

# RELATÓRIO DESOCUPAÇÃO

Évelyn Muniz

## SUMÁRIO

<b>1 INTRODUÇÃO E OBJETIVOS:</b>	<b>1</b>
<b>2 METODOLOGIA E RESULTADOS:</b>	<b>1</b>
2.1 Análise Exploratória dos Dados: . . . . .	1
2.2 Modelos: . . . . .	5
2.3 Análise de Resíduos: . . . . .	8
<b>3 SIMULAÇÕES:</b>	<b>11</b>

## 1 INTRODUÇÃO E OBJETIVOS:

O presente relatório visa demonstrar a análise exploratória de dados de Taxa de Desocupação do Total Brasil, no período de janeiro de 2012 ao quarto trimestre de 2022, bem como descrever as componentes da série, pelo método da decomposição.

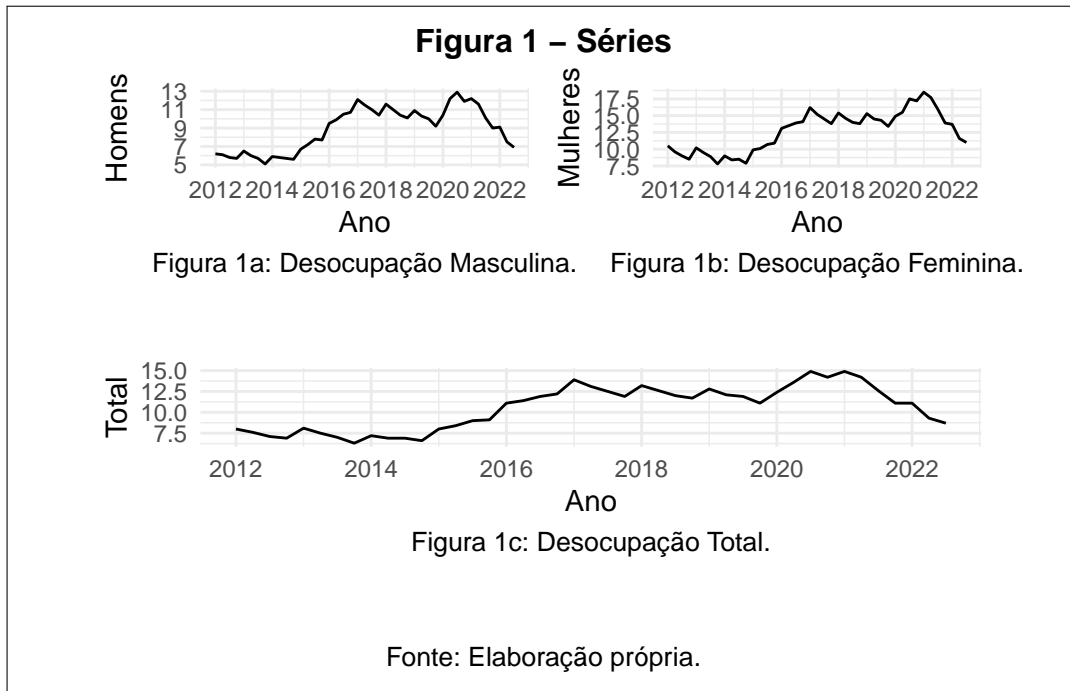
## 2 METODOLOGIA E RESULTADOS:

### 2.1 Análise Exploratória dos Dados:

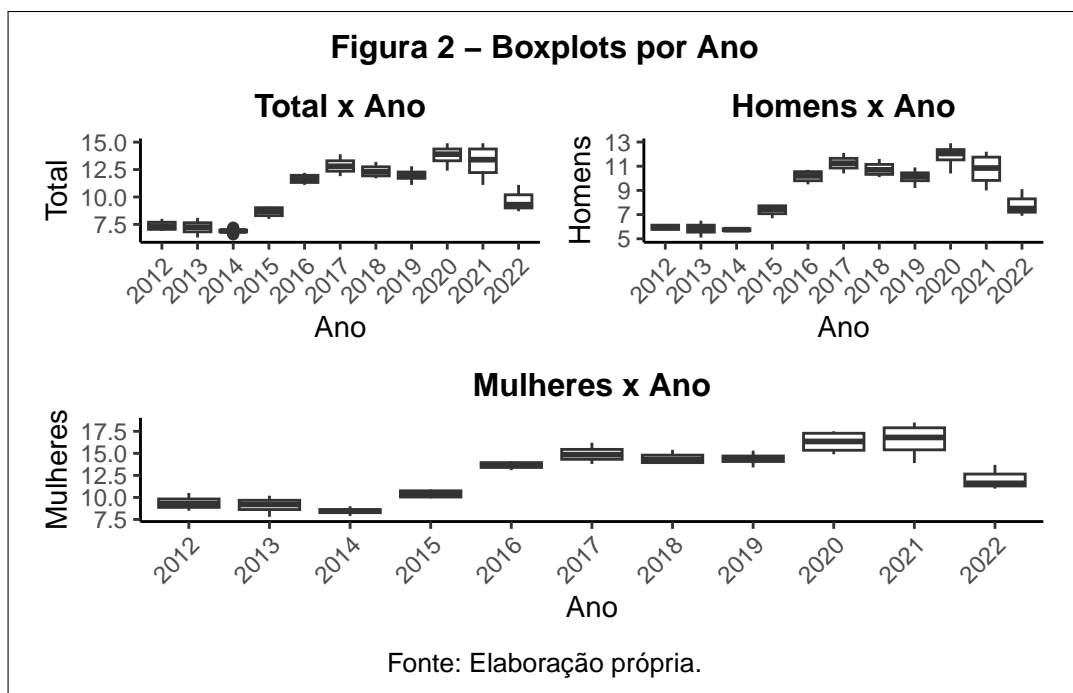
O banco de dados contém 5 colunas principais, a primeira é o ano, a segunda é o trimestre, a terceira é o valor da taxa por homens, a quarta é o valor da taxa por mulheres e a quinta representa a taxa total de desocupação.

