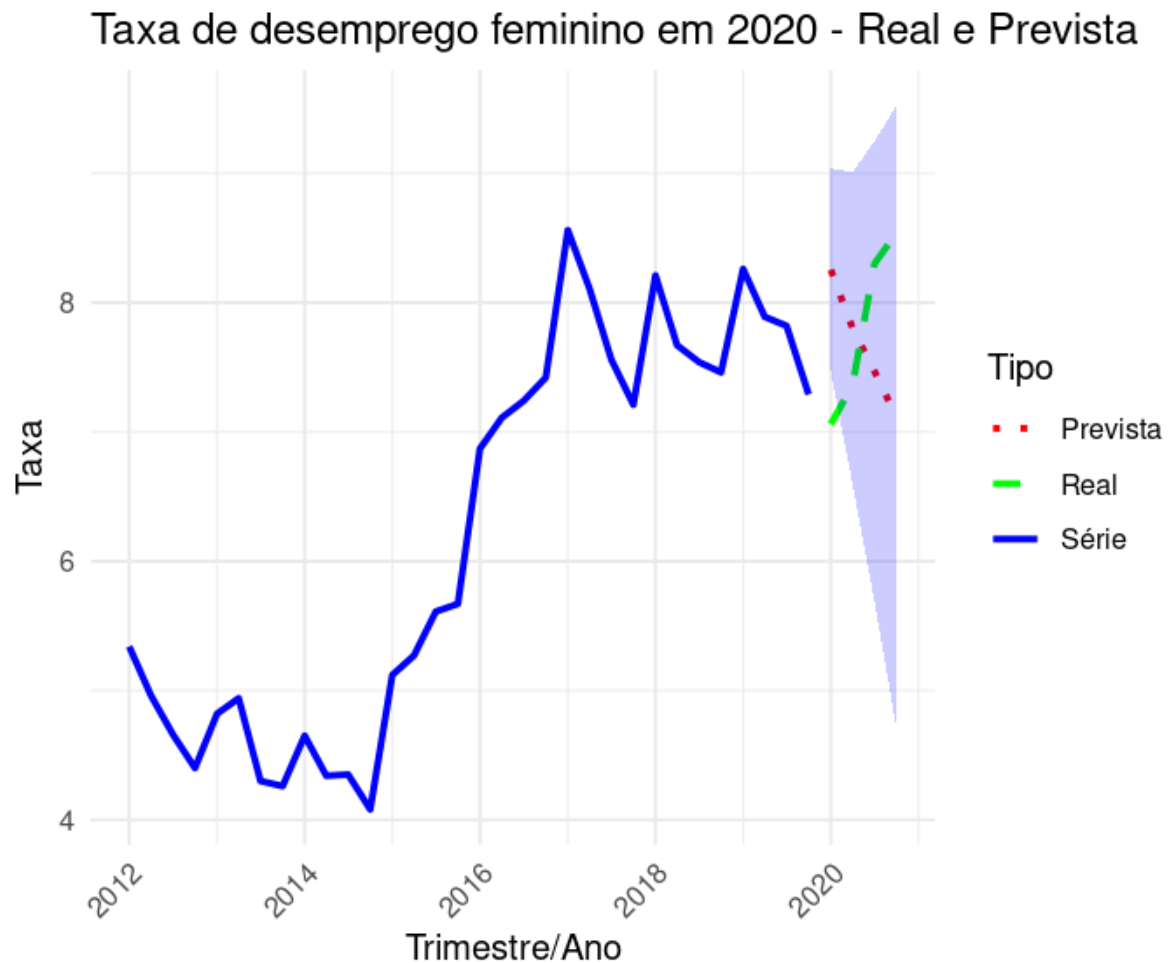


Figura 5.1: Gráfico da previsão da série de 2020 da taxa de desemprego feminino usando Holt-Winters

Observa-se na Figura 5.1 que a linha azul, representando a série temporal real da taxa de desemprego feminina até 2019, mostra uma tendência decrescente. Após o meio de 2018, houve um pico na taxa de desemprego, seguido por uma queda. No entanto, a linha verde indica que a série apresentou outro pico na taxa de desemprego. O método de Holt-Winters conseguiu prever esse pico, mas em situações opostas. A linha pontilhada vermelha, que representa a previsão do Holt-Winters, indicou que a taxa seria alta mas diminuiria em 2020. Na realidade, como mostrado pela linha pontilhada verde, a taxa estava baixa, mas houve um aumento. E isso pode ser explicado pela pandemia, que pode ter causado esse crescimento na taxa de desemprego.

