

Atividade 2 - Séries Temporais 2

Caroline Cogo*

João Inácio Scrimini†

Joelmir Junior‡

Renata F. Stone§

julho 2022

Sumário

1	Exercício 1: Números Mensais de Manchas Solares, 1749-1983.	2
1.1	Análise inicial	2
1.2	Aplicação dos Testes	4
1.3	Modelagem	5
1.4	Cálculo das medidas de acurácia	6
1.5	Análise de resíduo do modelo escolhido	7
1.6	Predição/Previsão do modelo escolhido	7
2	Exercício 2: Consumo Trimestral de Gás do Reino Unido.	9
2.1	Análise inicial	9
2.2	Aplicação dos Testes	11
2.3	Modelagem	12
2.4	Cálculo das medidas de acurácia	13
2.5	Análise de resíduo do modelo escolhido	13
2.6	Predição/Previsão do modelo escolhido	14

*carolcogo808@gmail.com

†joao.inacio.scrimini@gmail.com

‡moura22jr@hotmail.com

§renastan@gmail.com

1 Exercício 1: Números Mensais de Manchas Solares, 1749-1983.

1.1 Análise inicial

Nessa análise serão considerados dados mensais sobre Manchas Solares, o período da série temporal é entre os anos de 1749 até 1983. Na Figura 1 é apresentada a série com 2820 observações.

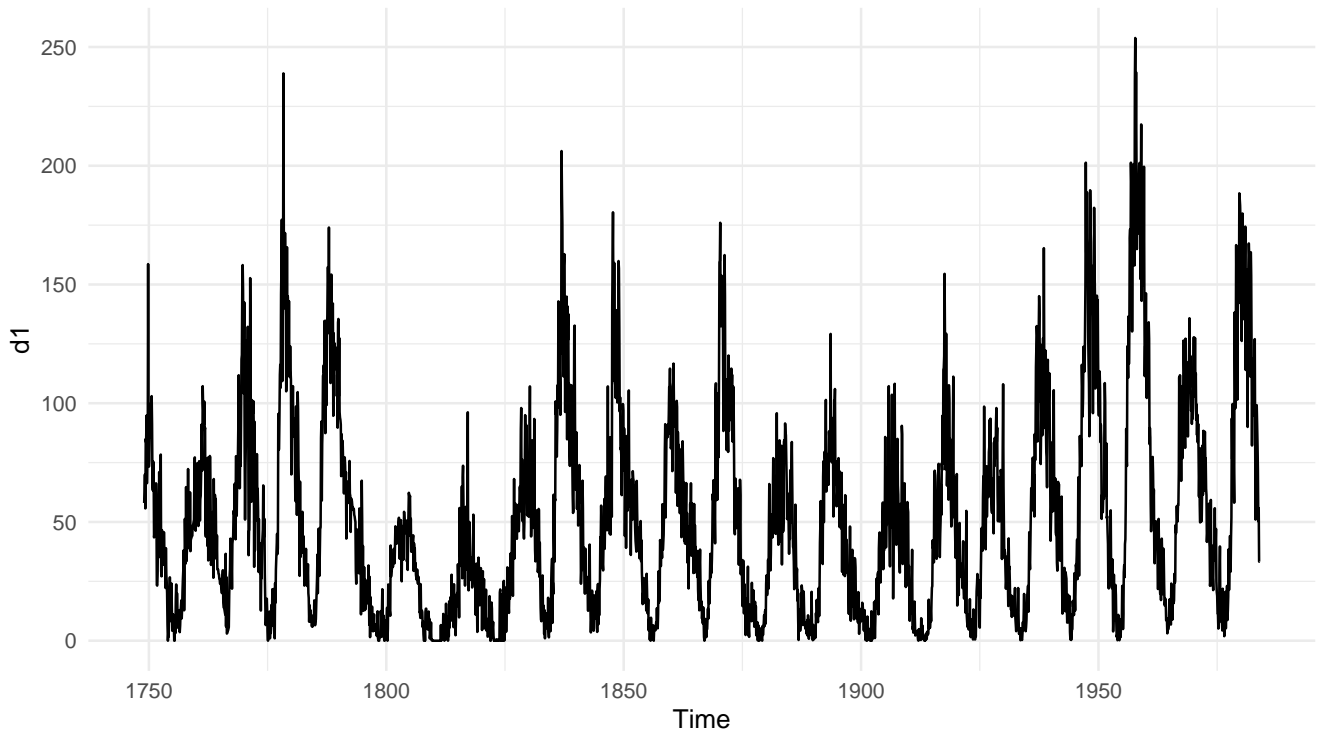


Figura 1: Gráfico da série amostral

Agora, considerando a Figura 2 (FAC), pode-se ver que existe autocorrelação, contendo um grande pico e significância no lag 1, diminuindo nos seguintes, e voltando a ser significativo em somente dois pontos, aproximadamente, perto de 36 e perto de 89. O decaimento da função de autocorrelação ocorre de forma lenta, indicando a dependência entre as observações. Agora, considerando a Figura 3 (FACP) nota-se que o primeiro lag é significativo e ao decorrer dos lags, até aproximadamente o lag 24 temos algumas correlações significativas. Mas na sequência, as correlações tornam-se não significativas, estando dentro dos limites estabelecidos.