Inicialmente, foi plotado o gráfico separado por gênero e total. A frequência observada é igual a 4, uma vez que os dados são trimestrais e o período inicial primeiro trimestre de 2012. Abaixo, a FIGURA 1a, FIGURA 1b e FIGURA 1c, em que é possível notar um crescimento entre os anos de 2014 e 2017, em 2018 inicia um decaimento da taxa que volta a crescer no final de 2019

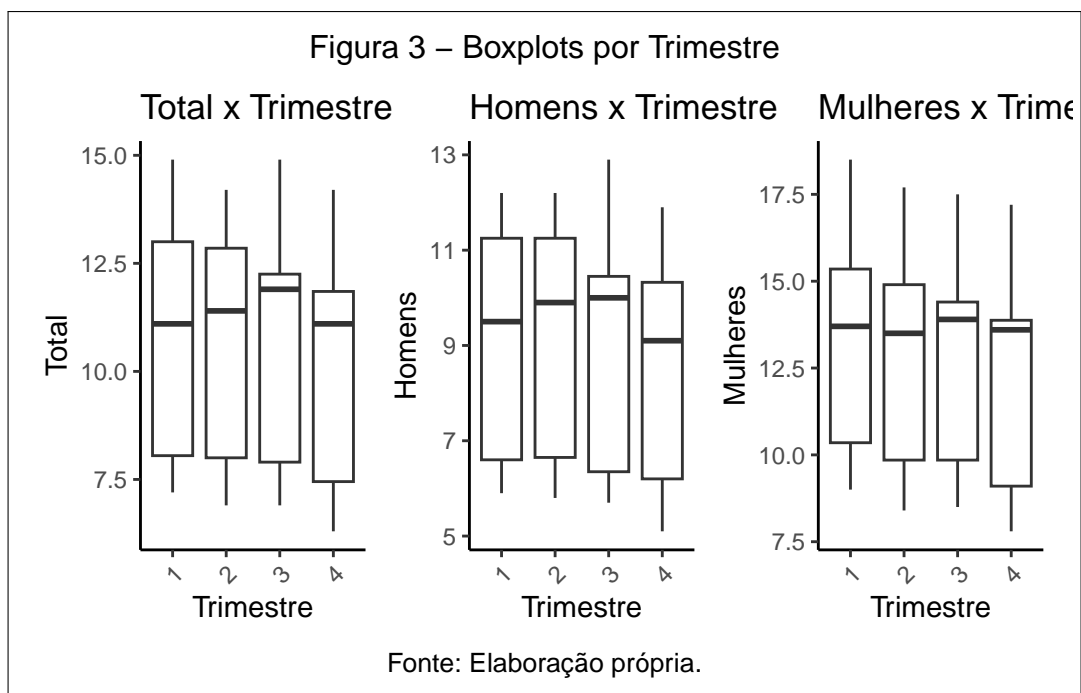
e entra em queda novamente no final de 2021. O comportamento foi o mesmo para homens, mulheres e a taxa total, indicando também a não-estacionariedade da série.



