

ExercicioExtra.R

joaopaulodecker

2025-11-18

```
# LEITURA DA BASE
base <- read.csv2("turismo_floripa.csv")

# a - Série temporal
turismo <- ts(base$turistas,
               frequency = 12,
               start = c(2015, 1))

# Separação Treino (até 12/2023) e Teste (01/2024 a 12/2024)
treino <- window(turismo, end = c(2023, 12))
teste <- window(turismo, start = c(2024, 1))

# b - AJUSTE DO MODELO
library(forecast)

## Registered S3 method overwritten by 'quantmod':
##   method           from
##   as.zoo.data.frame zoo

# Arima
mod.arima <- auto.arima(treino, seasonal = FALSE)
summary(mod.arima)

## Series: treino
## ARIMA(2,1,2) with drift
##
## Coefficients:
##             ar1      ar2      ma1      ma2     drift
##             1.6909  -0.9471  -1.9159  0.9465  4.8491
## s.e.    0.0272   0.0264   0.0465  0.0451  0.2322
##
## sigma^2 = 394.8: log likelihood = -473.07
## AIC=958.14   AICc=958.98   BIC=974.18
##
## Training set error measures:
##                  ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.255641 19.30894 14.96177 -0.7377991 4.230402 0.2510657
##                      ACF1
## Training set -0.3908359
```

```

# Sarima
mod.sarima <- auto.arima(treino, seasonal = TRUE)
summary(mod.sarima)

## Series: treino
## ARIMA(1,0,1)(2,1,0)[12] with drift
##
## Coefficients:
##             ar1      ma1      sar1      sar2     drift
##             -0.9301  0.7953  -0.8356  -0.3804  4.9594
## s.e.      0.0632  0.0855   0.1021   0.1069  0.0636
##
## sigma^2 = 272.5: log likelihood = -407.28
## AIC=826.57  AICc=827.51  BIC=841.95
##
## Training set error measures:
##                  ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.2579326 15.15144 11.35762 -0.2456343 2.718851 0.1905862
##                      ACF1
## Training set 0.1204093

# Comparação de AICc
mod.arima$aicc