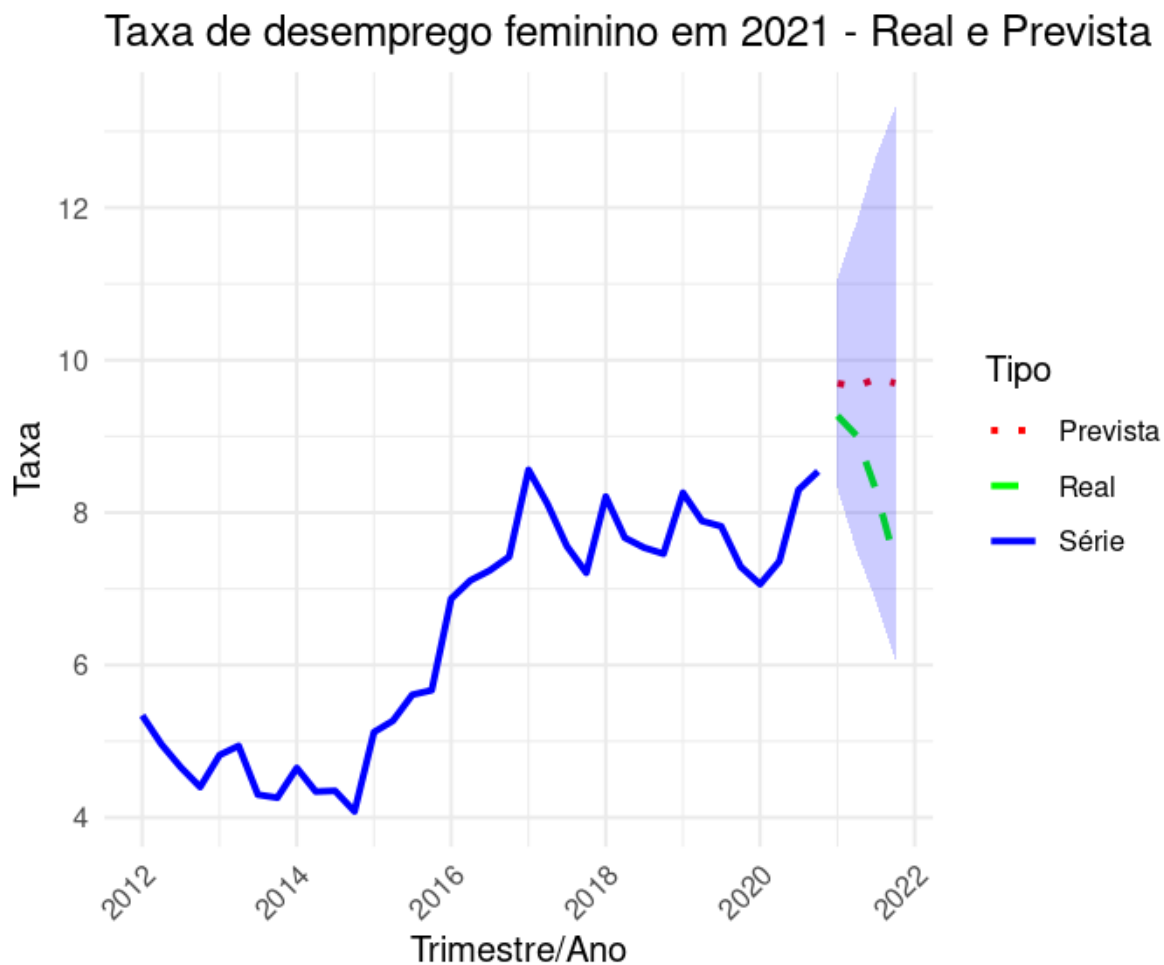


Figura 5.2: Gráfico da previsão da série de 2021 da taxa de desemprego feminino usando Holt-Winters

Na Figura 5.2, observamos que o método de Holt-Winters seguiu uma tendência mais linear, enquanto os valores reais mostraram uma queda na taxa de desemprego. No entanto, analisando a taxa de desemprego, percebemos que em 2021 ela atingiu os maiores níveis desde 2012, refletindo o reflexo da pandemia.

