

O GRANDE TÍTULO

Alisson Rosa e Vítor Pereira

Resumo

One Piece > Naruto

Sumário

1	Introdução	1
2	Informações	1
3	Testes	5
3.1	Tendência	5
3.2	Sazonalidade	5
3.3	Raiz unitária	6
3.4	Resultados	6
4	Modelagem	7
4.1	Médias Móveis Simples	7
4.2	Regressão LOESS	8
4.3	Suavização Exponencial	8
4.4	Calculando Métricas de Acurácia dos Modelos não-ModelTime	11
	Referências	11

1 Introdução

Naruto é uma animação japonesa (anime) que adapta a série de mangá escrita e ilustrada por Masashi Kishimoto, que conta a história de Naruto Uzumaki, um jovem ninja que constantemente procura por reconhecimento e sonha em se tornar Hokage, o ninja líder de sua vila.

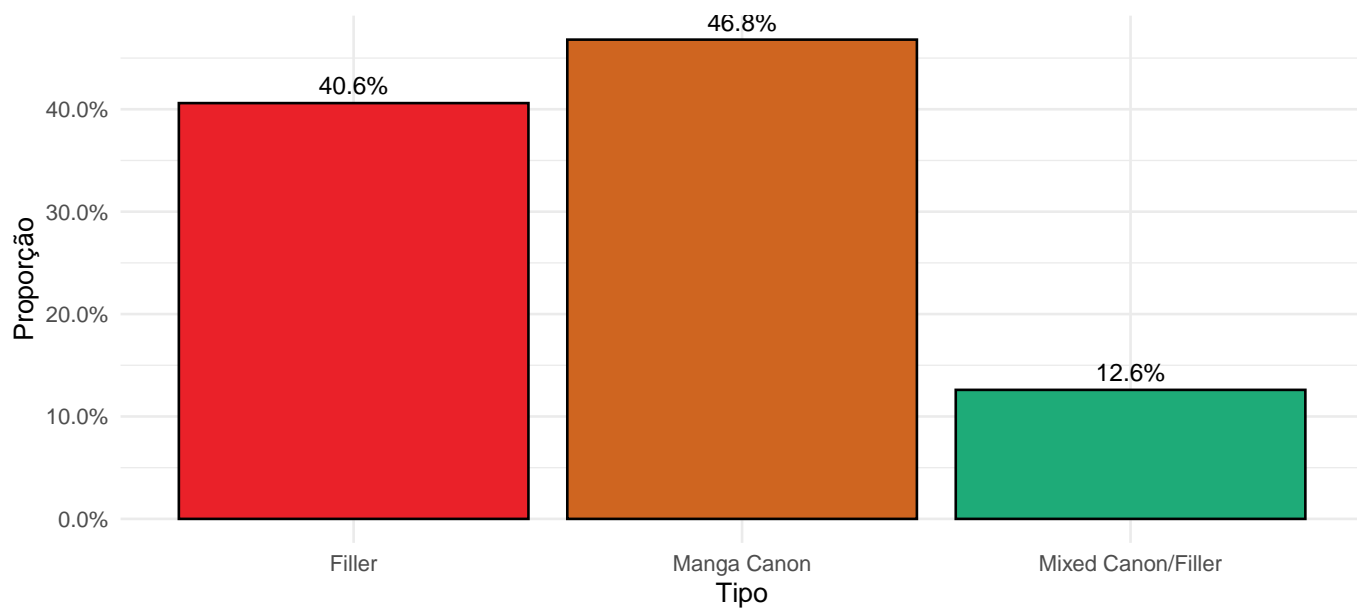
A história é dividida em duas partes, a primeira parte se passa nos anos da pré-adolescência de Naruto (clássico), e a segunda parte se passa em sua adolescência (Shippuden). Nesse trabalho trataremos sobre o Naruto Shippuden, desenvolvendo-o em uma perpestiva de séries temporais, vamos adotar como variável de interesse a avaliação dos episódios, assim veremos pontos importantes da saga ao longo do tempo.

O Banco de dados utilizado é totalmente original, foi criado fazendo *Web scraping* de dois sites diferentes, para aqueles interessados o banco foi disponibilizado aqui.