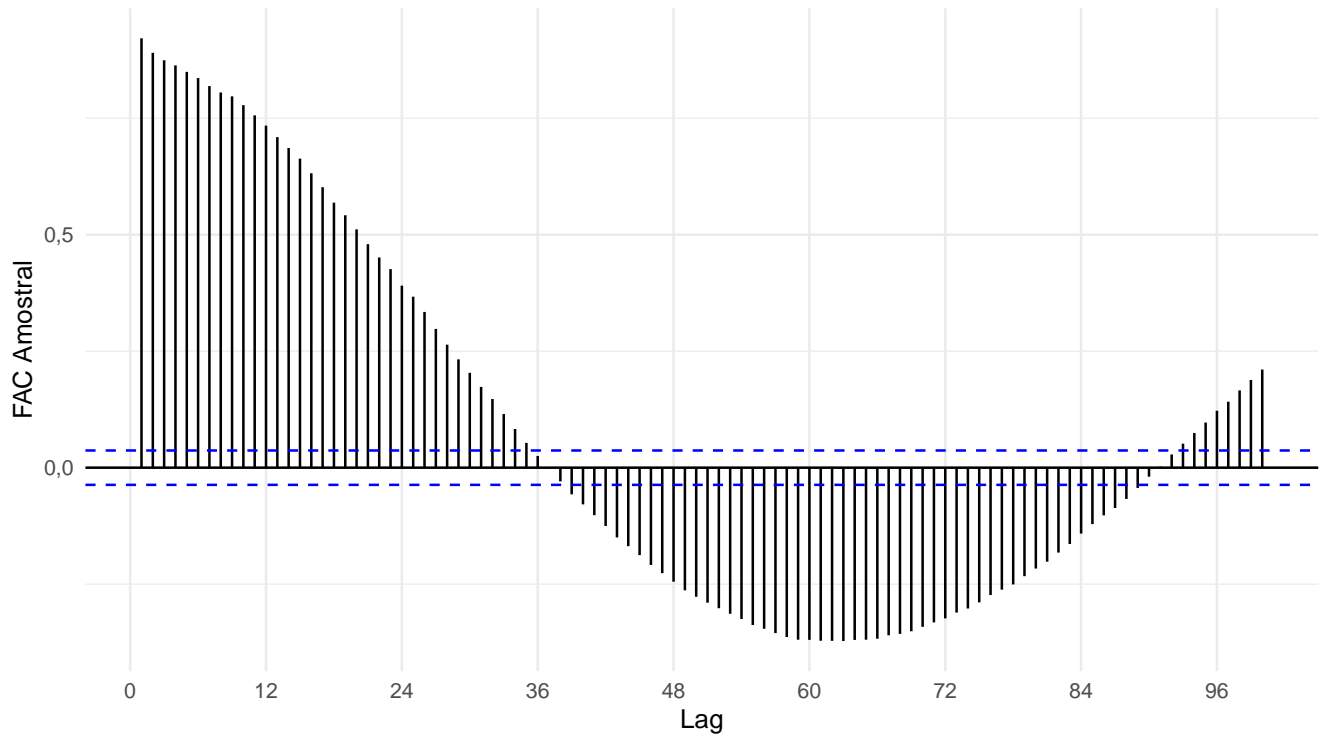


Figura 2: Gráfico da Função de autocorrelação amostral

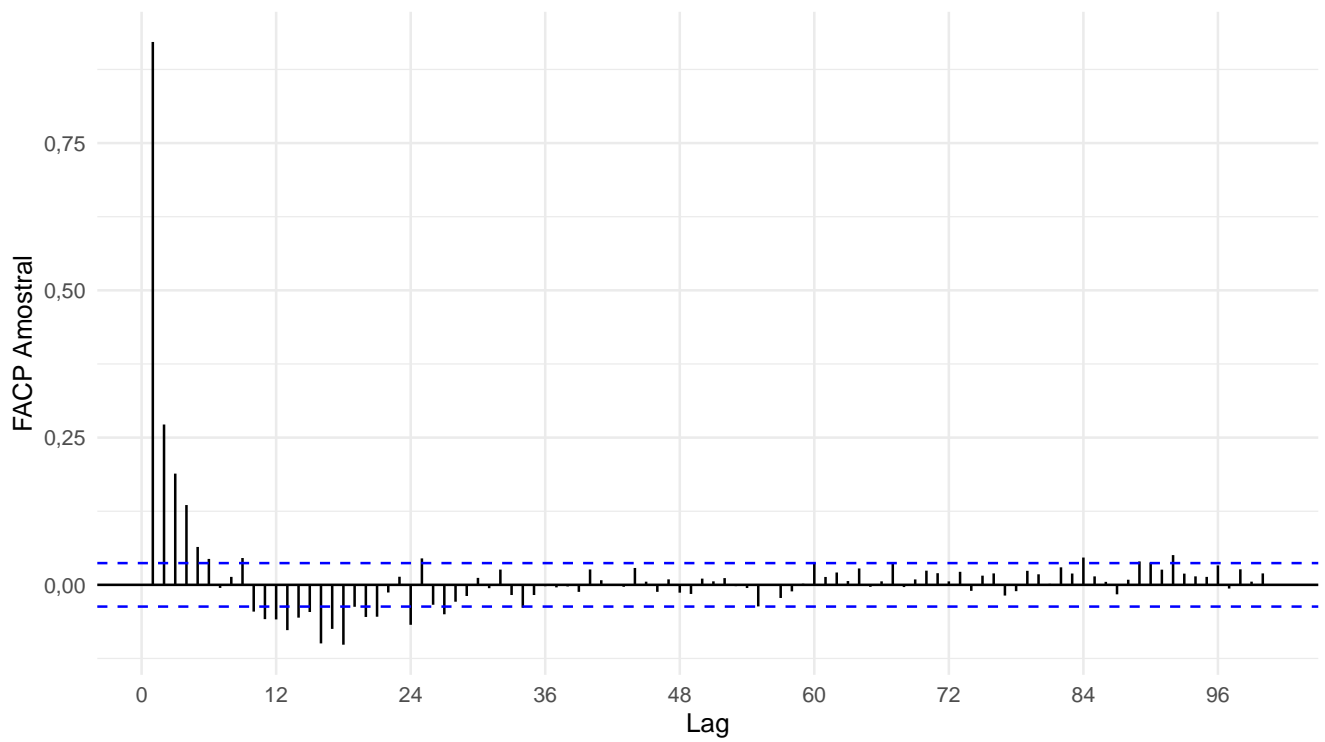


Figura 3: Gráfico da Função de autocorrelação parcial

Na sequência é feita a análise para avaliar se a série apresenta outliers, os quais podem influenciar na modelagem.

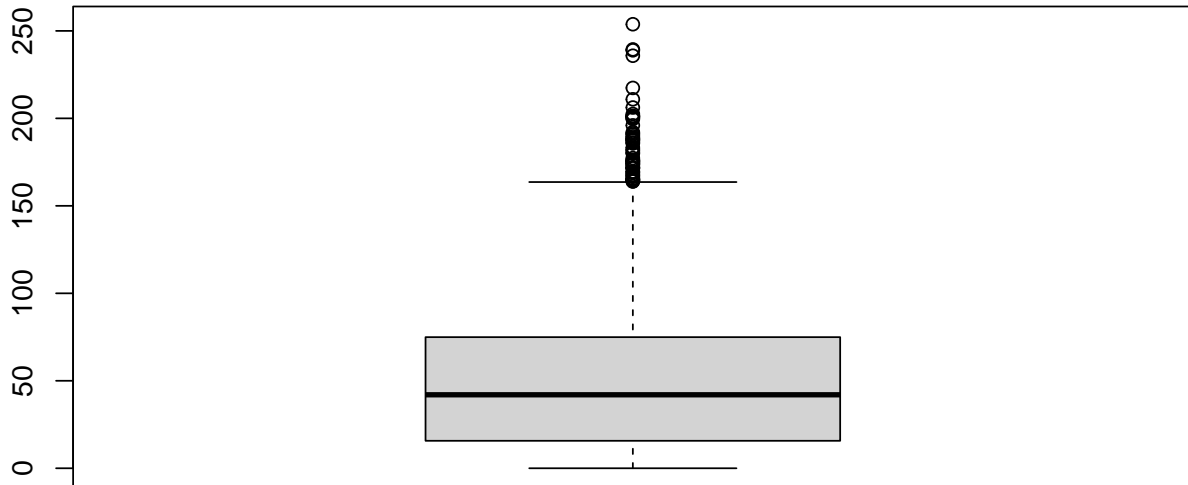


Figura 4: Gráfico Boxplot da série amostral

Na Figura 4, no Boxplot é indicado a existência de alguns pontos acima do limite, o que torna-os possíveis outliers. Para a confirmação foram testados cinco observações (2506, 2508, 353, 2505 e 2521), entre elas as observações 2506, 2508, 353 e 2505 foram consideradas outliers pelo Teste de Rosner e assim retiradas da série temporal, resultando em 2816 observações no total.

Após a identificação e retirada de outliers da série, será iniciados os testes para avaliar normalidade, tendência determinística, raiz unitária e sazonalidade.

1.2 Aplicação dos Testes

1.2.1 Testes de normalidade

Teste de Jarque-Bera, H_0 : Os dados possuem distribuição normal.

Com p valor igual a $< 2 \times e^{-16}$, ao nível de significância igual a $\alpha = 0.05$, conclui-se que rejeitamos **H_0** .

Teste de Shapiro Wilk, H_0 : Os dados possuem distribuição normal.

Com p valor igual a $4,5467 \times 10^{-38}$, ao nível de significância igual a $\alpha = 0.05$, conclui-se que rejeitamos **H_0** .

Como a série não apresenta normalidade, aplicou-se inicialmente a transformação de BoxCox nos dados, entretanto, após a transformação foi realizados os testes novamente e a série continuou não apresentando normalidade, por isso, continuamos as análises com os dados originais.

