

Descrição e Aplicações na Série Temporal de Prisioneiros em Prisões Federais nos Estado Unidos

Caroline Assis de Oliveira

Descritivas

O banco que será usado é composto pelas quantidades de encarcerados, de ambos os sexos, em Prisões Federais dos Estado Unidos de 1990 até 2019. O banco de dados foi retirado do site do Departamento Federal de Prisões (BOP - Federal Bureau of Prisons). O banco de dados, script e ao trabalho anterior são encontrados clicando aqui.

Tabela 1: Estatísticas Descritivas

Estatísticas	Valores
Minímo	65526
1º Quartil	104015
Mediana	171507
Média	157504
3º Quartil	197924
Máximo	217815

Série Temporal de Prisioneiros em Prisões Federais

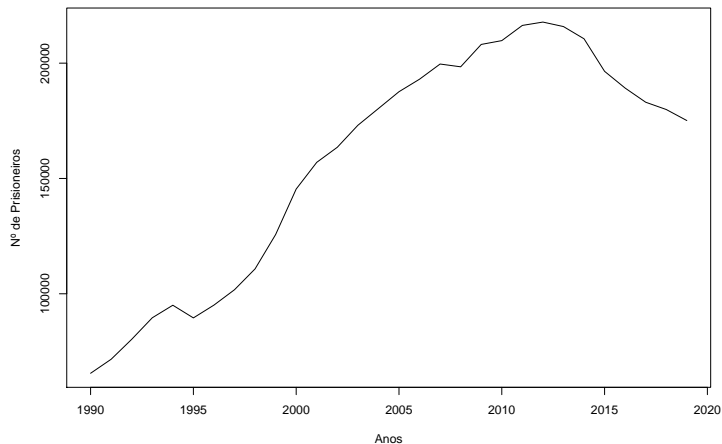


Figura 1: Gráfico da série temporal Federal

Com o gráfico podemos notar que em geral existe uma tendência crescente de encarcerados em prisões federais ao longo dos anos, porém entre 2010 e 2015 vemos que começa uma queda na série, segundo o BOP, acredita-se que esses números decaíram por conta de algumas políticas de segurança pública em que infratores com sentença não violenta, não séria e que não são infratores sexuais fossem transferidos para prisões locais, e em 2015, onde existe a maior queda, vem da política de libertação de infratores não-violentos envolvidos em crimes com drogas. Outro motivo de decaimento nos últimos anos da séries pode ser explicado pelo aumento do número de execuções durante o governo Trump.

ACF e PACF

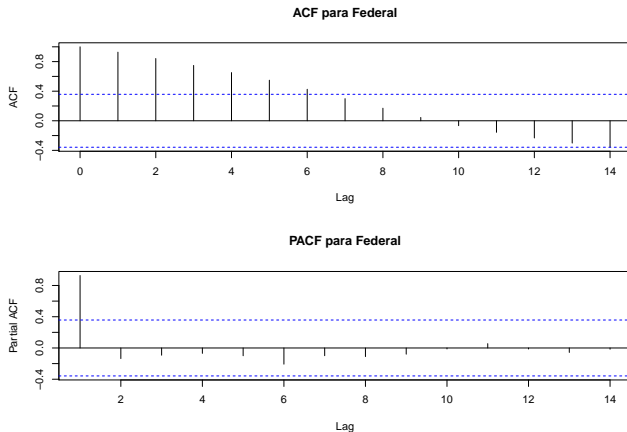


Figura 2: ACF e PACF da série temporal Federal

É possível notar no gráfico acf que as autocorrelações tem uma queda lenta, diferente no observado no gráfico pacf.

Estacionariedade

Para checar a estacionariedade de uma série existem alguns testes de hipóteses, aqui foi utilizado o teste de Phillips-Perron (também conhecido como teste pp) com as seguintes hipóteses:

H_0 : Existe pelo menos uma raiz dentro do círculo unitário

H_1 : Não existem raízes dentro do círculo unitário

Ao aplicar o teste obtemos um p-valor = 0,769, com nível de significância de 5% ($\alpha = 0,05$), rejeitamos a hipótese nula, portanto nossa série **não é estacionária**.

Quando nos deparamos com esse problema podemos seguir alguns caminhos, dois deles são: **transformação e diferenciação**.