## [1] 958.9818

mod.sarima$aicc

## [1] 827.5114

# Previsão para 12 meses (01/2024 a 12/2024)
prev_sarima <- forecast(mod.sarima, h = 12)
summary(prev_sarima)

##
## Forecast method: ARIMA(1,0,1)(2,1,0)[12] with drift
##
## Model Information:
## Series: treino
## ARIMA(1,0,1)(2,1,0)[12] with drift
##
## Coefficients:
##             ar1      ma1      sar1      sar2     drift
##             -0.9301  0.7953  -0.8356  -0.3804  4.9594
## s.e.      0.0632  0.0855   0.1021   0.1069  0.0636
##
## sigma^2 = 272.5: log likelihood = -407.28
## AIC=826.57  AICc=827.51  BIC=841.95
##
## Error measures:
##                  ME      RMSE      MAE      MPE      MAPE      MASE

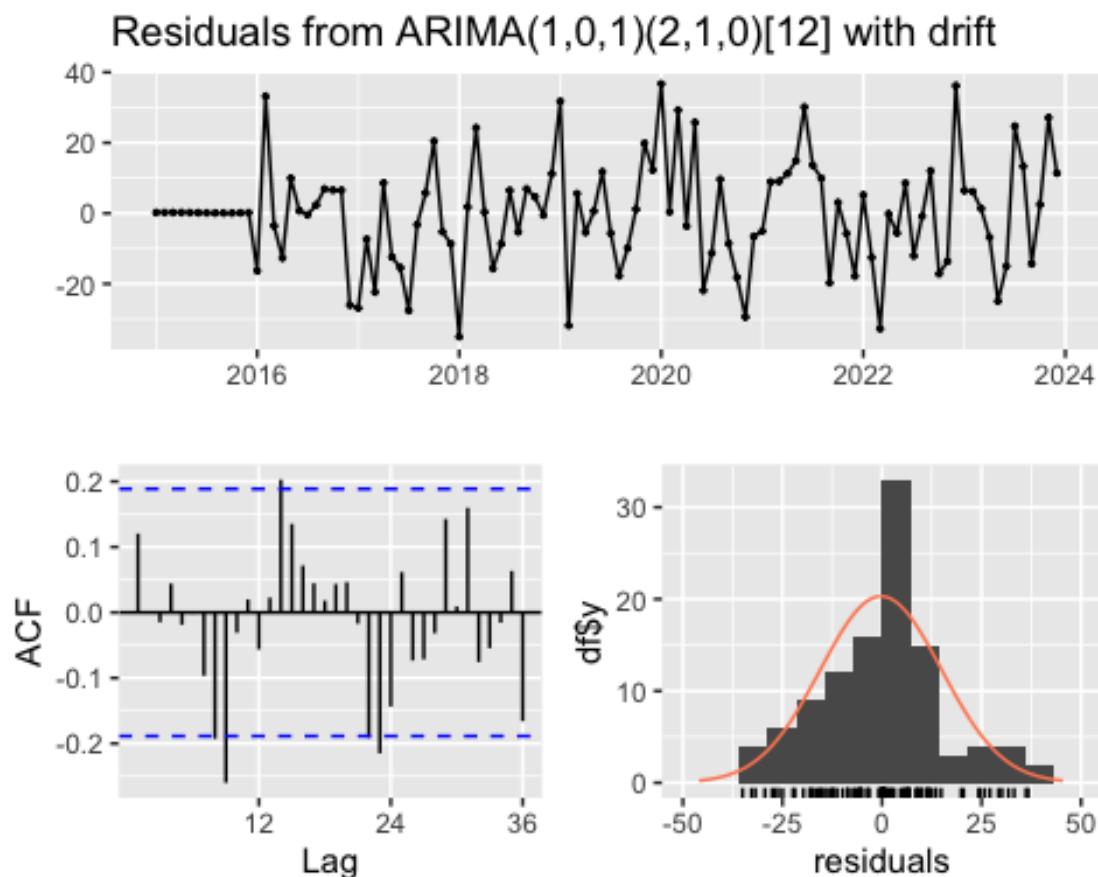
```

```

## Training set -0.2579326 15.15144 11.35762 -0.2456343 2.718851 0.1905862
##                               ACF1
## Training set 0.1204093
##
## Forecasts:
##             Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
## Jan 2024      742.0209 720.8674 763.1743 709.6694 774.3723
## Feb 2024      765.6358 744.2912 786.9804 732.9921 798.2795
## Mar 2024      772.5949 751.0864 794.1035 739.7005 805.4894
## Apr 2024      771.9997 750.3504 793.6491 738.8899 805.1095
## May 2024      752.0810 730.3106 773.8514 718.7860 785.3759
## Jun 2024      725.5201 703.6455 747.3947 692.0658 758.9744
## Jul 2024      678.5671 656.6028 700.5314 644.9756 712.1586
## Aug 2024      661.8026 639.7610 683.8442 628.0928 695.5123
## Sep 2024      650.4373 628.3290 672.5455 616.6256 684.2490
## Oct 2024      663.5892 641.4235 685.7550 629.6896 697.4889
## Nov 2024      700.8582 678.6428 723.0735 666.8827 734.8337
## Dec 2024      763.3268 741.0686 785.5850 729.2858 797.3678

# c - DIAGNÓSTICO DE RESÍDUOS DO MODELO ESCOLHIDO
checkresiduals(mod.sarima)