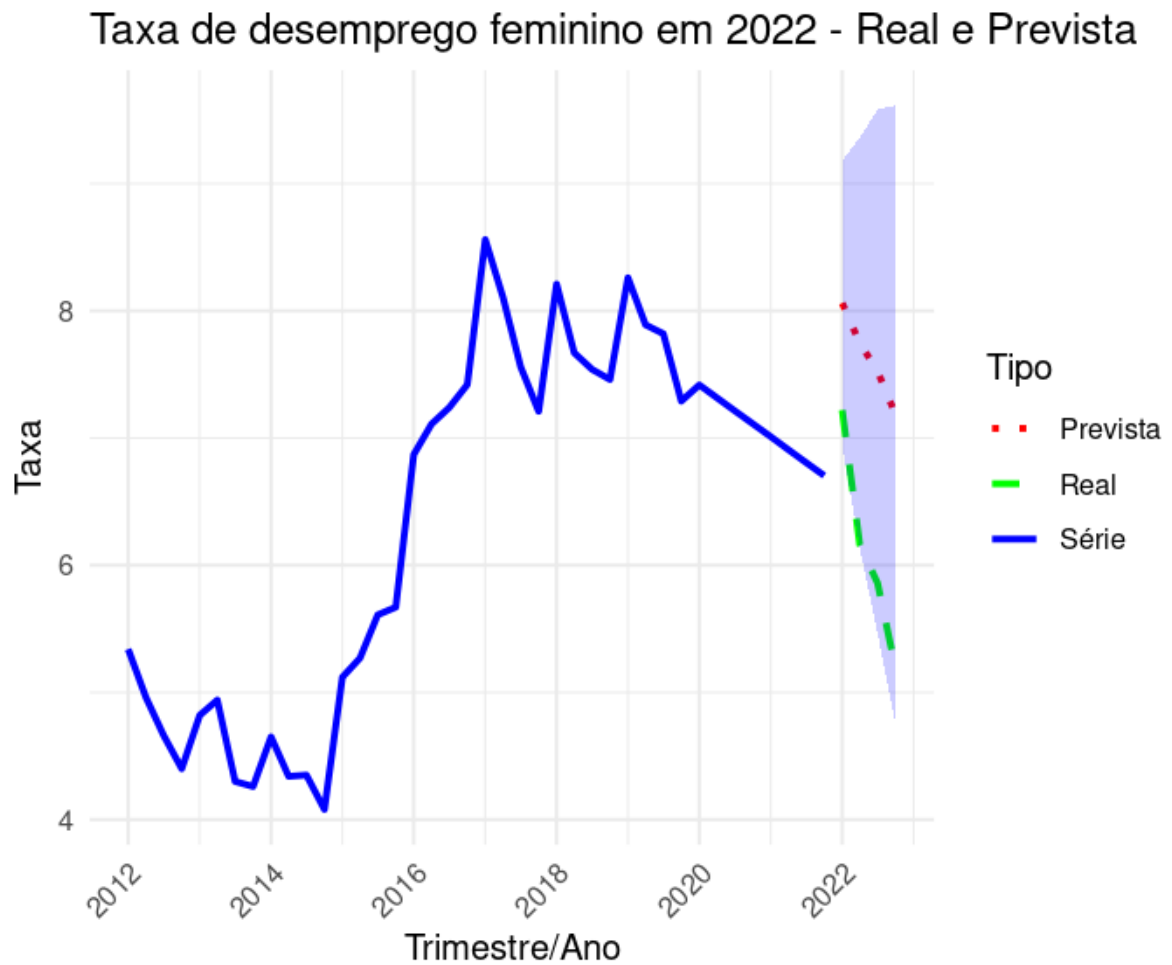


Figura 5.3: Gráfico da previsão da série de 2022 da taxa de desemprego feminino usando Holt-Winters

Na Figura 5.3, a previsão para 2022 é apresentada. Observando os valores reais, notamos que a taxa de desemprego diminuiu nesse ano. O método de Holt-Winters, que segue a tendência da série, ainda prevê taxas altas. No entanto, após dois anos de pandemia, as taxas de desemprego para o gênero feminino finalmente começaram a diminuir.

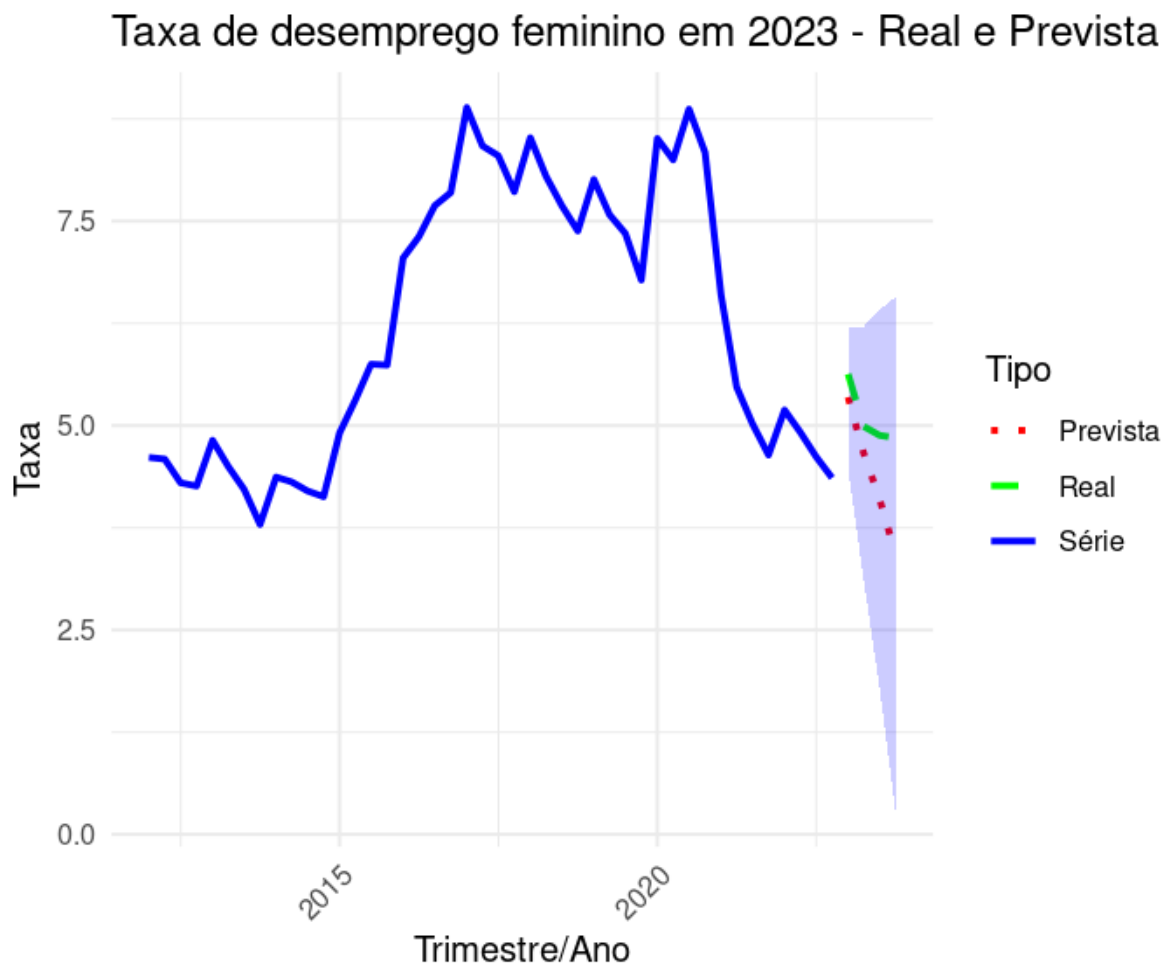


Figura 5.4: Gráfico da previsão da série de 2023 da taxa de desemprego feminino usando Holt-Winters