1.2.2 Teste de tendência determinística

Pela Tabela 1 abaixo, na maioria dos testes de tendência determinística aplicados, os p-valores são inferiores ao alpha ($\alpha = 5\%$). Portanto, rejeita-se a hipótese nula (H_0), **considerando como H_0 : Sem Tendência Determinística (STD)** e **H_1 : Possui Tendência Determinística (TD)**. Apenas a rotina *cox.stuart* não rejeitou H_0 .

Sendo assim, conclui-se pelos testes realizados que a série apresenta tendência determinística.

Tabela 1: Testes de Tendência Determinística

Teste	Rotina	H_0	H_1	P-valor	Conclusão
Cox-Stuart	<i>cox.stuart</i>	STD	TD	0,2862	STD
Cox-Stuart	<i>cs.test</i>	STD	TD	$1,0478 \times 10^{-8}$	TD
Wald-Wolfowitz	<i>runs.test</i>	STD	TD	$< 2 \times e^{-16}$	TD
Wald-Wolfowitz	<i>ww.test</i>	STD	TD	$< 2 \times e^{-16}$	TD
Mann-Kendall	<i>mk.test</i>	STD	TD	$1,4584 \times 10^{-9}$	TD
Mann-Kendall	<i>MannKendall</i>	STD	TD	$< 2.22 \times e^{-16}$	TD

1.2.3 Teste de raiz unitária

A partir da Tabela 2, considere RU como sendo a hipótese de existir raiz unitária e Estacionária como a hipótese de ter estacionariedade. Para o teste Aumentado de Dickey-Fuller (ADF) e teste de Phillips-Perron (PP), observa-se p-valor menor que 0,05, logo, rejeitamos a hipótese nula (H_0), demonstrando não haver Raiz unitária. Agora, segundo os testes de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), sendo o primeiro referente a tendência estocástica, a série apresentou raiz unitária, com p-valor sendo menor que 0,01, rejeitando a hipótese nula (H_0). Já no segundo, referente a tendência determinística, temos que a série apresenta tendência determinística, com p-valor menor que 0,01, rejeitando H_0 .

Tabela 2: Testes de Tendência Estocástica - Raiz Unitária

Teste	Rotina	H_0	H_1	P-valor	Conclusão
ADF	<i>adf.test</i>	RU	Estacionária	0,01	Estacionária
PP	<i>pp.test</i>	RU	Estacionária	0,01	Estacionária
KPSS	<i>kpss.test</i>	Estacionária	RU	0,01	RU
KPSS	<i>kpss.test</i>	Estacionária	TD	0,01	TD

Com os resultados dos testes das Tabelas 1 e 2, podemos verificar que a série apresenta as duas tendências, determinística e estocástica.

Na sequência é realizado o método de diferenciação na série em estudo, e após são aplicados novamente os mesmos testes das Tabelas 1 e 2, e conclui-se que a série diferenciada não possui tendência determinística e estocástica, e é estacionária.

1.2.4 Testes de Sazonalidade

Considere que para uma série temporal ser sazonal (possuir sazonalidade) é preciso que os fenômenos que ocorrem durante o tempo se repitam em um período idêntico de tempo. Logo, testes pra identificar sazonalidade na série ajustada foram feitos.

Com a série já diferenciada, testamos a presença de sazonalidade nos dados, através do Teste de sazonalidade de Kruskal-Wallis e o de Friedman, em que para ambos os testes, consideramos H_0 como a série não sendo sazonal. Na Tabela 3 estão os p-valor obtidos. Para os dois testes o p-valor foi menor que 0,05, rejeitando a hipótese nula (H_0) e indicando que a série possui sazonalidade.

Tabela 3: Testes de Sazonalidade

Teste	Rotina	H_0	P-valor	Conclusão
Kruskall Wallis	<i>kw</i>	Não Sazonal	0,0486	Sazonal
Friedman	<i>fried</i>	Não Sazonal	0,0261	Sazonal

1.3 Modelagem

Inicialmente, para a modelagem foi feita a divisão da série temporal entre treino e teste. Na série de treino ficou definida 2804 observações, enquanto que para a série de teste ficou 12 observações. Com o objetivo de aplicar sobre o

modelo de treino e depois avaliar a acurácia, no modelo de teste. Na Figura 5, através do gráfico, é possível perceber a divisão feita na série temporal.

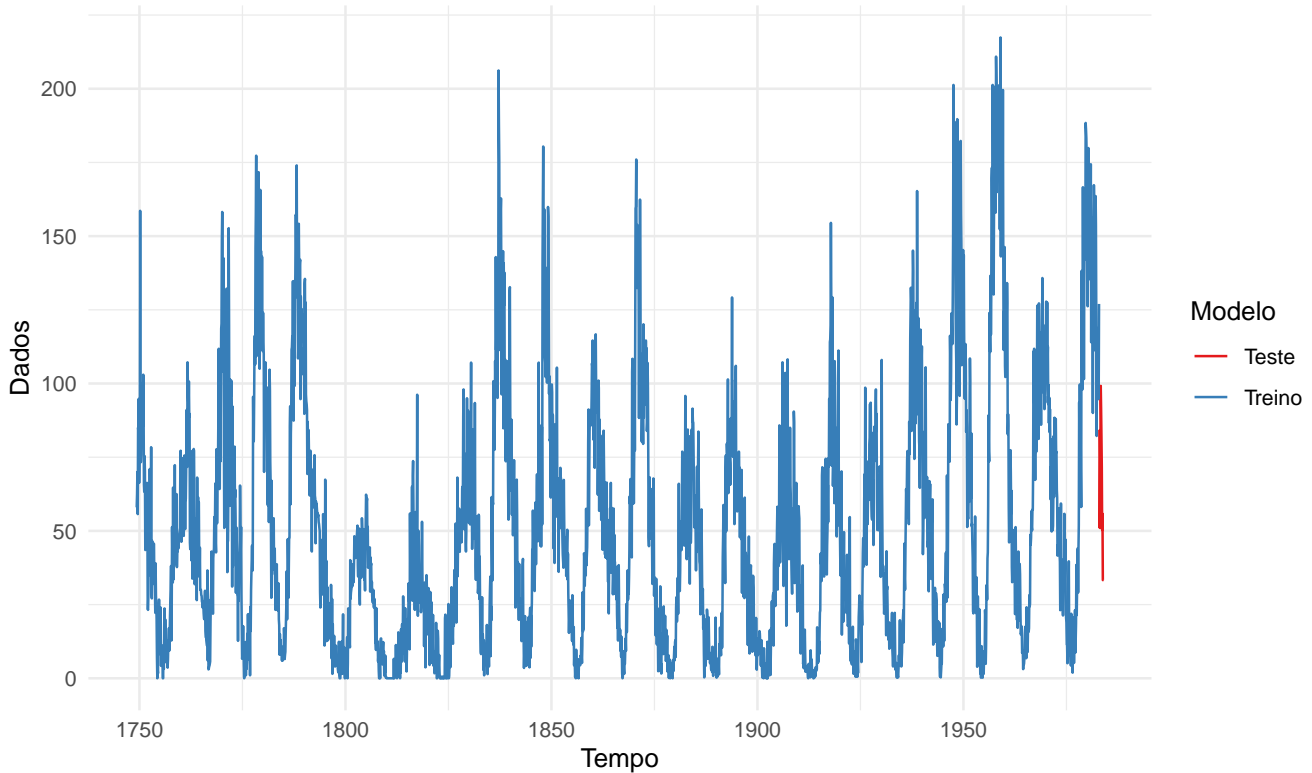


Figura 5: Gráfico da série dividida em treino e teste

Foram considerados dois modelos SARIMA:

- Modelo 1 = $(ar1, ar2, ma1, ma2, sar1, sma1, sma2) = (2,1,2)(1,0,2)[12]$
- Modelo 2 = $(ar1, ar2, ma1, ma2, sma1, sma2) = (2,1,2)(0,0,2)[12]$

Para cada modelo foi feito o cálculo das medidas de acurácia e a análise de resíduos, a fim de escolher o melhor modelo possível.