```



```

## 
## Ljung-Box test
## 
## data: Residuals from ARIMA(1,0,1)(2,1,0)[12] with drift
## Q* = 30.26, df = 18, p-value = 0.03499
## 
## Model df: 4. Total lags used: 22

```

Análise de Diagnóstico

O modelo selecionado foi o SARIMA, que apresentou $AICc = 827.51$, valor bem menor que o modelo não sazonal ($AICc = 958.98$). Isso indica que o modelo sazonal descreve melhor a série. Os coeficientes estimados também foram consistentes: o termo $AR(1) = -0.9301$ e o $MA(1) = 0.7953$ capturam a dinâmica de curto prazo, enquanto os termos sazonais $SAR(1) = -0.8356$ e $SAR(2) = -0.3804$ representam a forte sazonalidade anual presente nos dados.

Na análise de diagnóstico, os resíduos ficaram distribuídos de forma aleatória ao redor de zero, sem tendência ou padrão. A ACF dos resíduos não apresentou autocorrelações significativas, o que indica que o modelo conseguiu explicar bem a estrutura temporal da série.

Dessa forma, tanto os coeficientes obtidos quanto o comportamento dos resíduos indicam que o modelo ajustado é adequado e representa bem a série temporal utilizada.

```

# d - ANÁLISE GRÁFICA DO AJUSTE
library(ggplot2)

# Criação de data frames pro gráfico
df_treino <- data.frame(
  tempo = time(treino),
  valor = as.numeric(treino),
  serie = "Treino"
)

df_teste <- data.frame(
  tempo = time(teste),
  valor = as.numeric(teste),
  serie = "Teste"
)

df_prev <- data.frame(
  tempo = time(prev_sarima$mean),
  valor = as.numeric(prev_sarima$mean),

```

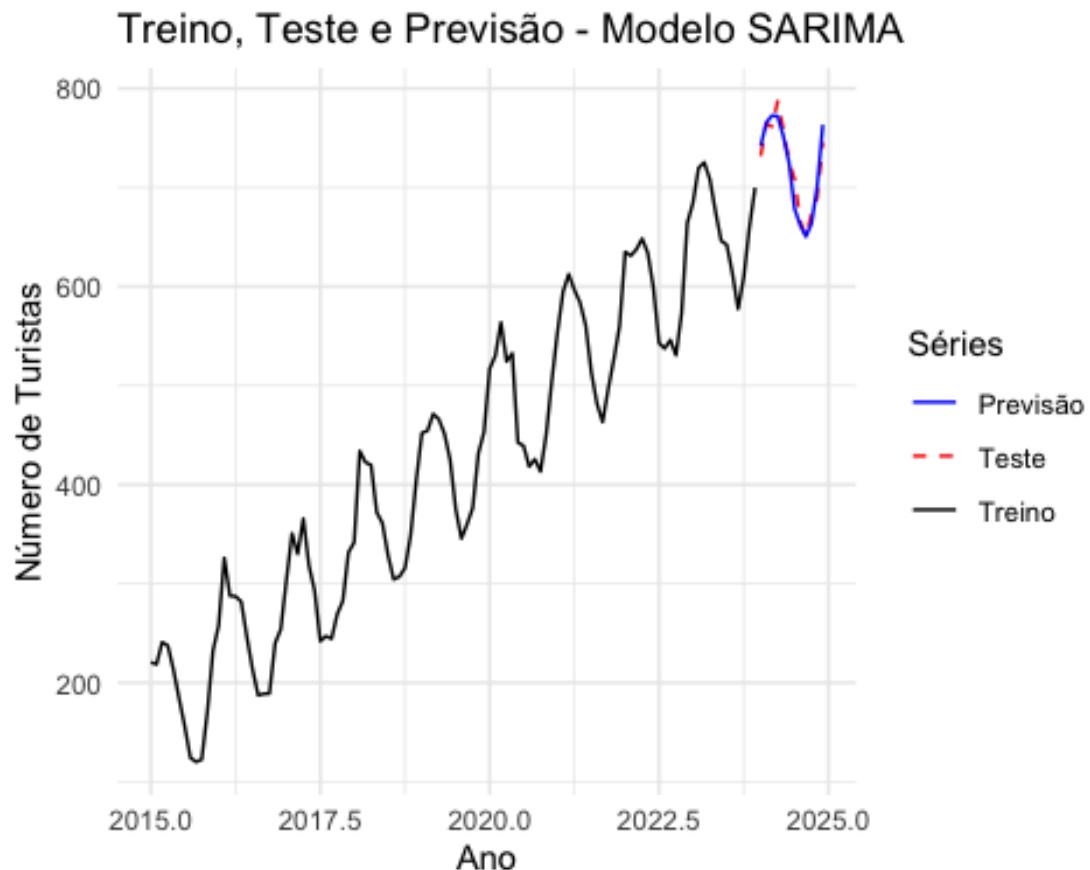
```

    serie = "Previsão"
)

# Gráfico
ggplot() +
  geom_line(data = df_treino, aes(x = tempo, y = valor, color = "Treino")) +
  geom_line(data = df_teste, aes(x = tempo, y = valor, color = "Teste"),
            linetype = "dashed") +
  geom_line(data = df_prev, aes(x = tempo, y = valor, color = "Previsão")) +
  scale_color_manual(values = c("Treino" = "black",
                                "Teste" = "red",
                                "Previsão" = "blue")) +
  labs(title = "Treino, Teste e Previsão - Modelo SARIMA",
       x = "Ano",
       y = "Número de Turistas",
       color = "Séries") +
  theme_minimal()

## Don't know how to automatically pick scale for object of type <ts>.
Defaulting
## to continuous.