2 Informações

Como o anime é uma adaptação, existem episódios fiéis ao mangá (Manga Canon) e episódios originais do próprio anime, em outras palavras não seguem o material original, esses episódios são chamados de Fillers, existem também episódios que seguem a trama do mangá mas além disso possuem elementos novos, esses chamados de Mixed Canon.

Vejamos pelo gráfico a seguir a porcentagem do tipo dos episódios



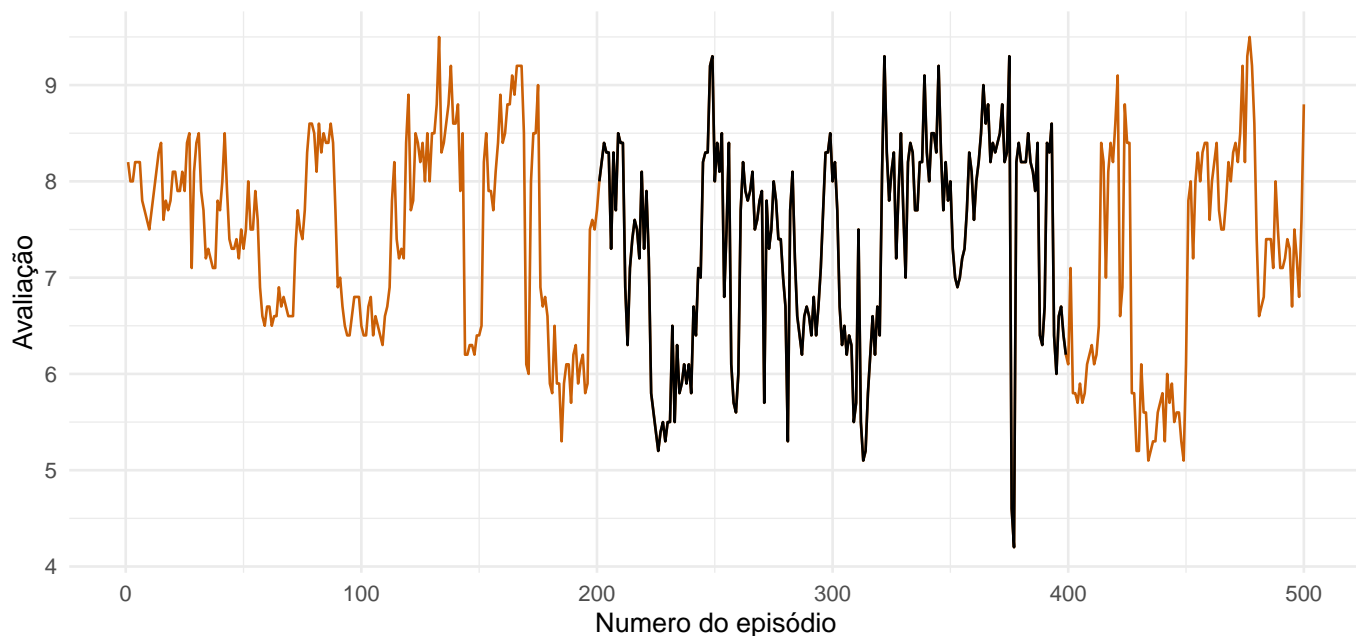
Fonte: Elaborado pelos autores

É um fato bastante curioso, Naruto Shippuden possui uma quantidade altíssima de episódios fillers, aproximadamente 41%, note pela tabela que são somente 31 episódios a mais canônicos.