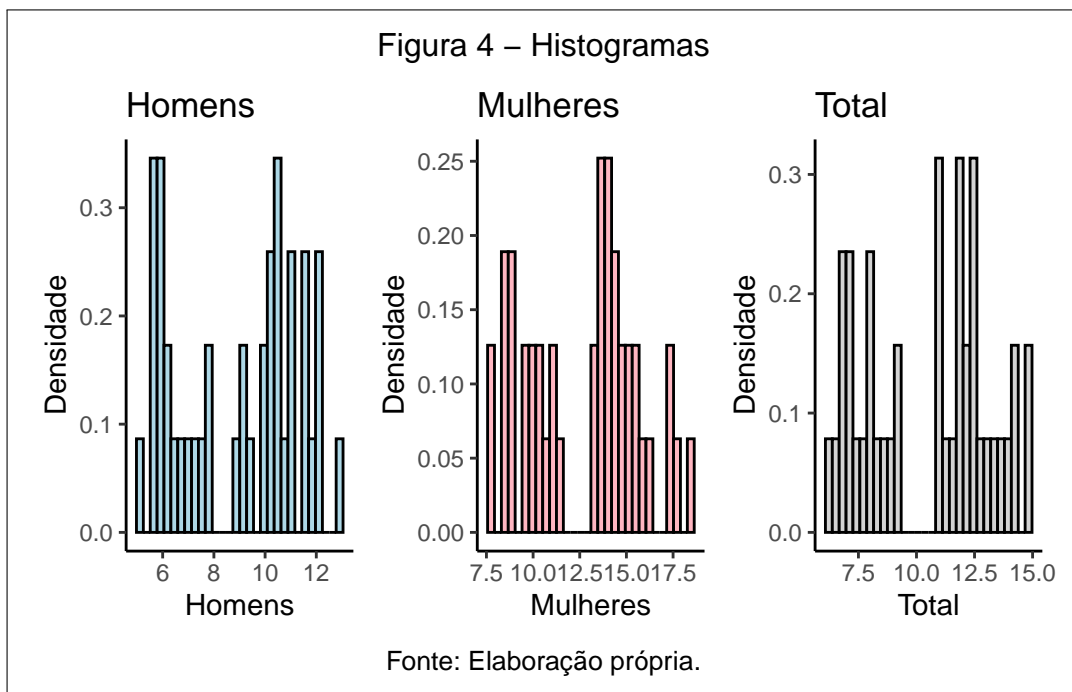
A FIGURA 2 mostra os *boxplots* por ano, em que é possível identificar o mesmo padrão visto na FIGURA 1, entretanto, o boxplot mostra uma maior variação da taxa no ano de 2021. Nota-se que não existe alinhamento entre as medianas, indicando variação na tendência central.



A FIGURA 3 mostra os *boxplots* por trimestre, em que percebe-se um certo grau de alinhamento entre as medianas, indicando uma tendência central, ou seja, não existe muita variação entre os trimestres em todos os cenários analisados.



A FIGURA 4 mostra os histogramas de densidade, não é possível identificar normalidade nos dados.



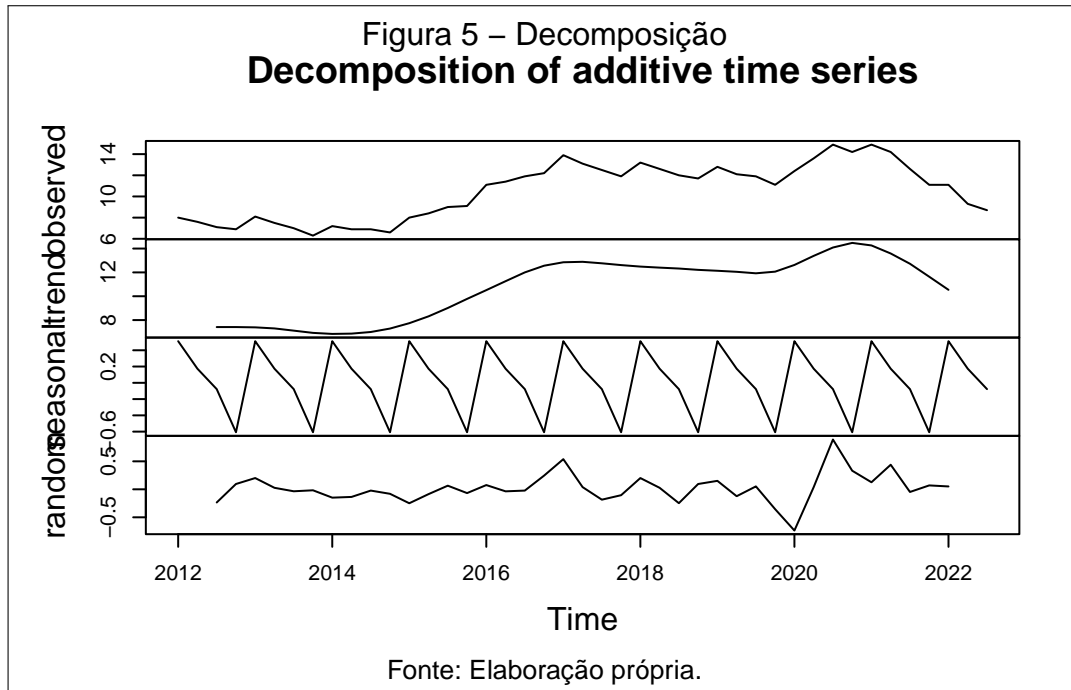
A TABELA 1 abaixo mostra as estatísticas descritivas para as variáveis analisadas, em que todas possuem assimetria negativa, o que subestima valores médios. Compreende-se também que as mulheres possuem a maior taxa de desocupação mínima, com 7,8 contra 5,1 dos homens e 8,0 do total. Os valores da média também são mais altos para mulheres (12,70) do que para os homens (8,89) e o total (10,53). O valor da curtose para todas as variáveis é maior do que 0, o que indica picos mais altos do que em uma curva normal gaussiana.