```

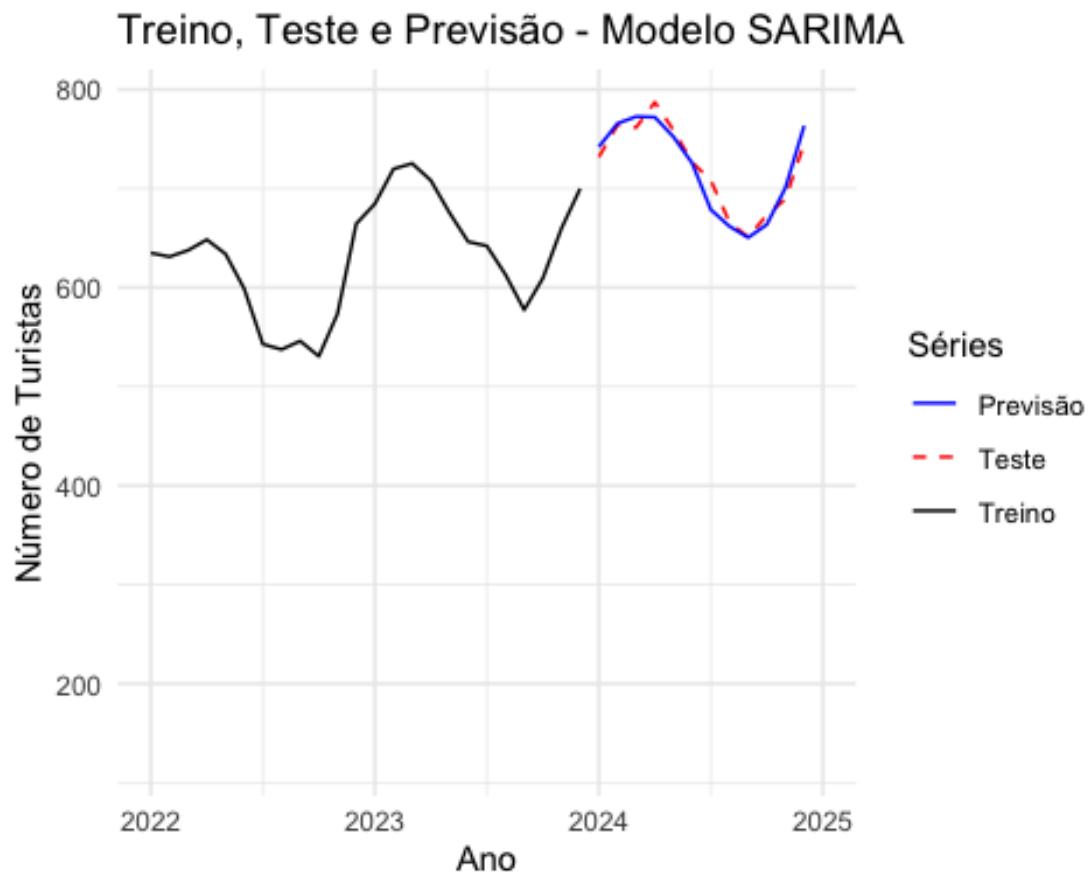


```

# Gráfico com zoom
ggplot() +
  geom_line(data = df_treino, aes(x = tempo, y = valor, color = "Treino")) +
  geom_line(data = df_teste, aes(x = tempo, y = valor, color = "Teste"),
            linetype = "dashed") +
  geom_line(data = df_prev, aes(x = tempo, y = valor, color = "Previsão")) +
  scale_color_manual(values = c("Treino" = "black",
                               "Teste" = "red",
                               "Previsão" = "blue")) +
  xlim(c(2022, 2025)) +
  labs(title = "Treino, Teste e Previsão - Modelo SARIMA",
       x = "Ano",
       y = "Número de Turistas",
       color = "Séries") +
  theme_minimal()

## Warning: Removed 84 rows containing missing values or values outside the
## scale range
## (`geom_line()`).