Tabela 1: Valores Absolutos do Tipo de Episódio

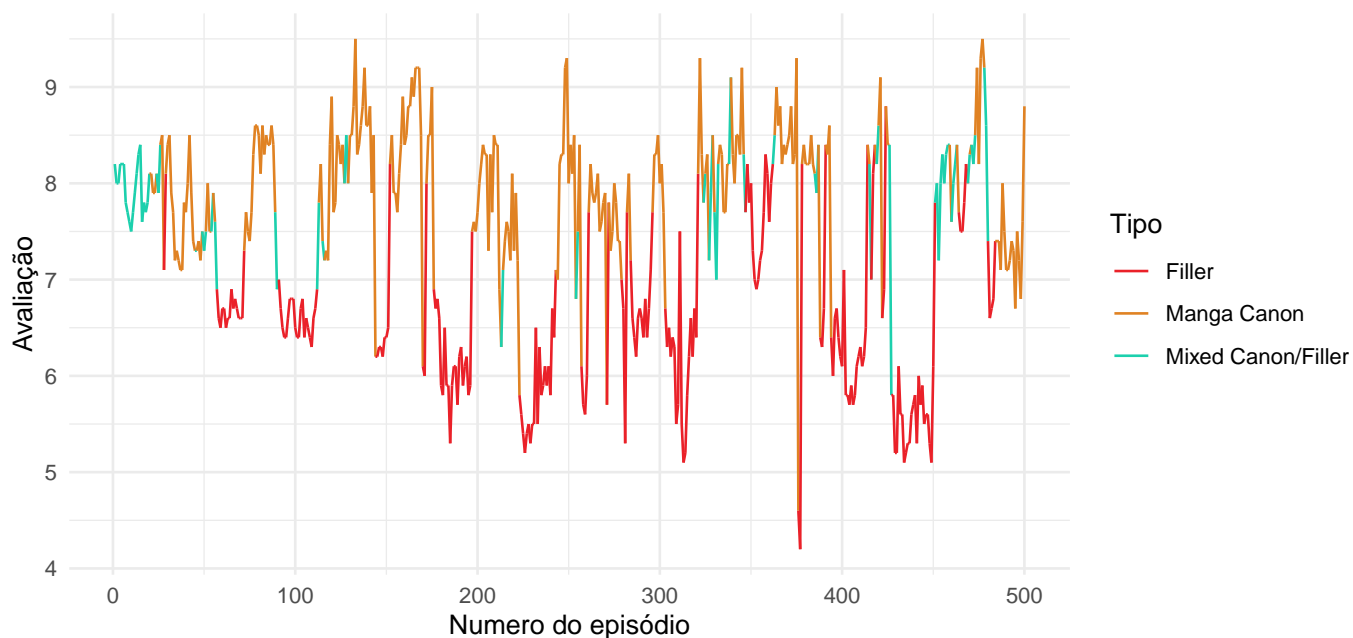
Tipo	Quantidade
Filler	203
Manga Canon	234
Mixed Canon/Filler	63

Em termos de avaliação ao longo do tempo, podemos utilizar como índice ordinal o próprio número do episódio já que que é uma função injetora no tempo.



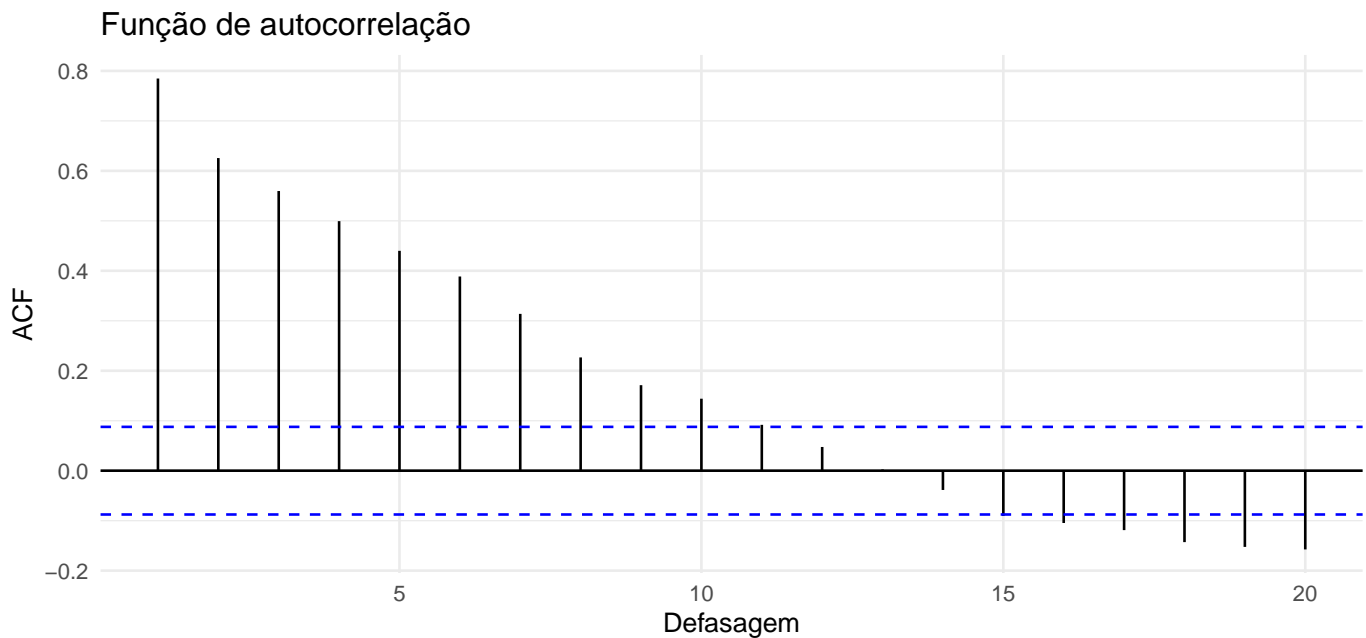
Há bastante coisa a se notar nesse gráfico, primeiramente pelos episódios 200 e 400, existe um indício de alteração na média das variáveis aleatórias, portanto furando o pressuposto de média constante ao longo do tempo, perto dos episódios finais da saga nota-se também inúmeras quedas bruscas na avaliação sem um contra-peso de avaliações com notas altas, outro indicio de não estacionariedade.