A FIGURA 5 mostra a decomposição das componentes principais da série: aleatória, sazonalidade, tendência e a série observada. É possível identificar a tendência de crescimento, vista nos gráficos de linhas (FIGURA1) e *boxplot* (FIGURA2).

Table 1: Tabela 1 - Estatísticas descritivas.

Estatística	Homens	Mulheres	Total
Min.	5.1000000	7.8000000	6.3000000
X1st.Qu.	6.3500000	10.0000000	8.0000000
Median	9.5000000	13.7000000	11.1000000
Mean	8.8930233	12.7069767	10.5348837
X3rd.Qu.	10.8000000	14.7500000	12.5500000
Max.	12.9000000	18.5000000	14.9000000
Var	5.7425692	9.0668549	7.0394684
Sd	2.3963658	3.0111219	2.6531997
Skewness	-0.1259017	-0.0440817	-0.1216666
Kurtosis	1.5580195	1.8653124	1.6447554

Fonte: Elaboração própria.



## 2.2 Modelos:

A seguir encontram-se os modelos propostos utilizando a função *tslm* e o teste de autocorrelação de resíduos de *Durbin-Watson* sob as hipóteses:

$H0$  : Não há autocorrelação dos resíduos  $H1$  : Os resíduos estão correlacionados.

Modelo 1 - Modelo com tendência.

Modelo 2 - Modelo com tendência e sazonalidade.

Call:

```
tslm(formula = Total ~ trend)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.0607	-1.0617	0.1754	1.3919	3.5187

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.15548	0.57235	12.502	1.42e-15 ***
trend	0.15361	0.02266	6.779	3.38e-08 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.844 on 41 degrees of freedom

Multiple R-squared: 0.5285, Adjusted R-squared: 0.517

F-statistic: 45.95 on 1 and 41 DF, p-value: 3.384e-08

Call:

```
tslm(formula = Total ~ trend + season)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.8103	-0.9484	0.0921	1.3909	2.9273

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.71645	0.74249	10.393	1.16e-12 ***
trend	0.15506	0.02301	6.738	5.57e-08 ***
season2	-0.51870	0.79752	-0.650	0.519
season3	-0.87376	0.79851	-1.094	0.281
season4	-1.01779	0.81719	-1.245	0.221

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.87 on 38 degrees of freedom  
Multiple R-squared: 0.5508, Adjusted R-squared: 0.5035  
F-statistic: 11.65 on 4 and 38 DF, p-value: 2.86e-06

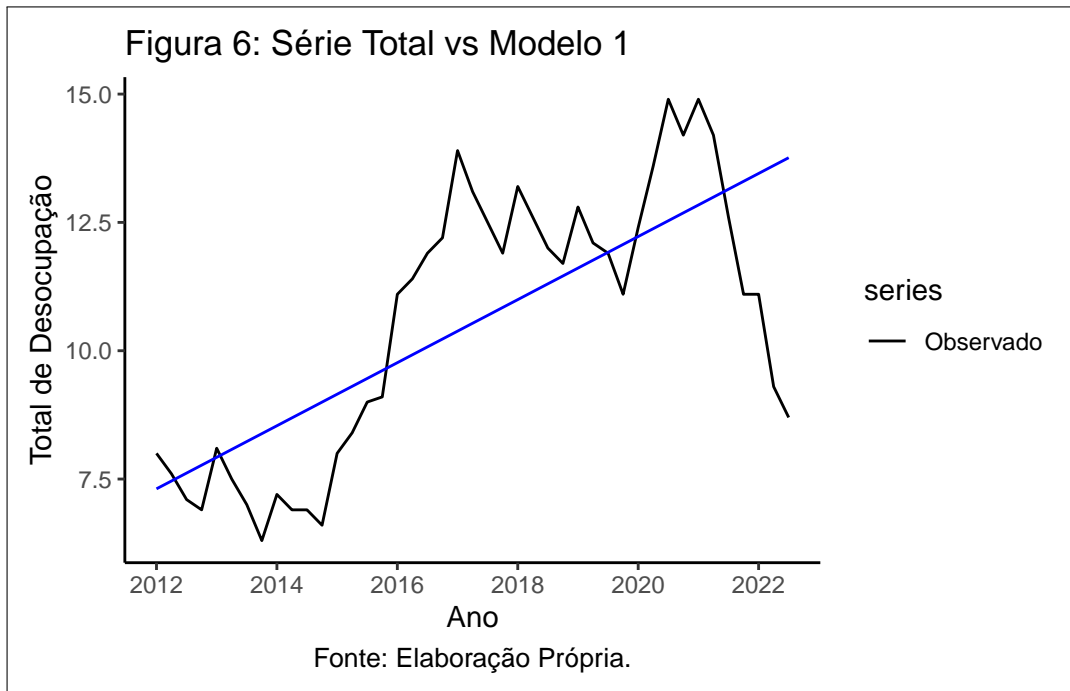
Durbin-Watson test

```
data: modelo1  
DW = 0.25469, p-value < 2.2e-16  
alternative hypothesis: true autocorrelation is greater than 0
```

