

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

Caroline Assis de Oliveira, 20180011385

Descritivas

O banco que será usado é composto pelas quantidades de encarcerados, de ambos os sexos, divididos em Prisões Federais e Prisões Estaduais dos Estado Unidos e por ano, indo de 1990 até 2019. O banco de dados e os sites de onde foram retirados, assim como esse documento podem ser encontrados clicando aqui

Importando os dados e fazendo uma breve análise descritiva

```
dados = file.choose()
dados = read.table(dados, header = T, sep = ";")
attach(dados)
```

```
str(dados)
```

```
## 'data.frame': 30 obs. of 3 variables:
## $ Ano : int 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...
## $ Federal : int 65526 71608 80259 89587 95034 89538 95088 101755 110793 125682 ...
## $ Estadual: int 708393 754011 802241 880857 958704 989004 1032440 1074809 1113676 1161490 ...
```

```
summary(dados)
```

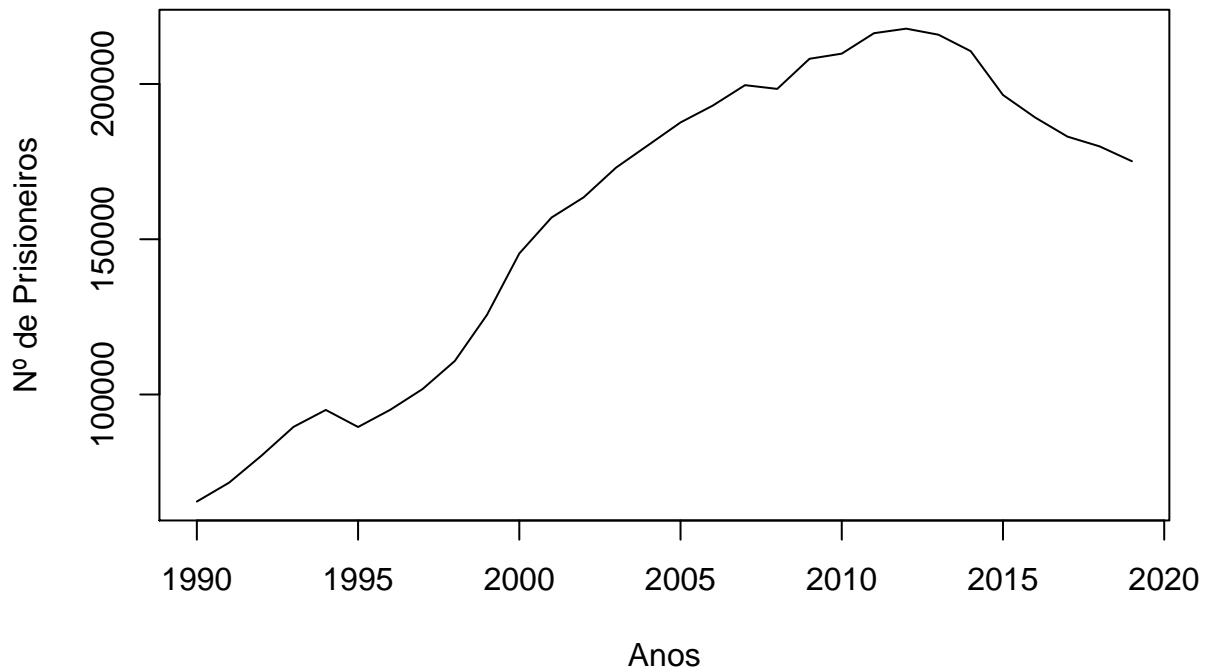
	Ano	Federal	Estadual
##	Min. :1990	Min. : 65526	Min. : 708393
##	1st Qu.:1997	1st Qu.:104015	1st Qu.:1084526
##	Median :2004	Median :177507	Median :1290015
##	Mean :2004	Mean :157504	Mean :1204552
##	3rd Qu.:2012	3rd Qu.:197924	3rd Qu.:1352375
##	Max. :2019	Max. :217815	Max. :1408479

Podemos verificar com as descritivas que não existe nenhum dado com características que fogem do padrão esperado, e que os dados para prisões estaduais estão mais próximos de ser simétricos do que os dados para as prisões federais. Para este trabalharei apenas com a série temporal de **prisioneiros em prisões federais**.