```



```

# e - GRÁFICO SOMENTE COM O TESTE E A PREVISÃO COM IC 95%
autoplot(prev_sarima) +
  autolayer(teste, series = "Real", color = "black") +

```

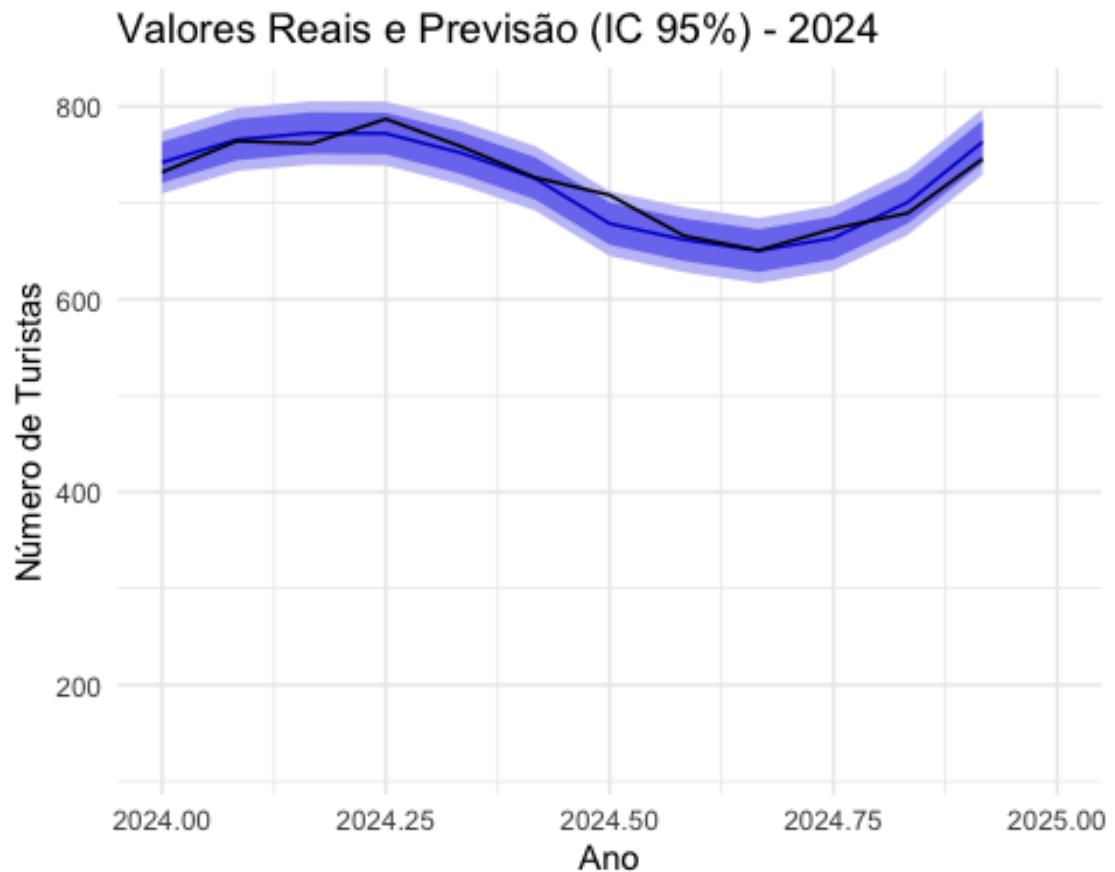
```

labs(title = "Valores Reais e Previsão (IC 95%) - 2024",
     x = "Ano",
     y = "Número de Turistas") +
  xlim(c(2024, 2025)) +
  theme_minimal()

## Scale for x is already present.
## Adding another scale for x, which will replace the existing scale.

## Warning: Removed 108 rows containing missing values or values outside the
## scale range
## (`geom_line()`).

```



Obs:

linha preta - valores reais do teste;

Linha azul - valores previstos com Arima;

Faixa azul mais escura - IC 80%

Fixa azul mais clara - IC 95%

f - Interpretação

Ao observar o gráfico, nota-se que a linha de previsão (em azul) acompanha bem o comportamento geral da série durante todo o ano. Além disso, todos os valores reais (linha preta) permanecem dentro do intervalo de confiança de 95%, indicado pela faixa azul mais clara ao redor da previsão.

Isso significa que, mesmo nos meses em que a previsão não coincidiu exatamente com os valores reais, o modelo ainda conseguiu prever dentro do nível de incerteza esperado. Assim, o SARIMA apresentou um desempenho satisfatório, capturando bem tanto a tendência quanto o padrão sazonal da série. De forma geral, o modelo pode ser considerado bem ajustado para fins de previsão.