Diferente dos outros, o gráfico 5.4 mostra que a taxa real ficou próxima da previsão de Holt-Winters, que previu uma diminuição nas taxas, o que de fato ocorreu. Após três anos de pandemia, as taxas de desemprego do gênero feminino estão alcançando os menores níveis desde 2012, indicando que os efeitos da pandemia estão se tornando mais controláveis.

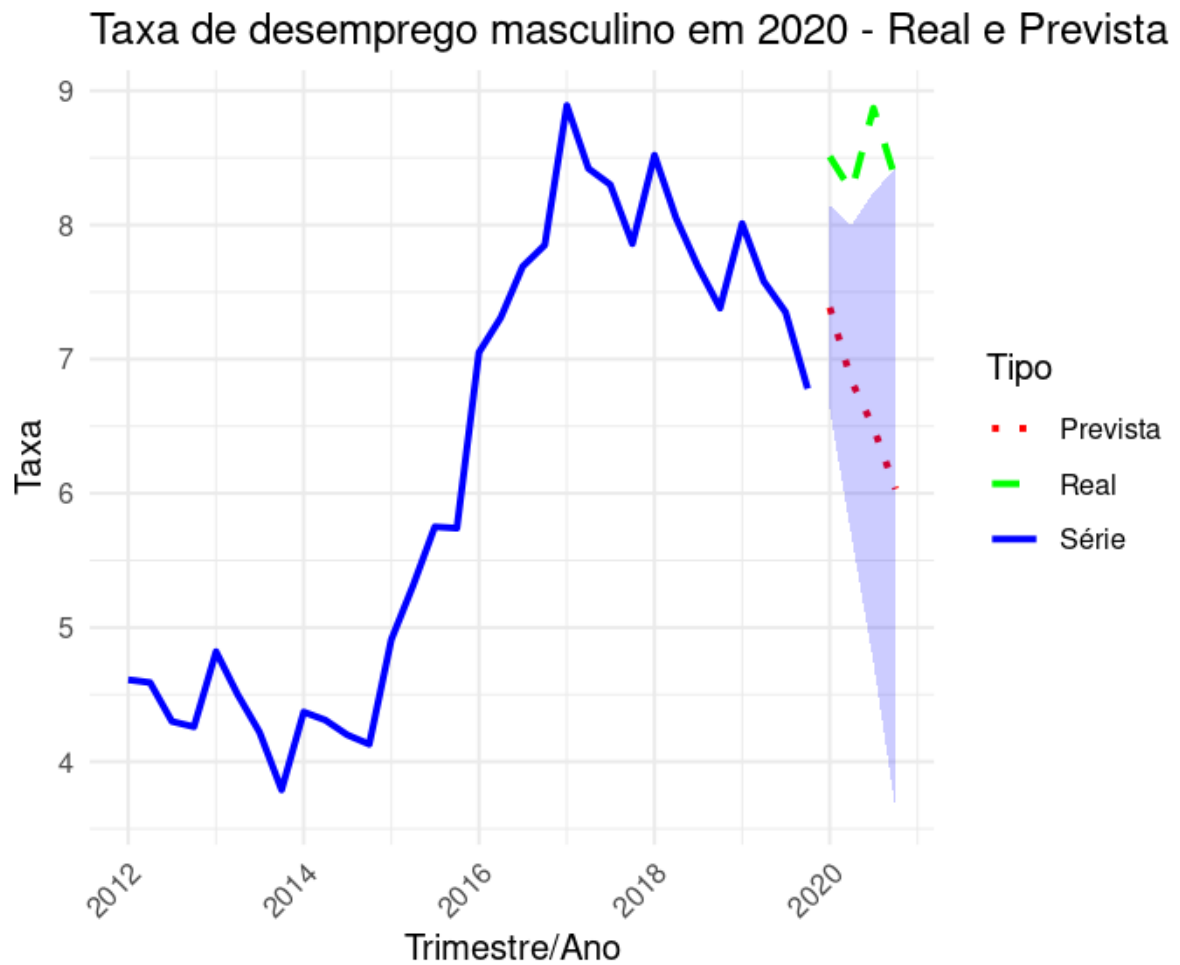


Figura 5.5: Gráfico da previsão da série de 2020 da taxa de desemprego masculino usando Holt-Winters

Ao examinarmos o gráfico mostrado na referência 5.5, percebemos que a taxa de desemprego real está significativamente acima da previsão de Holt-Winters. Isso ocorre porque o método seguiu a tendência decrescente da série, apesar dos picos cada vez menores, indicando que a taxa deveria continuar a cair. No entanto, a taxa real de desemprego foi bastante alta, quase alcançando o maior valor registrado em meados de 2017, o que pode ser atribuído ao primeiro ano da pandemia.