Transformação

Aplicando a transformação log na série:

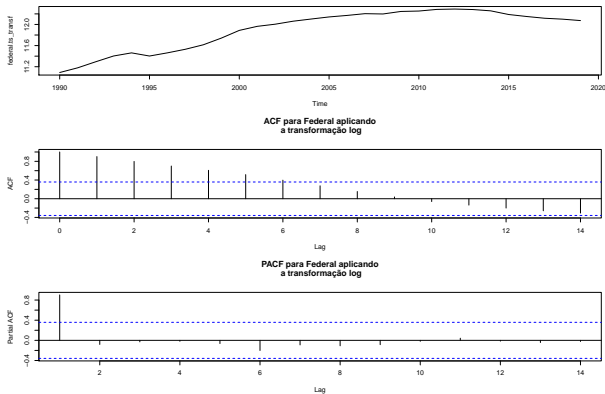


Figura 3: Gráficos para a série transformada

Podemos ver que mesmo aplicando a transformação não foi obtido o resultado desejado.

Diferenciação

Utilizando o comando `ndiffs()` sabe-se que são necessárias duas diferenças para tornar a série estacionária, logo, com a aplicação temos o seguinte resultado:

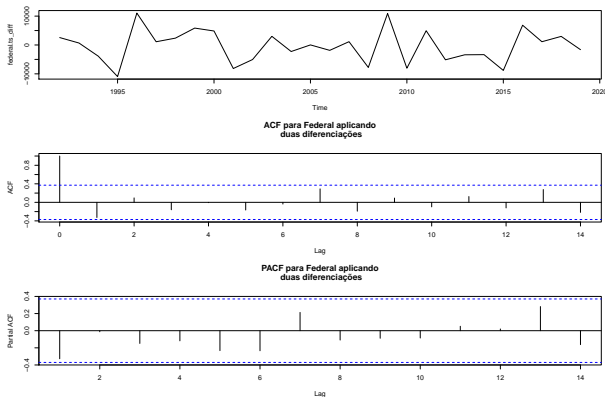


Figura 4: Gráficos para a série diferenciada

Podemos ver agora que o nosso gráfico acf tem uma queda rápida diferente de antes, o que nos indica estacionariedade na série. O teste pp nos confirma que a série é estacionária depois das diferenças, já que o nosso $p\text{-valor} = 0,01$.

Decomposição

Não foi possível fazer a decomposição da série pois ela não tem período, logo não é possível identificar tendência e sazonalidade, a saída que o R nos dá é

```
> decompose(federal.ts)  
Error in decompose(federal.ts) : série temporal não  
tem período, ou tem menos de 2
```

Estimação

Para a estimação será usado o comando `auto.arima()` para encontrar o melhor modelo para esses dados de acordo com os métodos **AIC** e **BIC**, e a previsão será feita para 3 dos últimos 5 anos dos dados (2014-2019): 2015, 2017 e 2019.

Para ambos os métodos, o melhor modelo é o $ARIMA(1,1,0)$.

Checagem do modelo

Fazendo os testes para os resíduos do modelo ARIMA(1,1,0), podemos ver que não há correlação entre eles, confirmado pelo **teste de Box-Pierce** que tem $p\text{-valor} = 0,2249$ (maior que nosso α), e que os resíduos seguem distribuição normal (de acordo com o **teste de lilliefors**, $p\text{-valor} = 0,9587 > \alpha = 0,05$).