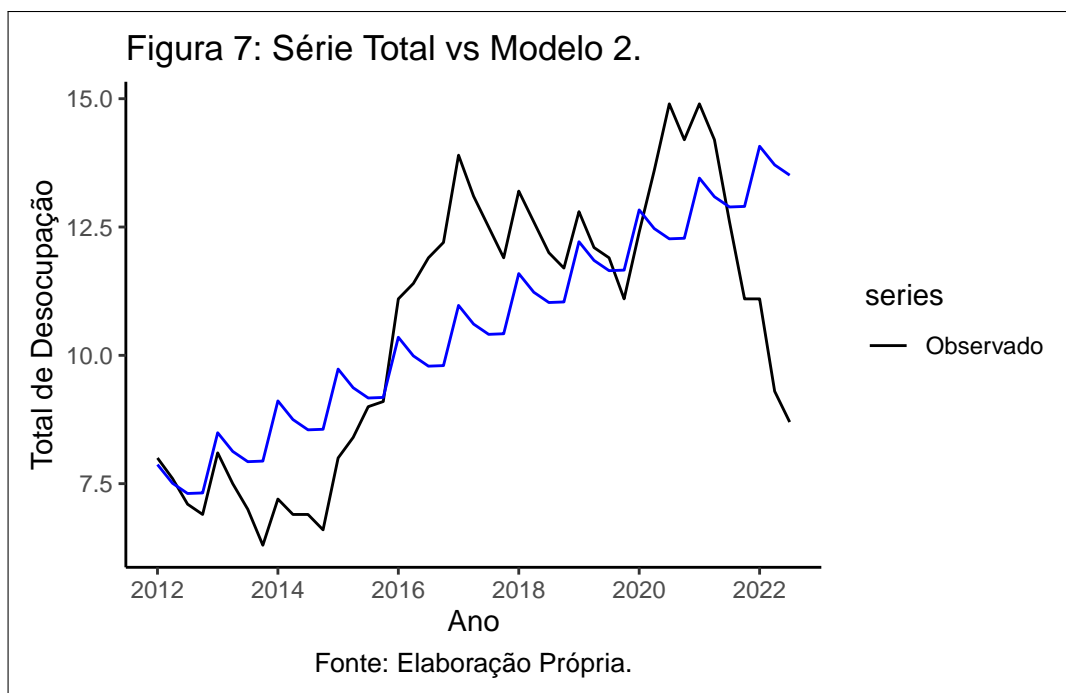
Os *p-values* observados indicam significância para a tendência e não possui significância para sazonalidade, sendo assim, o modelo mais parcimonioso é o modelo 1.

Com  $p\text{-value} = 2,2 \times 10^{-16}$ , rejeita-se a hipótese nula, indicando que os resíduos estão correlacionados.

A FIGURA 6 abaixo indica a adequação do modelo 1 a série, o resultado foi uma reta crescente que não se adequou bem a série proposta, porém conseguiu captar a tendência de crescimento.



A FIGURA 7 abaixo indica a adequação do modelo 2 a série, que captou melhor o comportamento da série do que o modelo 1, embora a componente sazonal não tenha sido significativa.