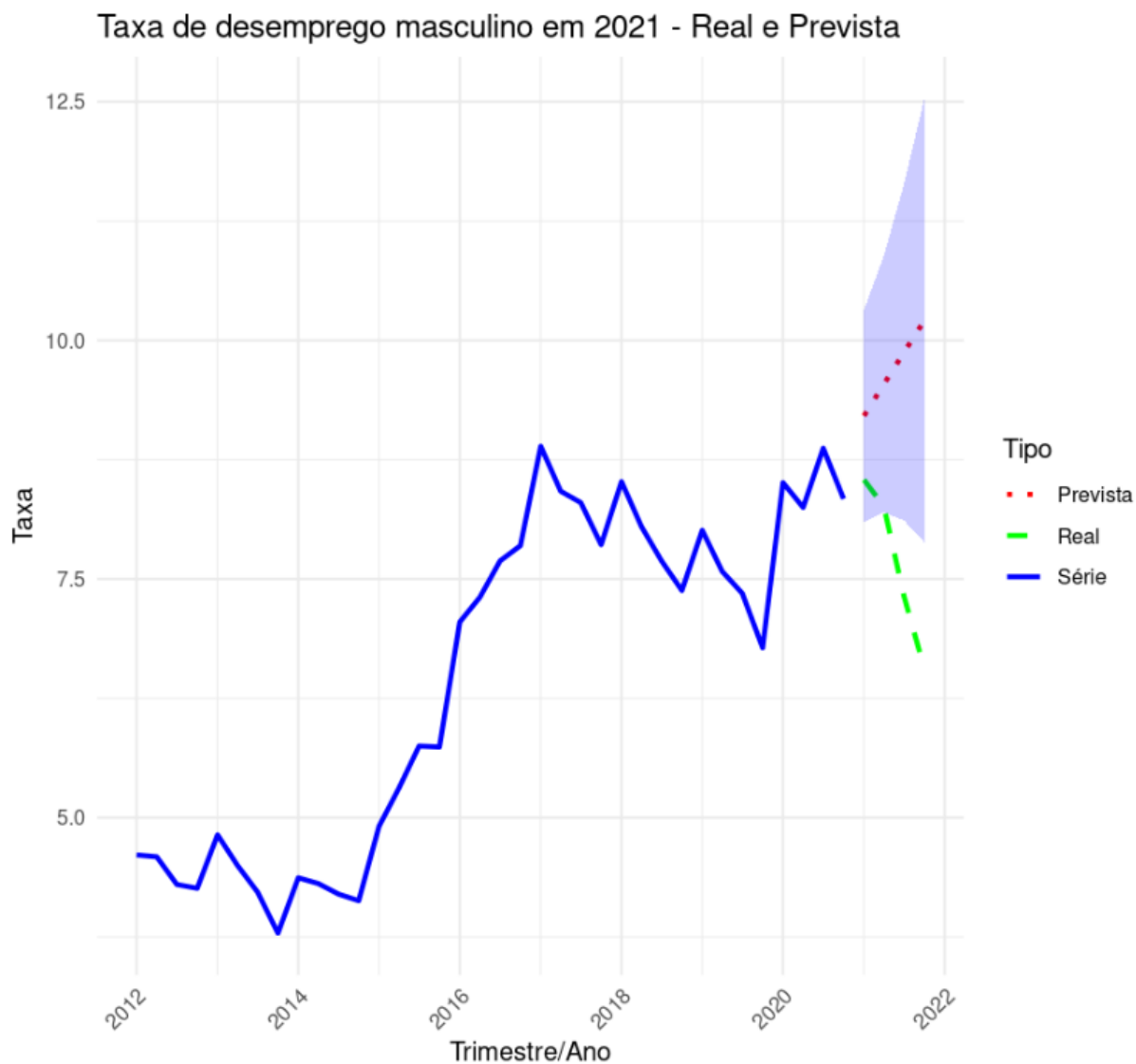


Figura 5.6: Gráfico da previsão da série de 2021 da taxa de desemprego masculino usando Holt-Winters

Ao analisarmos o gráfico em 5.6, observamos que a tendência de crescimento mencionada no gráfico anterior foi captada pelo método, que previu um aumento nas taxas para 2021. No entanto, o que ocorreu foi uma queda na taxa de desemprego. No segundo ano da pandemia, a taxa de desemprego do gênero masculino começou a mostrar menos reflexos negativos.