1.4 Cálculo das medidas de acurácia

Pode-se perceber que considerando o critério de informação de Akaike (AIC), ambos modelo estão próximos, afinal eles apresentaram uma pequena diferença, o modelo 1 resultou em 23290 enquanto que o do modelo 2 foi de 23291.

Agora, na Tabela 4, é apresentado as medidas de acurácia de cada modelo, realizadas sobre a série de treino, em um periodo de 12 meses, nota-se que o modelo 2 apresentou menores erros de previsão, apenas no teste ACF1 que o modelo 1 apresentou-se menor. Com essas análises, podemos perceber que o modelo 2 teve melhor ajuste e menores erros de previsão. Portanto utilizaremos o modelo 2 para realizar as previsões.

Tabela 4: Medidas de acurácia

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Modelo 1	-37,25	41,91	37,25	-74,65	74,65	0,5698	3,482
Modelo 2	-36,83	41,58	36,83	-74,00	74,00	0,5749	3,465

1.5 Análise de resíduo do modelo escolhido

Na Figura 6 é apresentado o gráfico com os círculos unitários, em que são calculadas as inversas das raízes de cada polinômio autoregressivo e de médias móveis. Nesse caso todas as raízes estão dentro do círculo unitário, logo eles são inversíveis e estacionários.

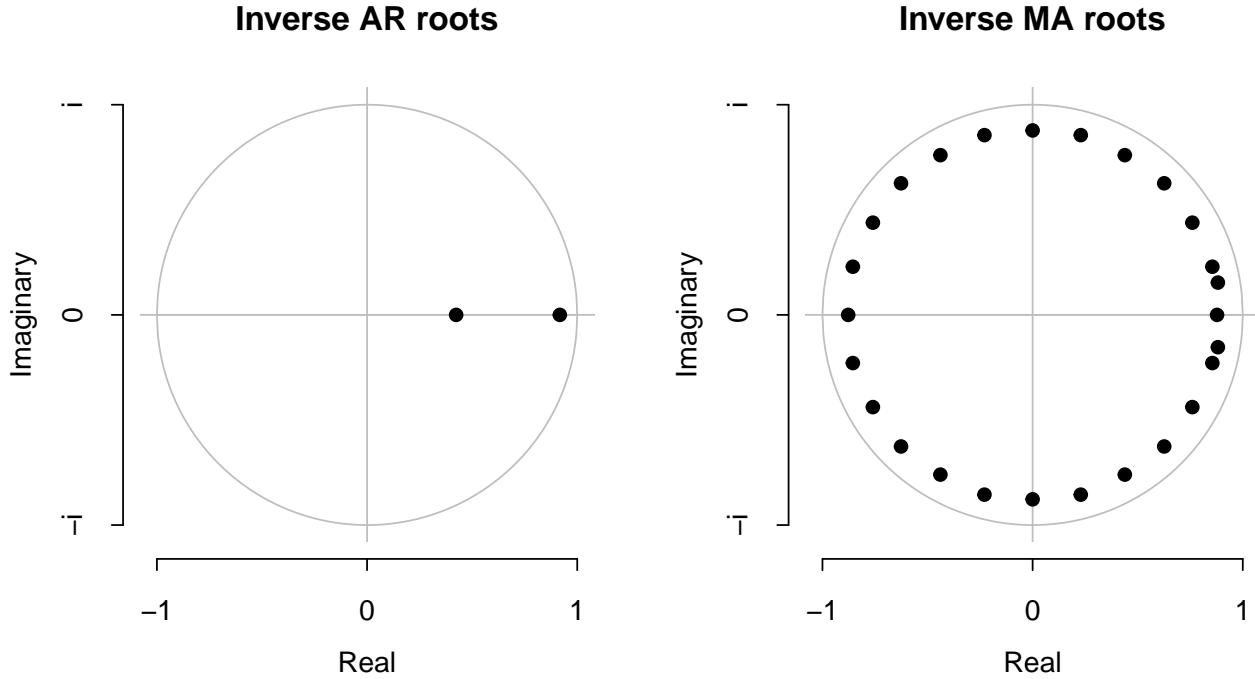


Figura 6: Representação do círculo unitário para o modelo 2

Pelo Box-Ljung test os resíduos não apresentaram correlação, com $p\text{-valor} = 0,1$, e utilizando $\text{lag} = 15$. Pelo Dickey-Fuller Test os resíduos apresentaram estacionariedade, com $p\text{-valor} = 0,01$. Os resíduos não apresentaram normalidade ($\text{shapiro.test}, p\text{-valor} < 2 \times e^{-16}$).

1.6 Predição/Previsão do modelo escolhido

Tabela 5: Previsão de 12 meses para o modelo 2

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2817	41,34	21,600	61,08	11,1514	71,52
2818	39,03	16,229	61,82	4,1617	73,89
2819	38,48	14,196	62,76	1,3416	75,62
2820	37,99	12,569	63,41	-0,8889	76,87
2821	37,72	11,194	64,25	-2,8500	78,30
2822	33,59	5,883	61,30	-8,7849	75,96
2823	31,38	2,391	60,36	-12,9528	75,70
2824	28,86	-1,502	59,23	-17,5764	75,30
2825	25,72	-6,121	57,55	-22,9746	74,41
2826	25,10	-8,287	58,49	-25,9621	76,16
2827	22,65	-12,355	57,65	-30,8860	76,18
2828	19,60	-17,076	56,27	-36,4896	75,68

Na Tabela 5 é apresentado as previsões para os próximos 12 meses utilizando o modelo 2, nota-se que a cada mês espera-se uma alteração entre 19,60 a 41,34, diminuído com o passar dos meses, com margem de erro médio de 19,74 até 36,67 para nível de confiança de 80% e margem de erro médio de 30,18 até 56,08 para nível de confiança de 95% e pela Figura 7 vemos esse ajuste da predição do modelo SARIMA(2,1,2)(0,0,2)[12] e da previsão de 12 meses para a série original, com os limites 95%.