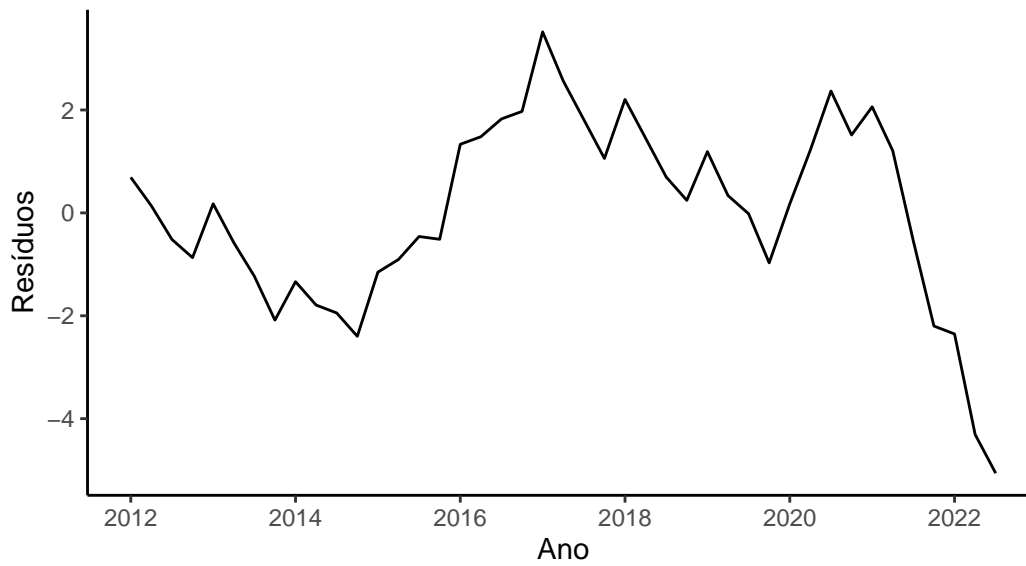


### 2.3 Análise de Resíduos:

AS FIGURAS 8-11 mostram a distribuição dos resíduos, o Gráfico de Dispersão (9) mostra que os resíduos estão distribuídos de forma aleatória, também é possível observar normalidade nos gráficos de Histograma (10) e Q-Q Plot (11).

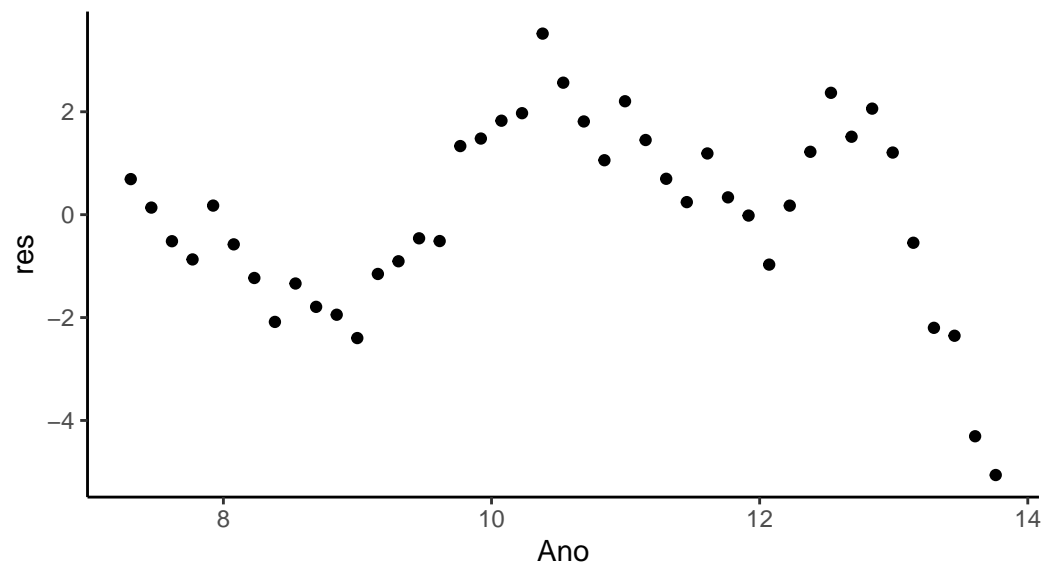


Figura 8: Gráfico de Linha dos Resíduos.