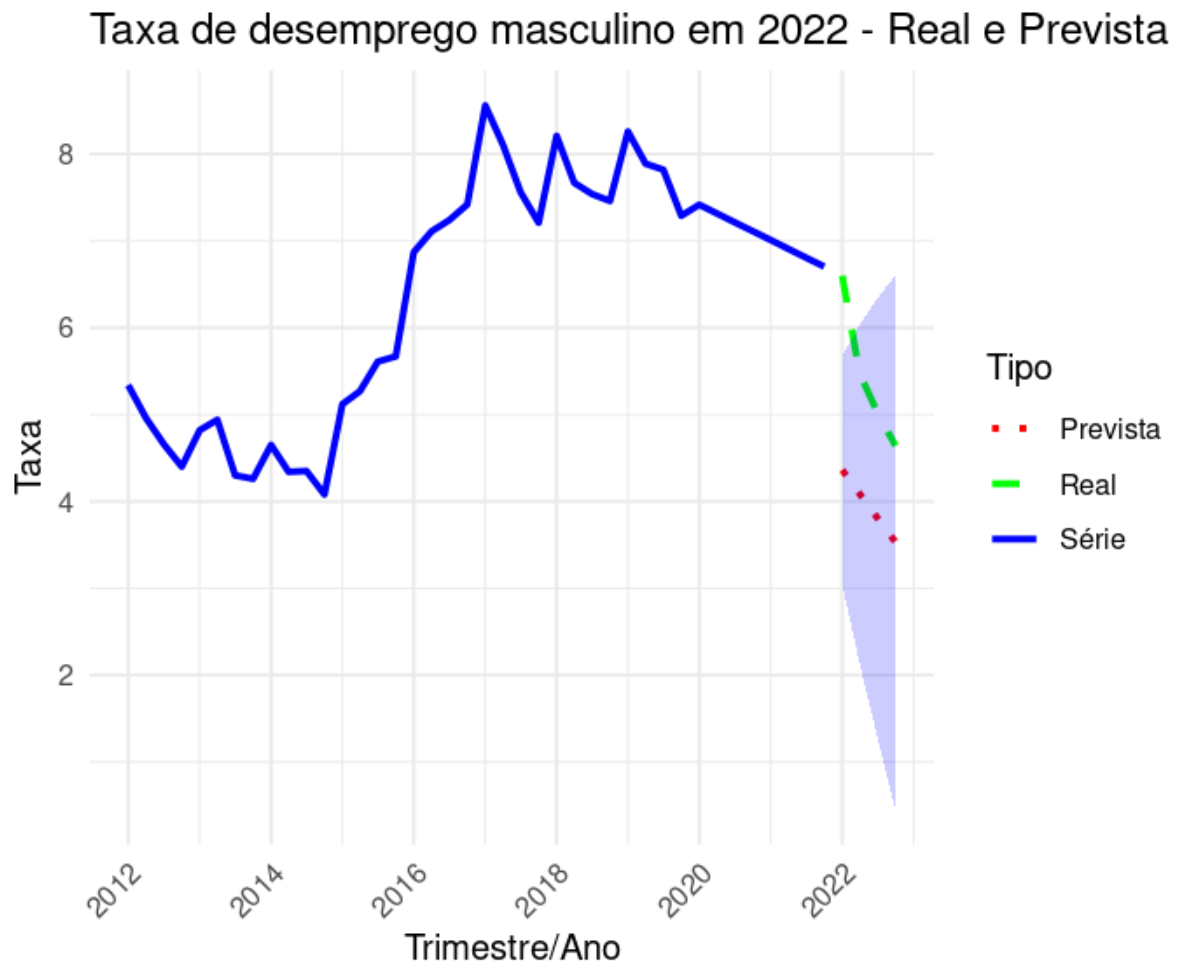


Figura 5.7: Gráfico da previsão da série de 2022 da taxa de desemprego masculino usando Holt-Winters

Ao analisarmos o gráfico apresentado na referência 5.7, observamos que o método Holt-Winters conseguiu captar a queda nas taxas vista no gráfico de 2021, seguindo um padrão semelhante às taxas reais. A única diferença foi que as taxas reais foram um pouco mais altas do que as previstas. Com essa diminuição, percebemos que os efeitos da pandemia nas taxas de desemprego do gênero masculino estão se dissipando.