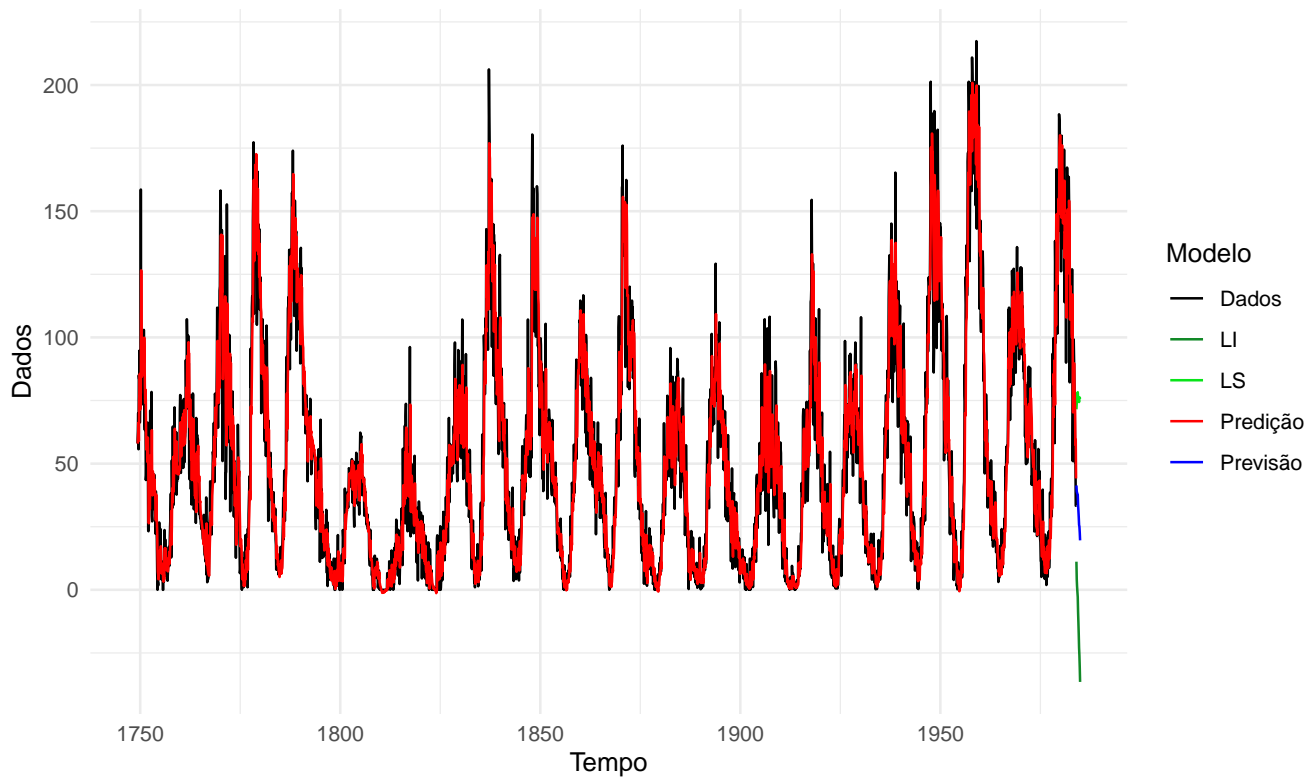


Figura 7: Gráfico da série contendo a predição do modelo final SARIMA(2,1,2)(0,0,2)[12] e predição de 12 meses, com limites superior e inferior de 95%.

2 Exercício 2: Consumo Trimestral de Gás do Reino Unido.

2.1 Análise inicial

Nessa análise serão considerados dados trimestrais sobre o consumo de gás do Reino Unido, o período da série temporal é entre os anos de 1960 até 1987. Na Figura 8 é apresentada a série com 108 observações.

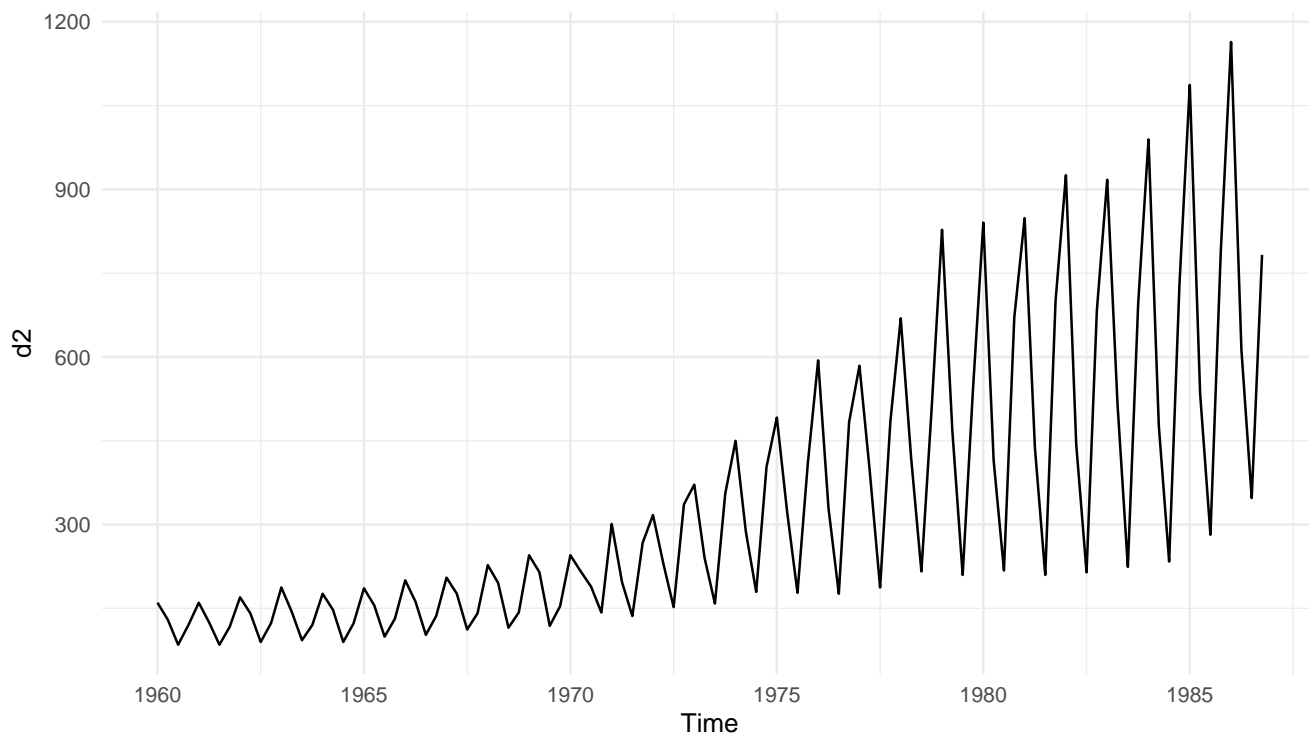


Figura 8: Gráfico da série amostral

Agora, considerando a Figura 9 (FAC), pode-se ver que existe autocorrelação, entretanto não existe autocorrelação constante, diminuindo durante os primeiros lags, e voltando a ser significativo em partes, aproximadamente, entre os lags 35 até 45. O decaimento da função de autocorrelação ocorre indicando um possível padrão. A partir do lag 87, aproximadamente, a série é significativa, sem autocorrelação. Agora, considerando a Figura 10 (FACP) nota-se que os primeiros lags são significativos, e partir do lag 6 temos correlações não significativas, dentro dos limites estabelecidos.