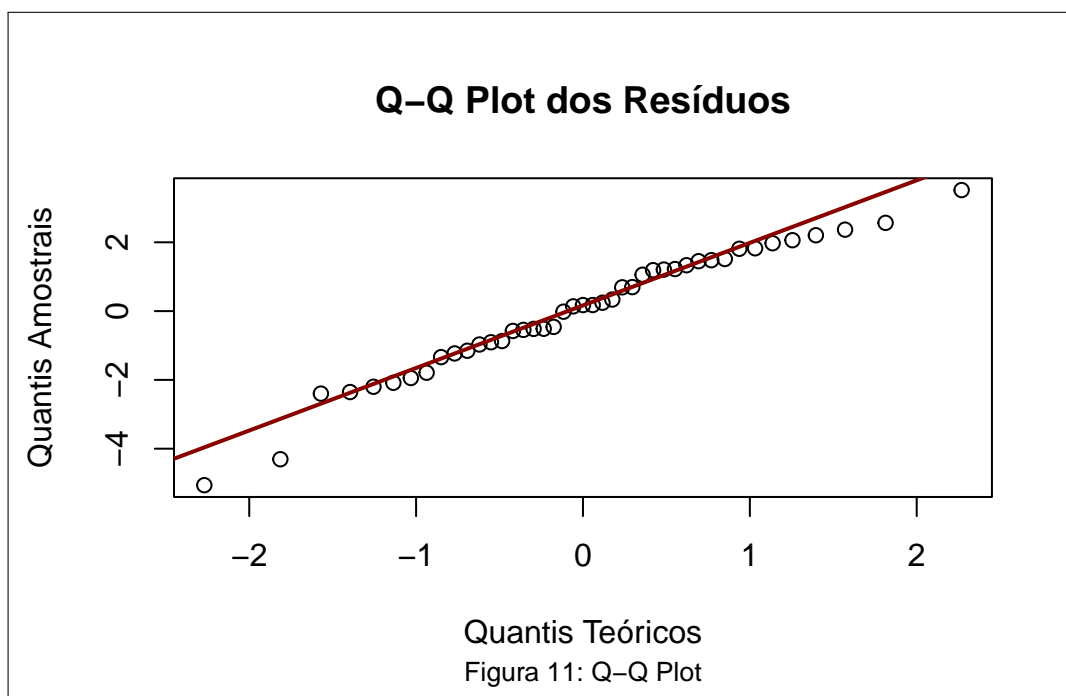
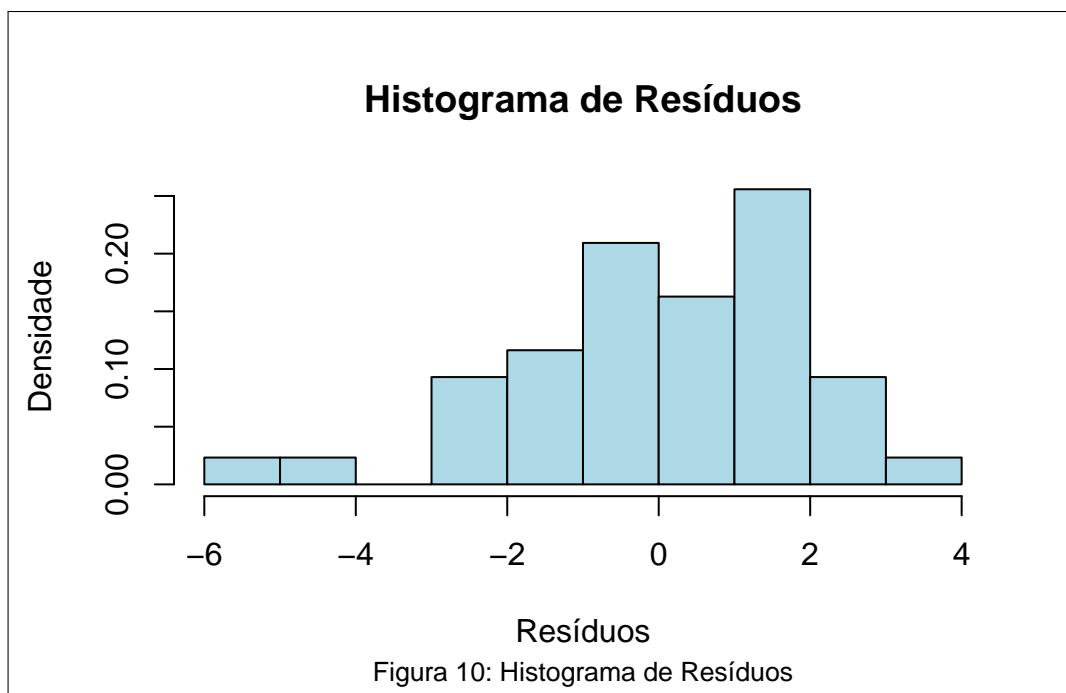


Fonte: Elaboração Própria.

Figura 9: Dispersão dos Resíduos



Fonte: Elaboração Própria.



Para conferir maior robustez a análise de resíduos, foi realizado o teste de normalidade *Anderson-Darling* e o teste *Shapiro-Wilk*, ambos sob as hipóteses:

$H_0$  : Os resíduos seguem distribuição normal  $H_1$  : Os resíduos não seguem distribuição normal.

Anderson-Darling normality test

```
data: res
A = 0.36743, p-value = 0.4159
```

Shapiro-Wilk normality test

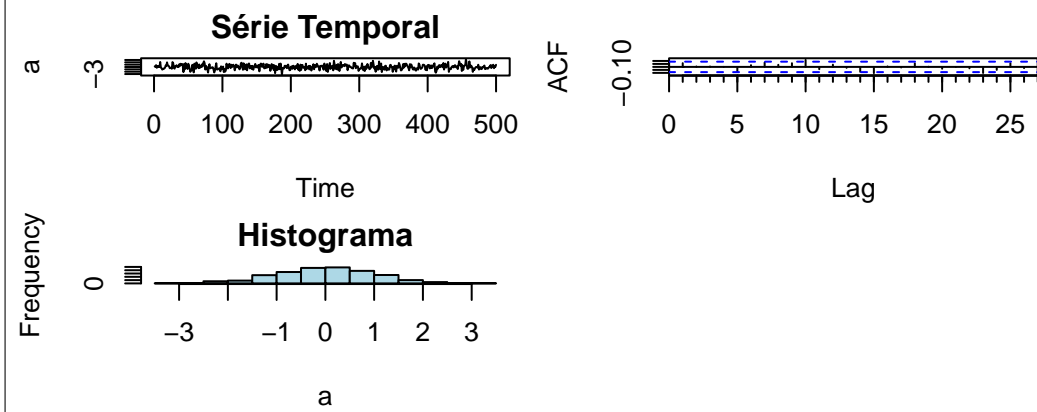
```
data: res
W = 0.96909, p-value = 0.2941
```

Ambos os testes apresentaram *p-values* não significativos, ou seja, não há evidências estatísticas para rejeição da hipótese nula, com nível de confiança de 95%.