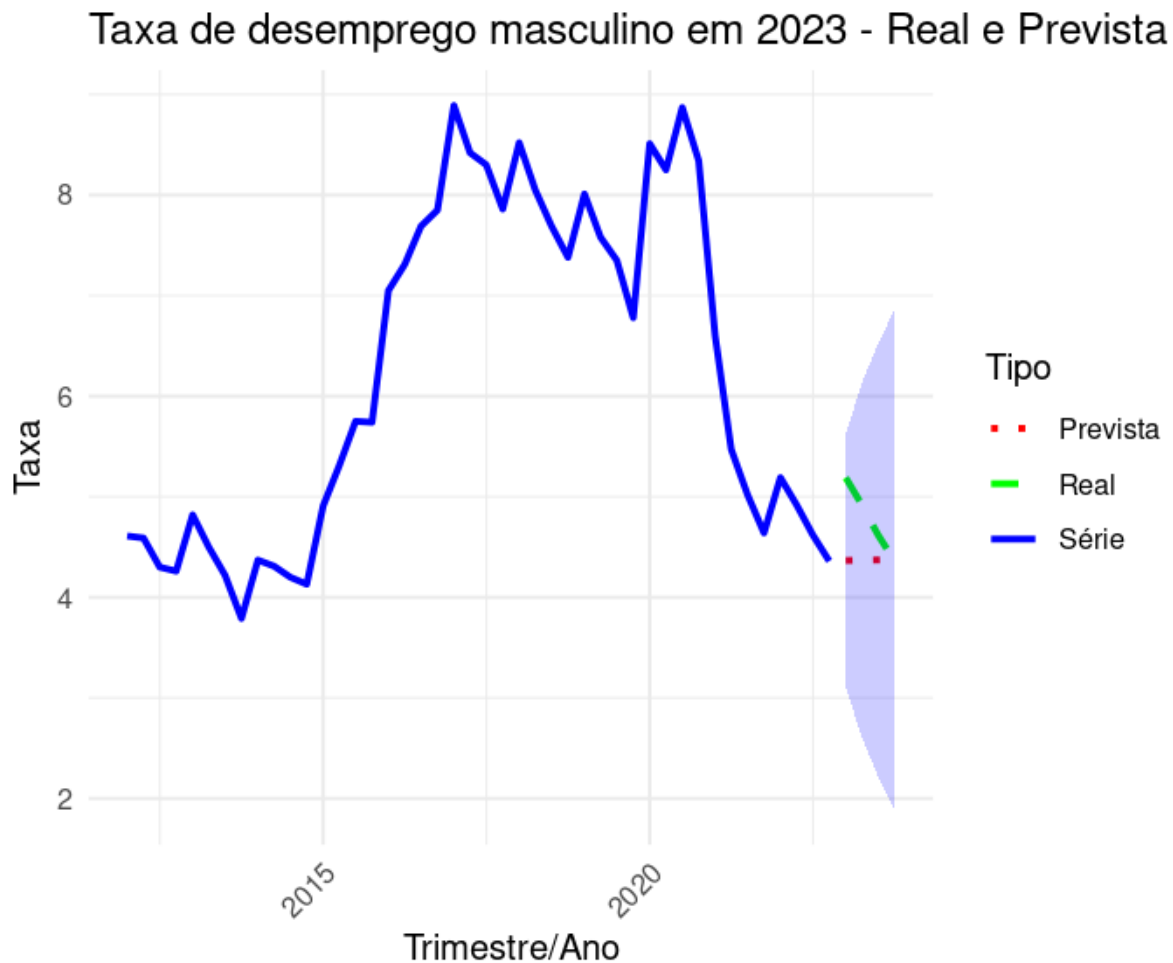


Figura 5.8: Gráfico da previsão da série de 2023 da taxa de desemprego masculino usando Holt-Winters

No último gráfico, referenciado como 5.8, observamos que a previsão foi mais linear, enquanto o resultado real mostrou um leve decréscimo. Mesmo assim, o método de Holt-Winters quase acertou na previsão. Com uma das menores taxas desde 2012, após três anos de pandemia, podemos concluir que não houve mais impactos significativos dela na taxa de desemprego masculino.

5.0.5 Conclusão

Neste trabalho, utilizamos o procedimento de Holt-Winters sazonal para especificar os métodos conforme desejávamos. Agora, com base nas métricas descritas (MAE, MSE, MPE, MASE e MAPE), identificamos qual foi o método que mais se adequou aos nossos dados.

Primeiro ano da pandemia ou seja, 2020

Temos as seguintes métricas:

Tabela 5.7: Comparação das métricas de métodos de previsão para dados masculinos e femininos no ano de 2020

Métrica	Masculino	Feminino
RMSE	0.45	0.51
MAE	0.37	0.41
MPE	-0.20	0.37
MAPE	6.22	6.61
MASE	0.47	0.63

Ao analisar a Tabela apresentada em 5.7, observamos que o método proposto apresenta menores erros nas métricas relacionadas à série da taxa de desemprego do gênero masculino, especialmente para o ano de 2020, indicando um desempenho superior nessa categoria.

Segundo ano da pandemia ou seja, 2021

Temos as seguintes métricas:

Tabela 5.8: Comparação das métricas de métodos de previsão para dados femininos e masculinos no ano de 2021

Métrica	Masculino	Feminino
RMSE	0.42	0.48
MAE	0.33	0.36
MPE	-0.18	0.33
MAPE	5.59	5.87
MASE	0.41	0.57

Na Tabela 5.8 observamos a mesma interpretação da tabela anterior, a série da taxa de desemprego do gênero masculino possui valores das métricas de erro menos do que no gênero feminino.

Terceiro ano da pandemia ou seja, 2022

Temos as seguintes métricas:

Tabela 5.9: Comparação das métricas de métodos de previsão para dados masculinos e femininos no ano de 2022

Métrica	Masculino	Feminino
RMSE	0.40	0.46
MAE	0.30	0.33
MPE	-0.15	0.30
MAPE	5.00	5.28
MASE	0.35	0.54

O mesmo se dispõe na Tabela 5.9 temos que as métricas de erro são menores na série da taxa de desemprego do gênero masculino.

Quarto ano da pandemia ou seja, 2023

Temos as seguintes métricas:

Tabela 5.10: Comparação das métricas de métodos de previsão para dados femininos e masculinos no ano de 2023

Métrica	Masculino	Feminino
RMSE	0.38	0.43
MAE	0.27	0.30
MPE	-0.14	0.27
MAPE	4.53	4.80
MASE	0.30	0.50

De modo similar, na tabela 5.10, observamos que os erros são menores na série da taxa de desemprego do gênero masculino.

Conclusão Geral

Com base na análise das métricas apresentadas nas tabelas, observamos que o método de Holt-Winters teve um desempenho superior ao prever a série da taxa de desemprego do gênero masculino. As métricas de erro indicam que o métodos errou menos para o gênero masculino, sugerindo uma maior precisão na previsão dos dados masculinos em comparação aos dados femininos. Portanto, podemos afirmar que as taxas de desemprego entre os homens foram previstas com maior exatidão do que as taxas de desemprego entre as mulheres, o que implica que os homens apresentaram taxas de desemprego efetivamente menores do que as das mulheres. E assim, temos que a pandemia afetou em sua maioria, as mulheres.