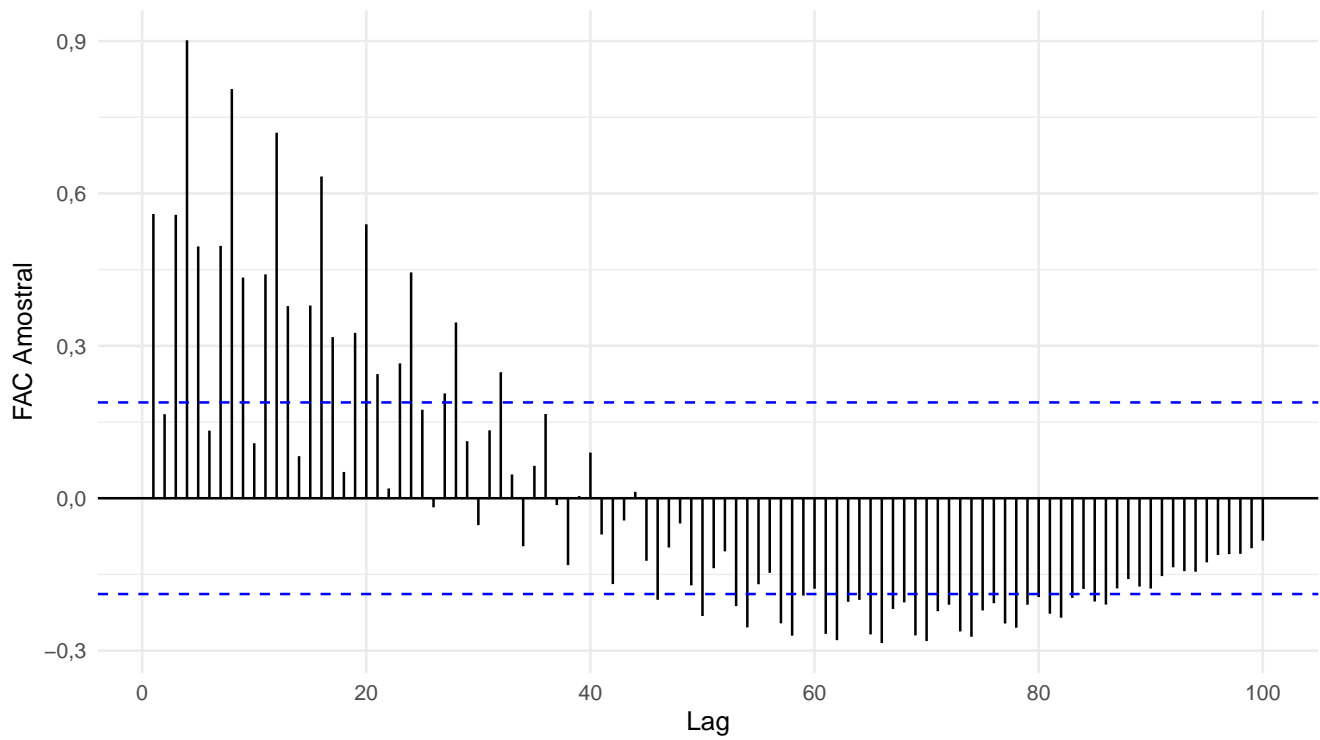


Figura 9: Gráfico da Função de autocorrelação amostral

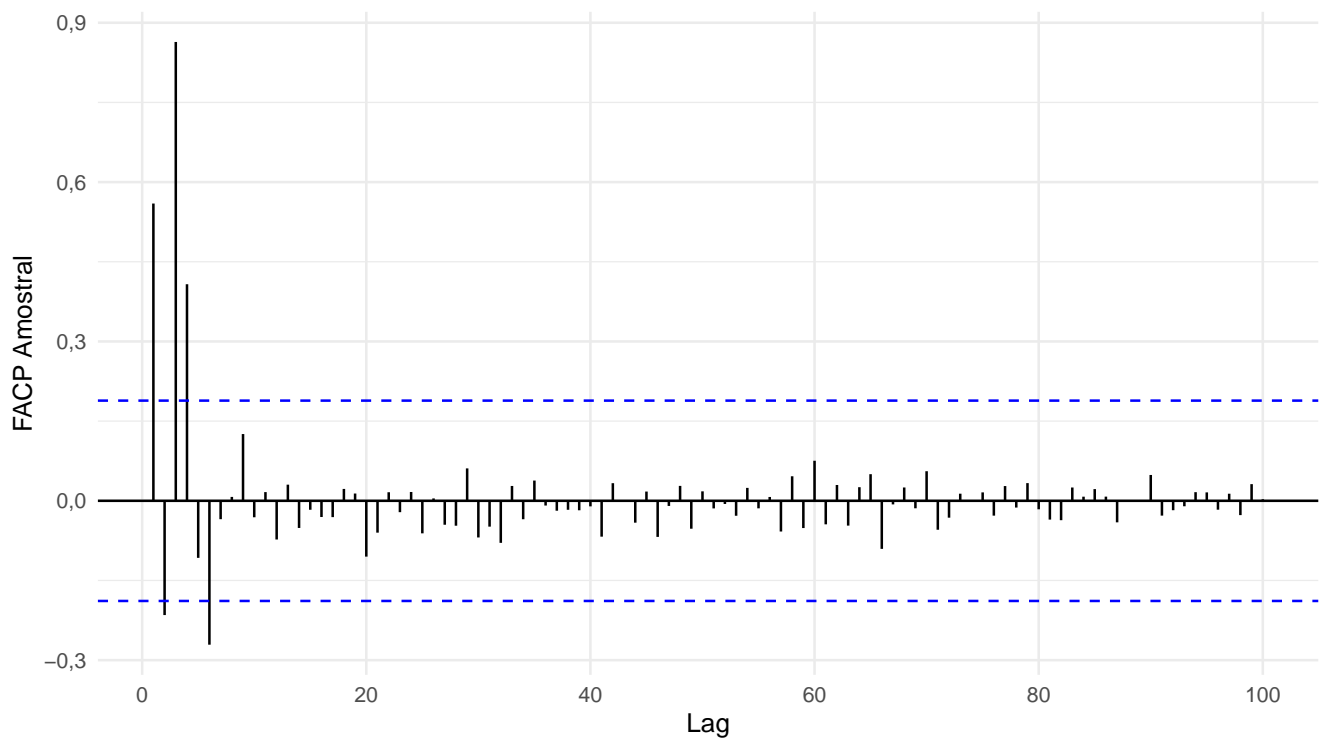


Figura 10: Gráfico da Função de autocorrelação parcial

Na sequência é feita a análise para avaliar se a série apresenta outliers, os quais podem influenciar na modelagem.

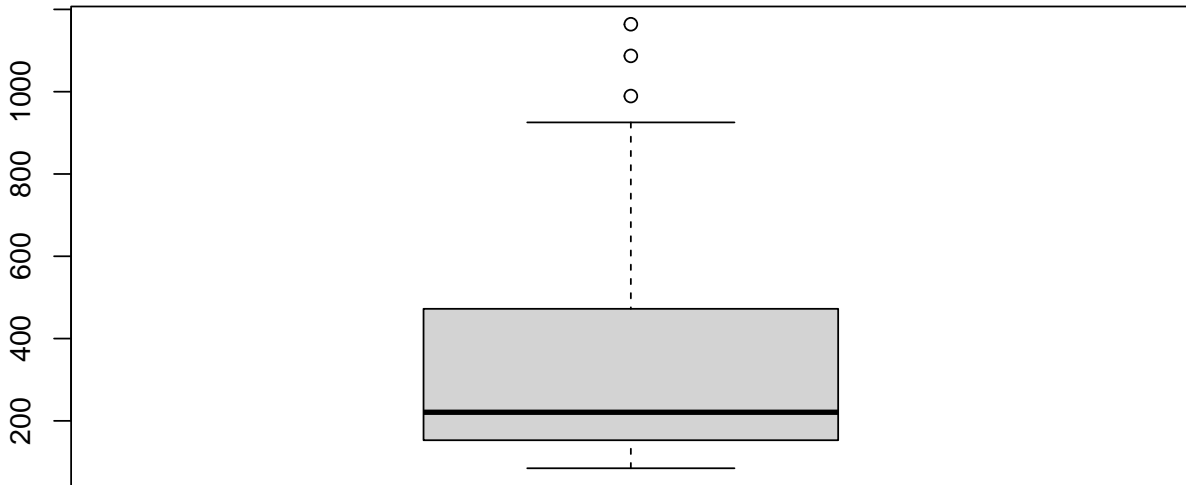


Figura 11: Gráfico Boxplot da série amostral

Na Figura 11, no Boxplot é indicado a existência de alguns pontos acima do limite, o que torna-os possíveis outliers. Para a confirmação foram testados três observações (105, 101, 97), mas nenhuma foi considerada outlier e assim não foram retiradas da série temporal. A partir de então será iniciados os testes para avaliar normalidade, tendência determinística, raiz unitária e sazonalidade.

2.2 Aplicação dos Testes

2.2.1 Teste de normalidade

Teste de Jarque-Bera, H_0 : Os dados possuem distribuição normal.

Com p valor igual a $< 3.619 \times 10^{-8}$, ao nível de significância igual a $\alpha = 0.05$, conclui-se que rejeitamos **H_0** .

Teste de Shapiro Wilk, H_0 : Os dados possuem distribuição normal.