Série Temporal de Prisioneiros em Prisões Federais

Vamos agora aplicar a função 'ts()' para o R entender que esses dados são uma série temporal

```
federal.ts = ts(Federal, start = 1990, end = 2019, frequency = 1)
plot(federal.ts, xlab = "Anos", ylab = "Nº de Prisioneiros")
```

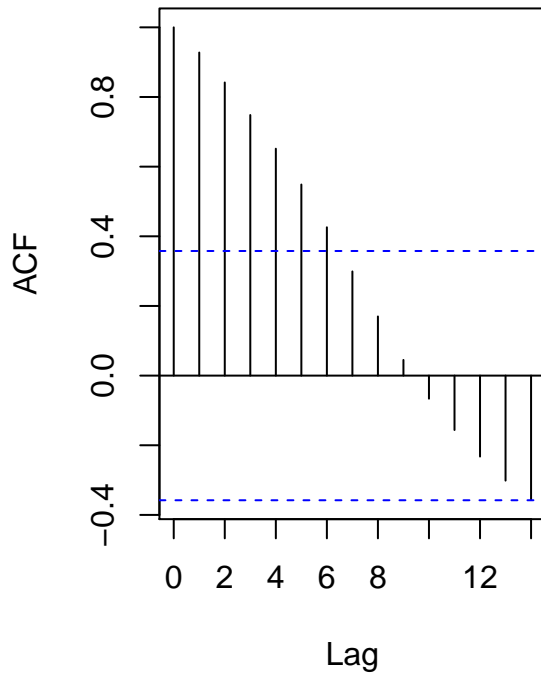


Com o gráfico podemos notar que em geral existe uma tendência crescente de encarcerados em prisões federais ao longo dos anos, porém entre 2010 e 2015 vemos que começa uma queda na série, segundo o BOP (Federal Bureau of Prisons) acredita-se que esses números decaíram por conta de algumas políticas de 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, são graças a política de libertação de infratores não-violentos envolvidos em crimes com drogas. Outro motivo de decaimento nos últimos anos da séries podem ser explicadas pelo aumento do número de execuções durante o governo Trump.

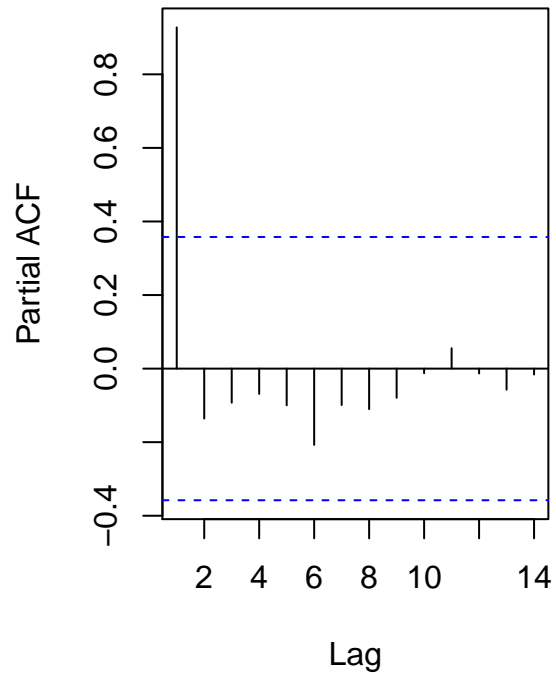
Agora iremos fazer algumas análises sobre as autocorrelações destes dados.

```
par(mfrow=c(1,2))
acf(federal.ts, main = "ACF para Federal")
pacf(federal.ts, main = "PACF para Federal")
```

ACF para Federal



PACF para Federal



Podemos notar no gráfico acf que as autocorrelações tem uma queda lenta, diferente no observado no gráfico pacf. Agora iremos relizar o teste Phillips-Perron (também conhecido apenas por 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

para ver se a série é estacionária ou não, com $\alpha = 0,05$

```
stationary.test(federal.ts, method = "pp")
```

```
## Phillips-Perron Unit Root Test
## alternative: stationary
##
## Type 1: no drift no trend
## lag Z_rho p.value
## 2 0.404 0.769
## -----
## Type 2: with drift no trend
## lag Z_rho p.value
## 2 -2.24 0.73
## -----
## Type 3: with drift and trend
## lag Z_rho p.value
## 2 1.56 0.99
## -----
## Note: p-value = 0.01 means p.value <= 0.01
```