Analisando os gráficos de previsão, notamos que, devido à tendência e à sazonalidade das séries, apenas um ano após o início da pandemia não foi suficiente para o método de Holt-Winters captar com precisão os valores reais. Um destaque é o ano de 2020 para ambos os gêneros. Na Figura 5.1, que representa os valores da série do gênero feminino, observamos que o método Holt-Winters, ao seguir a tendência e sazonalidade dos dados, indica uma queda na taxa de desemprego entre as mulheres. No entanto, o que realmente ocorreu foi que a pandemia começou a afetar esse grupo, resultando em um aumento na taxa de desemprego.

Para o gênero masculino, representado na Figura 5.5, podemos notar que, embora o método Holt-Winters capture a tendência da série, a pandemia teve um impacto signifi-

cativo sobre os homens, elevando a taxa de desemprego a um nível bem mais alto, fora até do intervalo de confiança de 95%.

Por fim, em 2023, o método conseguiu aprender melhor a série e estava mais próximo dos valores reais. Além disso, observou-se que o efeito da pandemia estava diminuindo na taxa de desemprego tanto para o gênero masculino quanto para o feminino.

Como a definição de desemprego neste trabalho é baseada na busca por emprego, é possível que uma pessoa que anteriormente trabalhava ocasionalmente, ou até com bastante frequência, não procurasse emprego. No entanto, com as restrições de isolamento, essa pessoa pode ter passado a procurar emprego.

Capítulo 6

Conclusão

Este estudo apresentou métodos de Holt-Winters com características de tendência e sazonalidade, realizando uma análise das taxas de desemprego por gênero. Escolhemos esse método por sua robustez em situações com poucos dados disponíveis.

Nosso objetivo era concluir se a pandemia afetou a taxa de desemprego e identificar qual gênero foi mais impactado. Aplicamos o método multiplicativo às séries femininas e o método aditivo às séries masculinas, utilizando métricas de classificação definidas no texto para decidir essa abordagem. A análise foi atualizada anualmente para entender o comportamento durante o primeiro, segundo, terceiro e quarto ano após a pandemia.

Os gráficos de previsão utilizando o método de Holt-Winters indicaram que:

- Embora o método previsse a série conforme seu comportamento histórico, a pandemia influenciou alterações significativas, fazendo com que a taxa real de desemprego divergisse da prevista.
- No primeiro ano da pandemia, os homens foram os mais afetados, com uma taxa de desemprego real significativamente alta.
- No segundo ano, as taxas reais começaram a cair, mas as mulheres sofreram mais, atingindo a maior taxa de desemprego desde 2012.
- No terceiro e quarto ano, as taxas de desemprego diminuíram significativamente para ambos os gêneros. Contudo, as mulheres ainda enfrentaram maiores taxas de desemprego, embora a pandemia estivesse perdendo seu papel como principal fator.

Concluimos que, no geral, o público feminino foi mais impactado pelo desemprego após a pandemia, embora os homens tenham sofrido taxas recordes no primeiro ano.

Como mencionado anteriormente, isso pode ter ocorrido devido à busca de emprego por parte de pessoas que perderam trabalhos não fixos, como freelancers, cujos locais de trabalho podem ter fechado durante a pandemia. Como resultado, essas pessoas passaram a procurar emprego ativamente ao serem entrevistadas, resultando na classificação como desempregadas.

Para pesquisas futuras, propomos segmentar o público por classes sociais para identificar quais delas foram mais atingidas pelo desemprego. Uma hipótese possível é que, após a pandemia, o desemprego afete prioritariamente as classes sociais mais pobres.

Referências Bibliográficas

- Almeida de Castro, E. (2022). Análise dos indicadores de desemprego e trabalho remoto durante a pandemia de covid-19 na américa latina, com ênfase nos grupos de maior vulnerabilidade.
- Fernandes Silva, R. (2023). O impacto da pandemia no mercado de trabalho nordestino no ano de 2020.
- Ferreira, P. C., Mattos, Daiane Marcolino de Oliveira, I. C. L. d., Lima, L. F. e Duca, V. E. L. d. A. (2017). *Análise de Séries Temporais em R: Um Curso Introductório*. GEN Atlas.
- Hyndman, R. J. e Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. Otexts.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2023a). Desemprego. <https://www.ibge.gov.br/explica/desemprego.php>.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2023b). Pesquisa Nacional por Amostra de Domicílio Contínua (PNAD Contínua). <https://www.ibge.gov.br/estatisticas/sociais/populacao/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?edicao=37665>.
- Morettin, P. A. e Toloi, C. M. C. (2006). *Análise de séries Temporais*. ABE - Projeto Fisher.
- Organização Pan-Americana de Saúde (Acessado em 2023). Histórico da pandemia de COVID-19. <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19#:~:text=Em%2011%20de%20mar%C3%A7o%20de,pa%C3%ADses%20e%20regi%C3%B5es%20do%20mundo>.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shumway, R. H. e Stoffer, D. S. (2011). *Time Series Analysis and Its Applications: With R Examples*. Springer.