Com p valor igual a $1,4344 \times 10^{-9}$, ao nível de significância igual a $\alpha = 0.05$, conclui-se que rejeitamos **H_0** .

Como a série não apresenta normalidade, aplicou-se inicialmente a transformação de BoxCox nos dados, entretanto, após a transformação foi realizados os testes novamente e a série continuou não apresentando normalidade, por isso, continuamos as análises com os dados originais.

2.2.2 Teste de tendência determinística

Ao observarmos a Tabela 6, concluímos que a série apresenta tendência determinística a partir dos testes realizados, em que os p-valores são todos menores que o alpha ($\alpha = 5\%$). Sendo assim, rejeitamos H_0 , considerando H_0 : **Sem Tendência Determinística (STD)** e H_1 : **Possui Tendência Determinística (TD)**

Tabela 6: Testes de Tendência Determinística

Teste	Rotina	H_0	H_1	P-valor	Conclusão
Cox-Stuart	<i>cox.stuart</i>	STD	TD	$1,3843 \times 10^{-7}$	TD
Cox-Stuart	<i>cs.test</i>	STD	TD	$1,9732 \times 10^{-9}$	TD
Wald-Wolfowitz	<i>runs.test</i>	STD	TD	$9 \times e^{-06}$	TD
Wald-Wolfowitz	<i>ww.test</i>	STD	TD	$4 \times e^{-09}$	TD
Mann-Kendall	<i>mk.test</i>	STD	TD	$2,7529 \times 10^{-22}$	TD
Mann-Kendall	<i>MannKendall</i>	STD	TD	$< 2.22 \times e^{-16}$	TD

2.2.3 Teste de raiz unitária

Na Tabela 7, consideramos “RU” como a hipótese para raiz unitária e “Estacionária” para hipótese de estacionariedade. Pelo teste Aumentado de Dickey-Fuller (ADf) e teste de Phillips-Perron (PP), com $\alpha = 0,05$, não rejeitamos a hipótese nula (H_0), o que nos indica Raiz Unitária. Agora, pelos testes de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sendo o primeiro referente a tendência estocástica, indica raiz unitária, com p-valor sendo menor que 0,01, rejeitando, portanto, a hipótese nula (H_0). Enquanto o segundo teste nos indica que a série apresenta tendência determinística, com p-valor menor que 0,01, rejeitando H_0 .

Tabela 7: Testes de Tendência Estocástica - Raiz Unitária

Teste	Rotina	H_0	H_1	P-valor	Conclusão
ADF	<i>adf.test</i>	RU	Estacionária	0,7393	RU
PP	<i>pp.test</i>	RU	Estacionária	0,01	Estacionária
KPSS	<i>kpss.test</i>	Estacionária	RU	0,01	RU
KPSS	<i>kpss.test</i>	Estacionária	TD	0,01	TD

2.2.4 Testes de Sazonalidade

Logo abaixo, realizaremos testes para identificar sazonalidade na série ajustada.

Com a série já diferenciada, testamos a presença de sazonalidade nos dados, através do Teste de sazonalidade de Kruskal-Wallis e o de Friedman, em que para ambos os testes, consideramos H_0 como a série não sendo sazonal. Na Tabela 8 estão os p-valor obtidos. Para os dois testes o p-valor foi menor que 0,05, rejeitando a hipótese nula (H_0) e indicando que a série possui sazonalidade.

Tabela 8: Testes de Sazonalidade

Teste	Rotina	H_0	P-valor	Conclusão
Kruskall Wallis	<i>kw</i>	Não Sazonal	$1,1102 \times 10^{-16}$	Sazonal
Friedman	<i>fried</i>	Não Sazonal	$1,4574 \times 10^{-12}$	Sazonal

2.3 Modelagem

Inicialmente, para a modelagem foi feita a divisão da série temporal entre treino e teste. Na série de treino ficou definida 100 observações, enquanto que para a série de teste ficou 8 observações. Com o objetivo de aplicar sobre o modelo de treino e depois avaliar a acurácia, no modelo de teste. Na Figura 12, através do gráfico, é possível perceber a divisão feita na série temporal.

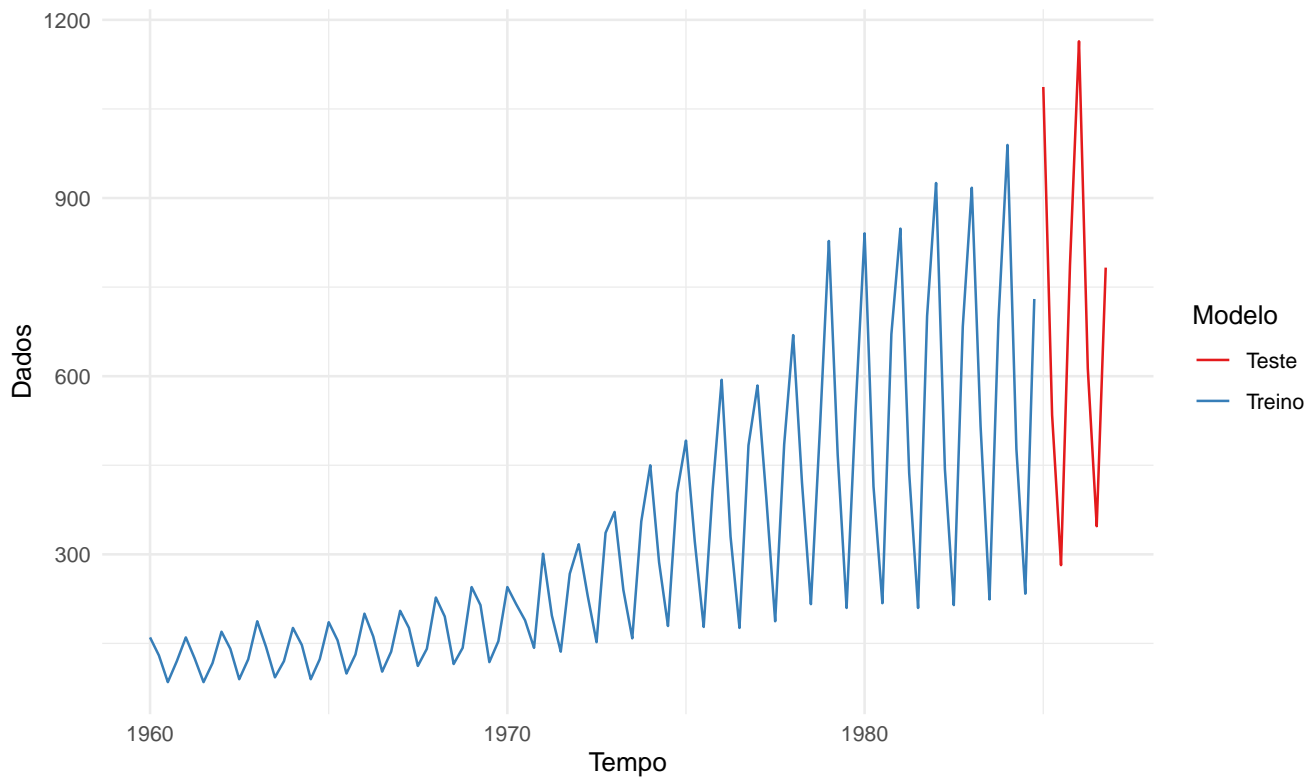


Figura 12: Gráfico da série do exercício 2 dividida em treino e teste

Foram considerados dois modelos SARIMA:

- Modelo 1 = SARIMA(3,1,1)(0,1,1)[4]
- Modelo 2 = SARIMA(0,1,1)(0,1,0)[4]
- Modelo 3 = SARIMA(0,1,1)(1,1,1)[4]

2.4 Cálculo das medidas de acurácia

Considerando o critério de informação de Akaike (AIC), todos os modelos estão próximos, afinal eles apresentaram uma pequena diferença, o modelo 1 resultou em 935,3 enquanto que o modelo 2 foi de 947,7 e por último o modelo 3 com 951,4.