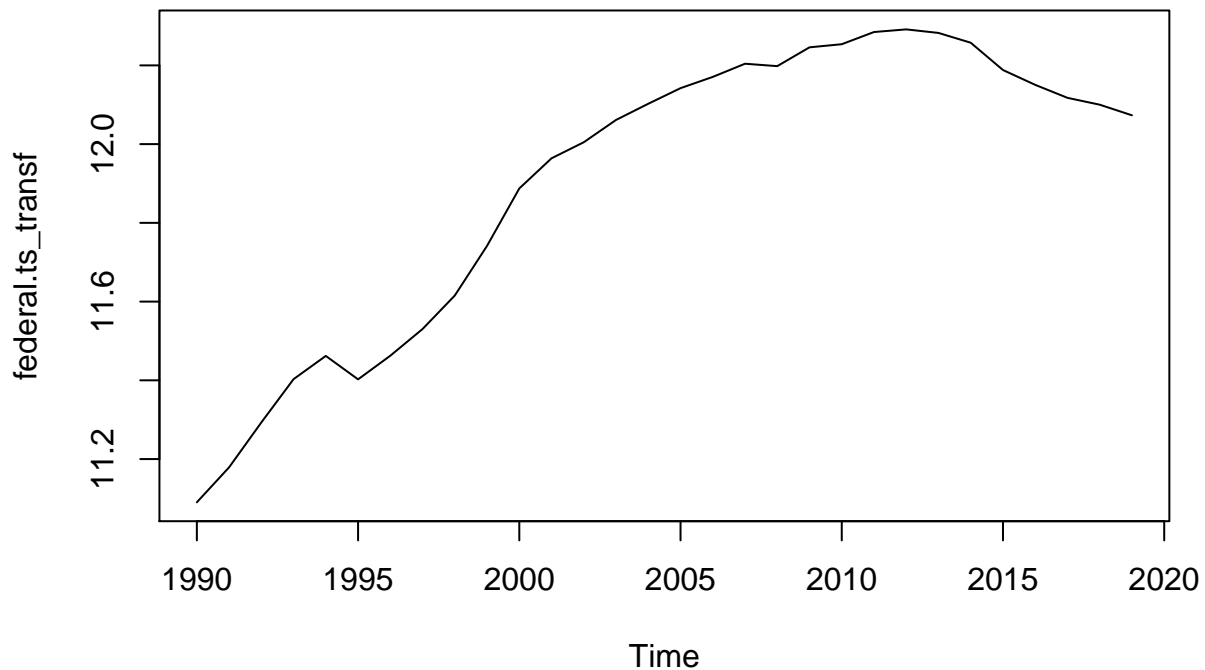
```
ndiffs(federal.ts)
```

```
## [1] 2
```

Pelos p-valores (todos maiores que nosso α) temos que essa série não é estacionária, o que corroborado pela função 'ndiffs()', que nos retorna um valor diferente de 0. Quando nos deparamos com esse problema podemos seguir alguma caminhos, aqui seguiremos por dois: **transformação e diferenciação**.

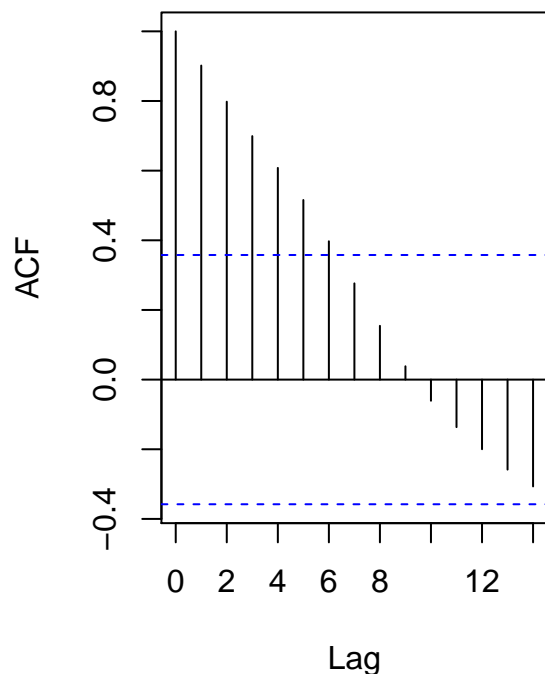
Faremos primeiro a **transformação** dos dados para ver se conseguimos transformar essa série em estacionária.

```
federal.ts_transf = ts(log(Federal), start = 1990, end = 2019,  
                        frequency = 1)  
plot(federal.ts_transf)
```