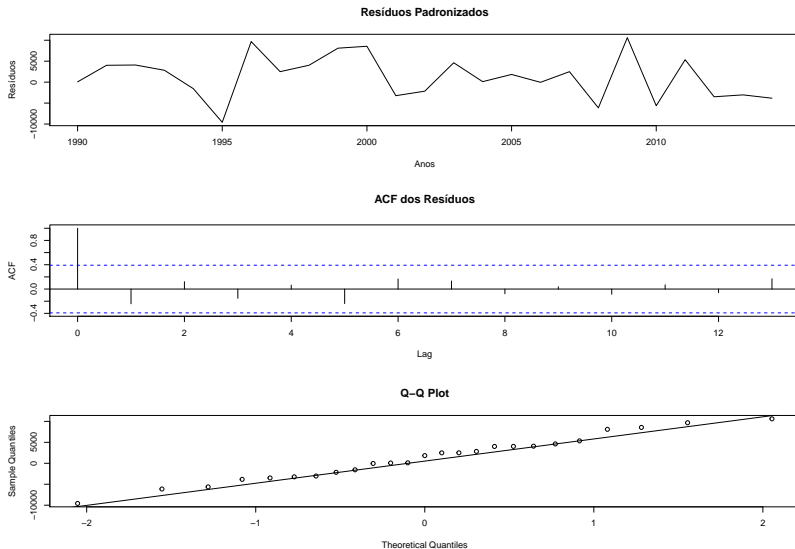


Figura 5: Gráficos dos Resíduos

Na tabela abaixo é possível comparar os valores reais e os valores estimados pelo modelo ARIMA(1,1,0). É possível observar que os valores são próximos, porém os valores preditos são maiores que os valores reais.

Tabela 2: Valores reais e previstos do modelo ARIMA(1,1,0)

	Dados Reais	Valores Previstos
2015	196455	206589.5
2017	183058	201362.9
2019	175116	198418.2

Métricas

Como não houve diferença entre os métodos, não existe a necessidade de comparação entre os métodos, mas será necessário para a comparação entre o modelo ARIMA e o Alisamento Exponencial que será feito a seguir. As métricas que serão utilizadas são **RMSE**, **MAE** e **MPE**.

O RMSE é dado por:

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

O MAE é dado por:

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

O MPE é dado por:

$$MPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \times 100$$

O modelo ARIMA(1,1,0) tem variância $\sigma^2 = 28459974$ e os valores para as métricas podem ser observados na tabela abaixo.

Tabela 3: Métricas para o modelo ARIMA(1,1,0)

	RMSE	MAE	MPE(%)
2015	10134.50	10134.50	-5.158688
2017	14668.36	14283.79	-7.591947
2019	17775.45	17187.27	-9.415846

Alisamento Exponencial

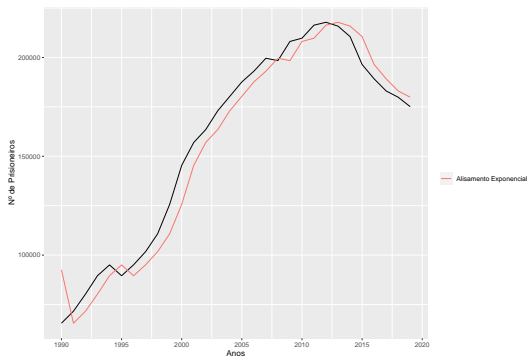


Figura 6: Gráfico comparando a Série Temporal com o Alisamento Exponencial

Estimação com Alisamento Exponencial

Na tabela abaixo é possível comparar os valores reais e os valores estimados pelo Alisamento Exponencial. É possível observar que os valores estimados são os mesmos para todos os anos, um possível motivo para isso é o fato da série não ter tendência e nem sazonalidade.

Tabela 4: Valores reais e previstos pelo Alisamento Exponencial

	Dados Reais	Valores Previstos
2015	196455	210567.5
2017	183058	210567.5
2019	175116	210567.5

O alisamento exponencial tem variância $\sigma^2 = 10176,49$, e as métricas são dadas na tabela abaixo.

Tabela 5: Métricas para o Alisamento Exponencial

	RMSE	MAE	MPE(%)
2015	14112.53	14112.53	-7.183594
2017	21701.39	20999.20	-11.169896
2019	26871.11	25823.73	-14.160515

Comparações entre modelos

Tabela 6: Comparação de métricas

	RMSE	MAE	MPE(%)
ARIMA-2015	10134.50	10134.50	-5.158688
ARIMA-2017	14668.36	14283.79	-7.591947
ARIMA-2019	17775.45	17187.27	-9.415846
AE-2015	14112.53	14112.53	-7.183594
AE-2017	21701.39	20999.20	-11.169896
AE-2019	26871.11	25823.73	-14.160515

Conclusões

Vimos que mesmo apresentando uma variância menor, o alisamento exponencial exibe valores maiores para as métricas de qualidade de ajuste, exceto pelo MPE (tabela 6), sendo assim é certo dizer que para a predição de valores o modelo $ARIMA(1,1,0)$ é a melhor opção para esses dados.