Agora, na Tabela 9, é apresentada as medidas de acurácia de cada modelo, realizadas sobre a série de treino, nota-se que o modelo 2 apresentou menores erros de previsão, apenas no teste ACF1 que o modelo 1 apresentou-se menor. Com essas análises, podemos perceber que o modelo 2 teve melhor ajuste e menores erros de previsão. Portanto utilizaremos o modelo 2 para realizar as previsões.

Tabela 9: Medidas de acurácia para a série do exercício 2

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Modelo 1	59,67	71,27	59,73	10,116	10,123	-0,0256	0,1083
Modelo 2	57,45	68,50	57,45	9,038	9,038	0,1407	0,1031
Modelo 3	57,56	69,04	57,56	8,973	8,973	0,1223	0,1041

2.5 Análise de resíduo do modelo escolhido

Na Figura 13 é apresentado o gráfico com os círculos unitários, em que são calculadas as inversas das raízes de cada polinômio autoregressivo e de médias móveis. Nesse caso todas as raízes estão dentro do círculo unitário, logo eles

são inversíveis e estacionários.

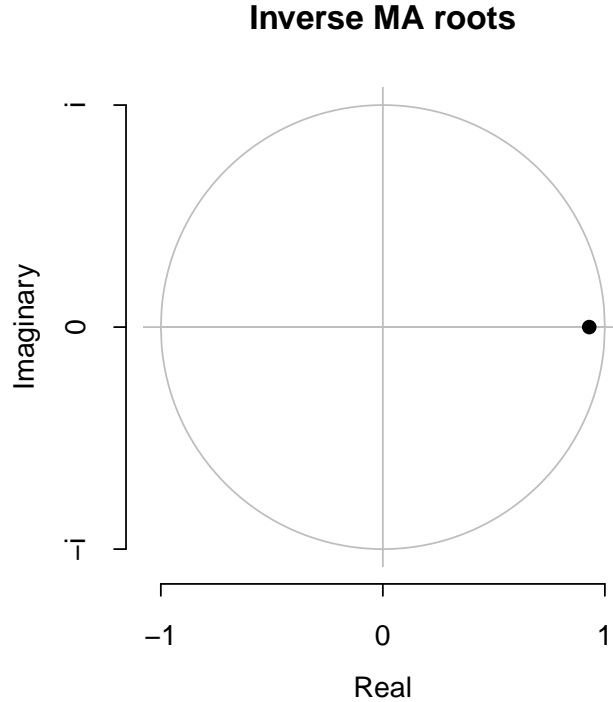


Figura 13: Representação do círculo unitário para o modelo 2

Pelo Box-Ljung test os resíduos não apresentaram correlação, com $p\text{-valor} = 0,1$, e utilizando $\text{lag} = 15$. Pelo Dickey-Fuller Test os resíduos apresentaram estacionariedade, com $p\text{-valor} = 0,01$. Os resíduos não apresentaram normalidade (shapiro.test , $p\text{-valor} = 2 \times e^{-05}$).

2.6 Predição/Previsão do modelo escolhido

Na Tabela 10 é apresentado as previsões para os próximos 8 trimestres utilizando o modelo 2, nota-se que a cada mês espera-se uma alteração entre 385,4 a 1239,9 e pela Figura 14 vemos esse ajuste da predição do modelo SARIMA(0,0,1)(0,1,0)[4] e da previsão de 8 trimestres para a série original, com os limites 95%.

Tabela 10: Previsão de 8 trimestres para o modelo 2

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1987 Q1	1201,9	1157,0	1246,7	1133,3	1270,5
1987 Q2	651,1	606,1	696,1	582,3	719,9
1987 Q3	385,4	340,3	430,5	316,5	454,3
1987 Q4	820,8	775,6	866,0	751,7	889,9
1988 Q1	1239,9	1174,0	1305,8	1139,1	1340,7
1988 Q2	689,1	622,9	755,3	587,8	790,4
1988 Q3	423,4	356,9	489,9	321,7	525,1
1988 Q4	858,8	792,0	925,6	756,6	961,0

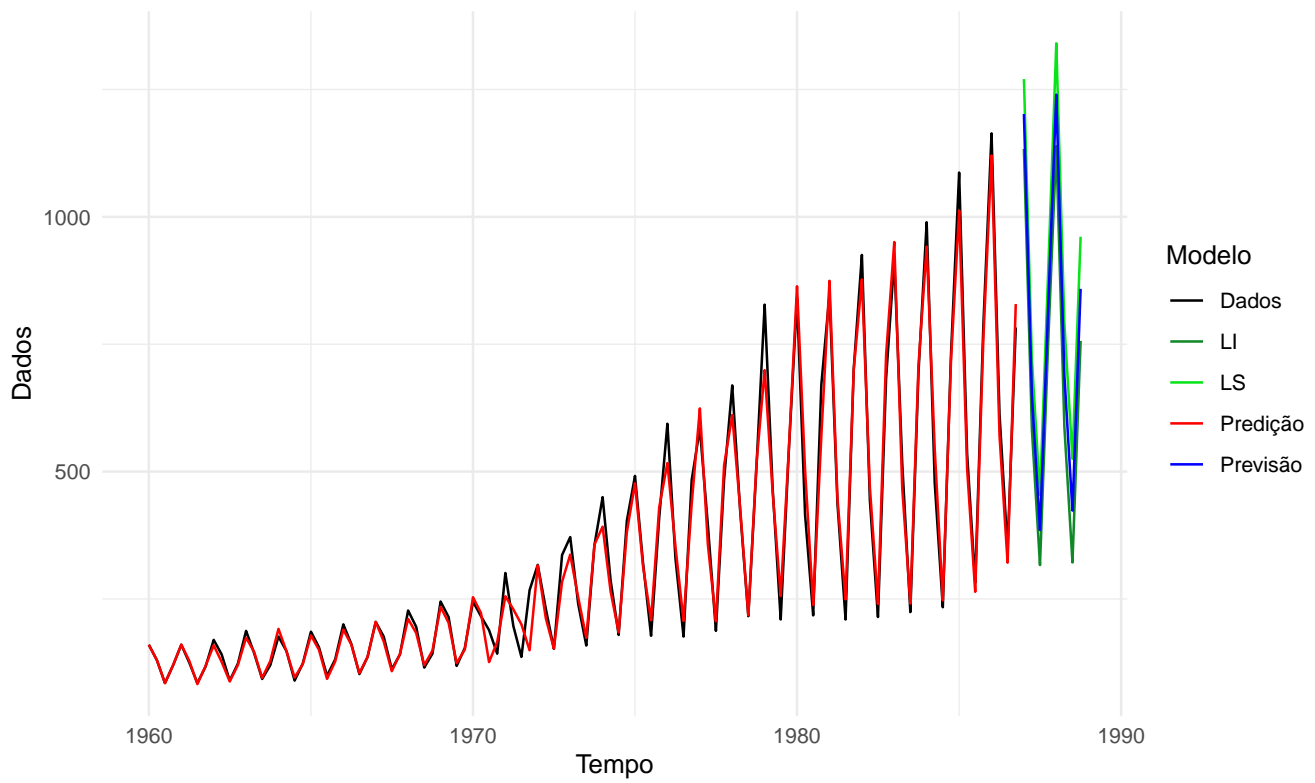


Figura 14: Gráfico da série contendo a predição do modelo final SARIMA(0,0,1)(0,1,0)[4] e predição de 8 trimestres, com limites superior e inferior de 95%.