### 3 SIMULAÇÕES:

Para a simulação P1:  $\{at\}$  com médio zero e variância 1, o resultado pode ser visto na FIGURA 12 abaixo. É possível identificar uma série estacionária em torno da média zero, enquanto o gráfico ACF não indica autocorrelação significativa, ou seja, é um ruído branco. O histograma apresenta o comportamento de uma curva normal, já esperado pois os valores da simulação são de uma distribuição normal padrão.

Figura 12 – EDA Processo 1

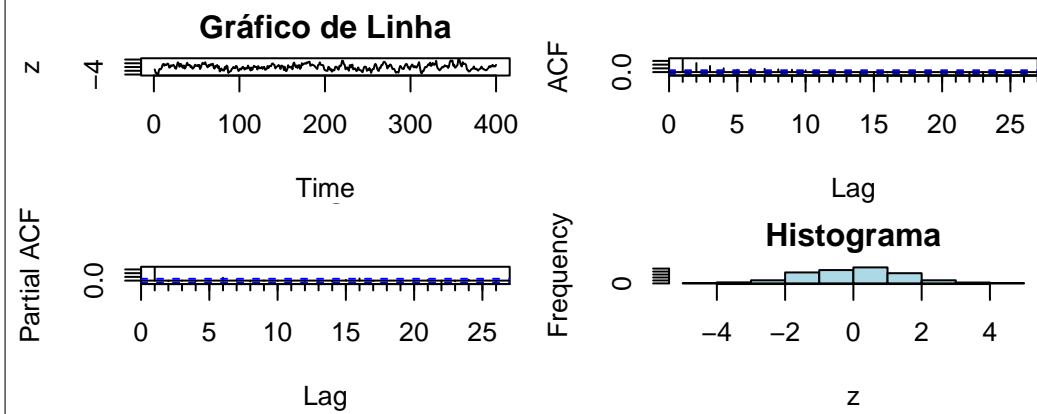


Fonte: Elaboração própria.

A simulação do Processo 2 é possível identificar:

- ACF decai exponencialmente (típico de AR(1))
- PACF tem apenas 1 lag significativo (confirma AR(1))
- Processo estacionário pois  $|\phi| = 0.7 < 1$

Figura 13 – EDA Processo 2



Fonte: Elaboração própria.

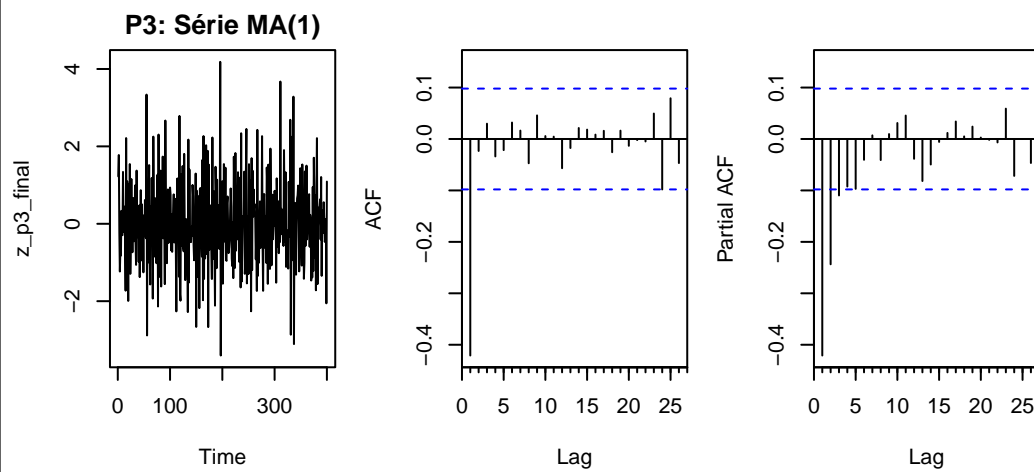
Abaixo os valores das estatísticas descritivas:

```
[,1]
Min.    -4.199660219
1st Qu. -1.023096752
Median   0.057515880
Mean    -0.007918512
3rd Qu.  0.977560741
Max.     4.333991673
Var      2.025816066
Sd       1.423311654
skewness 0.036848565
kurtose  2.993744029
```

Simulação do Processo 3:

- ACF tem 1 lag significativo (típico de MA(1)).
- PACF decai exponencialmente.
- Processo sempre estacionário (qualquer theta em MA(1)).

Figura 14 – EDA Processo 3



Fonte: Elaboração própria.