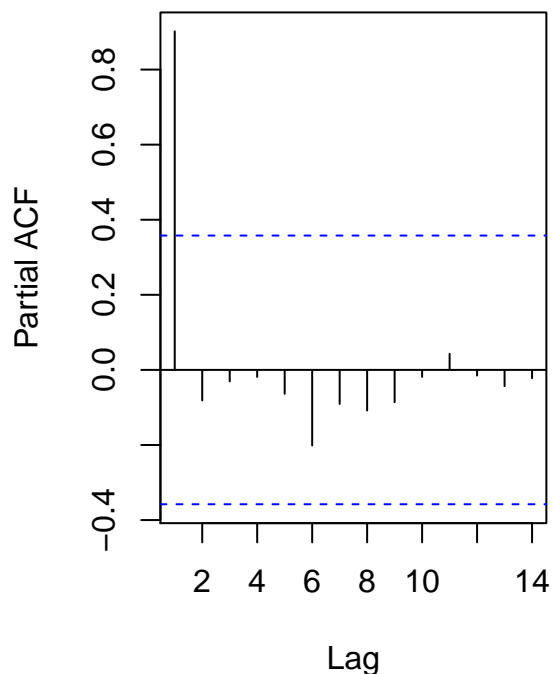


```
par(mfrow=c(1,2))  
acf(federal.ts_transf,  
    main = "ACF para Federal aplicando\n a transformação log")  
pacf(federal.ts_transf,  
     main = "PACF para Federal aplicando\n a transformação log")
```

ACF para Federal aplicando a transformação log



PACF para Federal aplicando a transformação log



```
stationary.test(federal.ts_transf, method = "pp")
```

```
## Phillips-Perron Unit Root Test
## alternative: stationary
##
## Type 1: no drift no trend
## lag Z_rho p.value
## 2 0.0795 0.698
## -----
## Type 2: with drift no trend
## lag Z_rho p.value
## 2 -2.8 0.665
## -----
## Type 3: with drift and trend
## lag Z_rho p.value
## 2 0.302 0.99
## -----
## Note: p-value = 0.01 means p.value <= 0.01
```

```
ndiffs(federal.ts_transf)
```

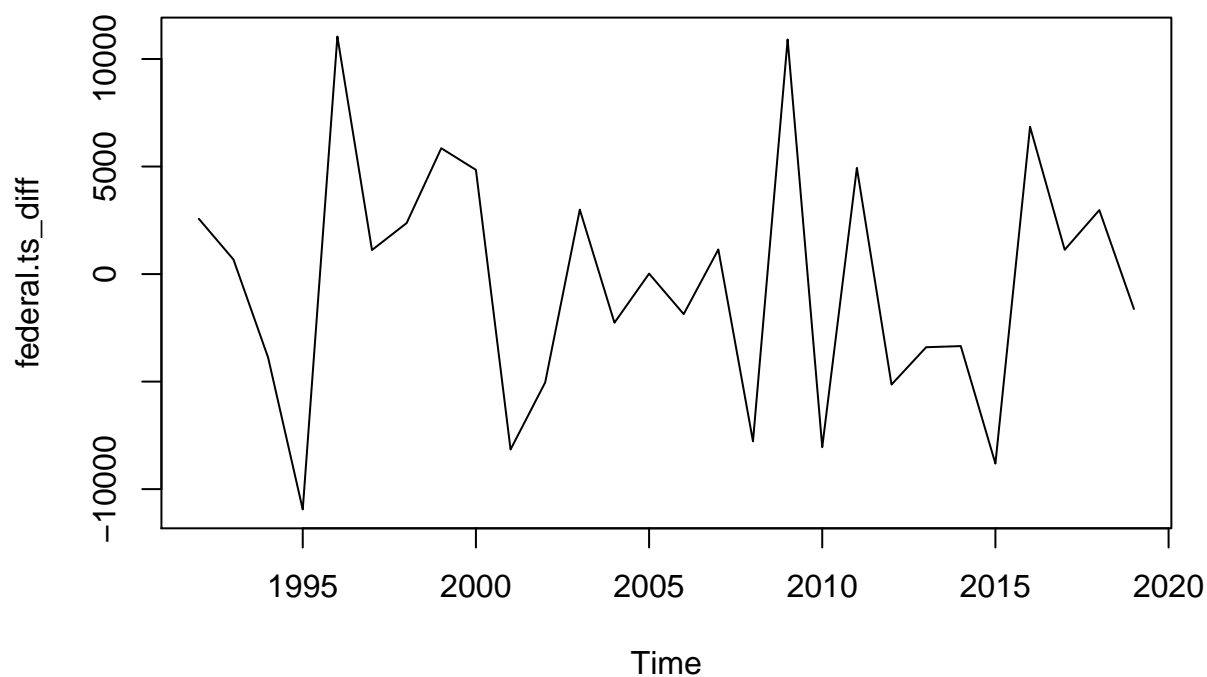
```
## [1] 2
```

Vemos que com a transformação a série continua com aparente tendência crescente, acf e pacf continuam com as mesmas características, e nosso teste de estacionariedade continua nos indicando que a série não é estacionária.

Então partiremos para a **diferenciação**, o comando 'ndiffs()' nos retorna o valor 2, que no caso é a quantidade

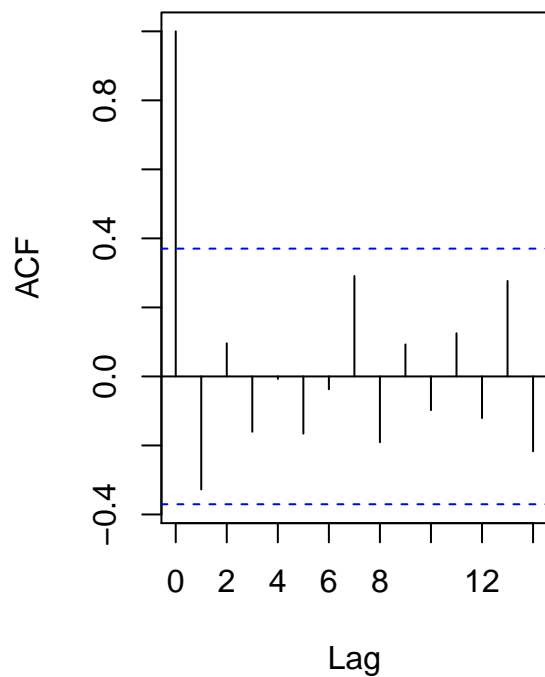
de diferenciações que são necessárias para transformar essa série em estacionária, portanto faremos as duas diferenciações seguidas dos testes feitos acima.

```
federal.ts_diff = diff(diff(federal.ts))  
plot(federal.ts_diff)
```

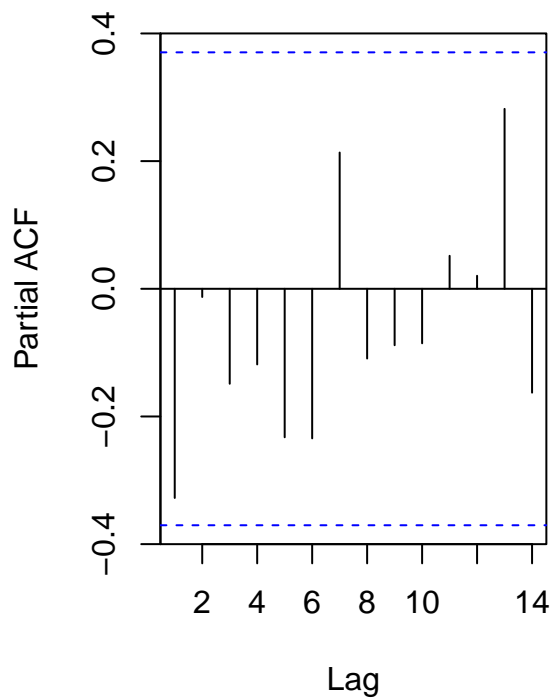


```
par(mfrow=c(1,2))  
acf(federal.ts_diff,  
    main = "ACF para Federal aplicando\n duas diferenciações")  
pacf(federal.ts_diff,  
     main = "PACF para Federal aplicando\n duas diferenciações")
```

ACF para Federal aplicando duas diferenciações



PACF para Federal aplicando duas diferenciações



```
stationary.test(federal.ts_diff, method = "pp")
```

```
## Phillips-Perron Unit Root Test
## alternative: stationary
##
## Type 1: no drift no trend
## lag Z_rho p.value
## 2 -35.4 0.01
## -----
## Type 2: with drift no trend
## lag Z_rho p.value
## 2 -35.3 0.01
## -----
## Type 3: with drift and trend
## lag Z_rho p.value
## 2 -35.3 0.01
## -----
## Note: p-value = 0.01 means p.value <= 0.01
```

```
ndiffs(federal.ts_diff)
```

```
## [1] 0
```

Aplicando as duas diferenciações vemos que a tendência dos nossos dados mudou, não tendo como reconhecer nenhum tipo de tendência. Podemos ver agora que o nosso gráfico acf tem uma queda rápida diferente de antes, e que o teste pp agora nos indica estacionariedade.