É importante tentarmos entender um pouco sobre esses episódios com notas baixas, assim vamos ver esse gráfico pelo tipo de episódio.



Assim fica fácil ver que os episódios com menor avaliação são em sua maior parte fillers, toda queda brusca de avaliação tem um episódio filler envolvido.

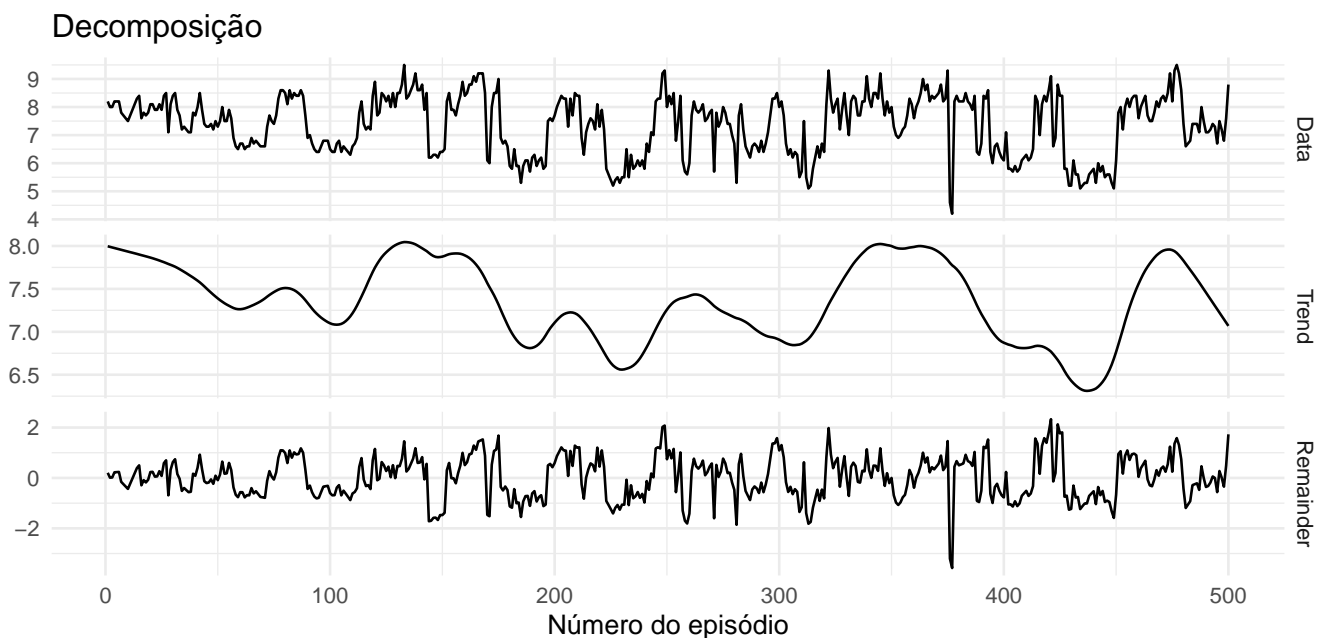
É necessário também avaliar a correlação nas avaliações:



Note que o gráfico anterior as autocorrelações tendem a ficar significativas até 10 defasagens, pois em média a duração dos arcos é 23.8, assim 10 defasagens significa em geral que:

- o episódio está tendo influência do final do arco anterior;
- o episódio está gozando consequências dos episódios iniciais do arco que está situado;
- o episódio está apanhando desfechos dos episódios pós metade do arco que está situado

Uma abordagem importante também é decompor a série temporal, apriori supõe-se a existência de três elementos, a saber: Tendência, sazonalidade e resíduo. Todos esses serão explorados cuidadosamente nas seções posteriores, assim vamos decompor a série para ter um vislumbre de tais elementos:



A questão de curiosidade vamos ver se os resíduos seguem uma distribuição normal, para isso utilizamos o teste de Shapiro Wilk que possui como hipótese:

$$H_0 : \text{Os dados seguem uma distribuição Normal}$$

Efetuando o teste obtemos um p-valor de 0.006, assim portanto rejeitando a hipótese H_0 , porém os resíduos não possuírem distribuição normal não afetará as análises daqui pra frente.

3 Testes

Os gráficos de seção anterior evidenciaram a possibilidade da série não ser estacionária, assim faz-se necessário verificar se vale de fato para o processo estocástico, com isso vamos precisar aplicar testes de hipóteses para averiguar algumas propriedades, como existência de tendência e sazonalidade.

3.1 Tendência

Tendência refere-se a um algum comportamento não - estocástico da série em algum momento do tempo, se tal comportamento só acontece em alguns momentos específicos do tempo, chamamos de tendência estocástica, do contrário é dita determinística. Vamos começar pelos testes de tendência determinística, que em termos sumarizados possuem como hipótese:

$$H_0 : \text{A série não possui tendência determinística}$$

Tabela 2: Testes de Tendência Determinística

Testes	P-valor
Cox-Stuart	0.050
Wald-Wolfowitz	0.000
Mann-Kendall	0.039

Conforme mostrado pela Tabela anterior podemos ver que pelos testes de Cox-Stuart, Wald-Wolfowitz, Mann-Kendall, temos que todos os p-valores são menores pequenos, então concluímos que existe tendência determinística na série temporal.

3.2 Sazonalidade

Sazonalidade acontece quando a série possui um comportamento que se repete frequencialmente, vamos nessa subseção testar se existe sazonalidade nas avaliação dos episódios, assim temos como hipótese:

$$H_0 : \text{A série não possui sazonalidade}$$

Os testes aplicados foram Kruskal-Wallis, Friedman e Autocorrelação em lags Sazonais, assim gerando a seguinte tabela:

Tabela 3: Testes de Sazonalidade

Testes	P-valor
Kruskal-Wallis	0.745
Friedman	0.878
Autocorrelação em lags Sazonais	0.701

Assim, como os testes obtiveram p-valor > 0.5, concluímos portanto que não existem indícios de sazonalidade.

3.3 Raiz unitária

Para começar tal seção, primeiros vamos definir um passeio aleatório em sua forma simplificada:

$$Y_t = Y_{t-1} + \epsilon_t$$

Onde t refere-se aos índices de ordenação, e ϵ um termo aleatório, para facilidade vamos assumir que $E(\epsilon_t) = \mu$ e $var(\epsilon_t) = \sigma^2 \forall t$, assim subtraindo-se Y_{t-1} em ambos os lados tem-se

$$Y_t - Y_{t-1} = \epsilon_t$$

que é um processo estacionário, o processo anterior é dito processo estacionário em diferença, em outros casos também chamado de processo integrado. O exemplo anterior trata-se de um caso mais geral

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

onde evidentemente $\phi = 1$, é fácil mostrar que se $|\phi| < 1$ tem-se um processo estocástico, aqui portanto, estamos interessados nesse caso, para isso utilizaremos testes de hipótese.

Fala-se também que o caso anterior na forma simplificada possui raiz unitária, por causa do operador Lag, aqui não definido, porém pode-se ler sobre em [1]

Tabela 4: Testes de Tendência Estocástica

Testes	P-valor
Kwiatkowski-Phillips-Schmidt-Shin (KPSS)	0.1
Augmented Dickey-Fuller (ADF)	0.01*
Phillips-Perron (PP)	0.01*

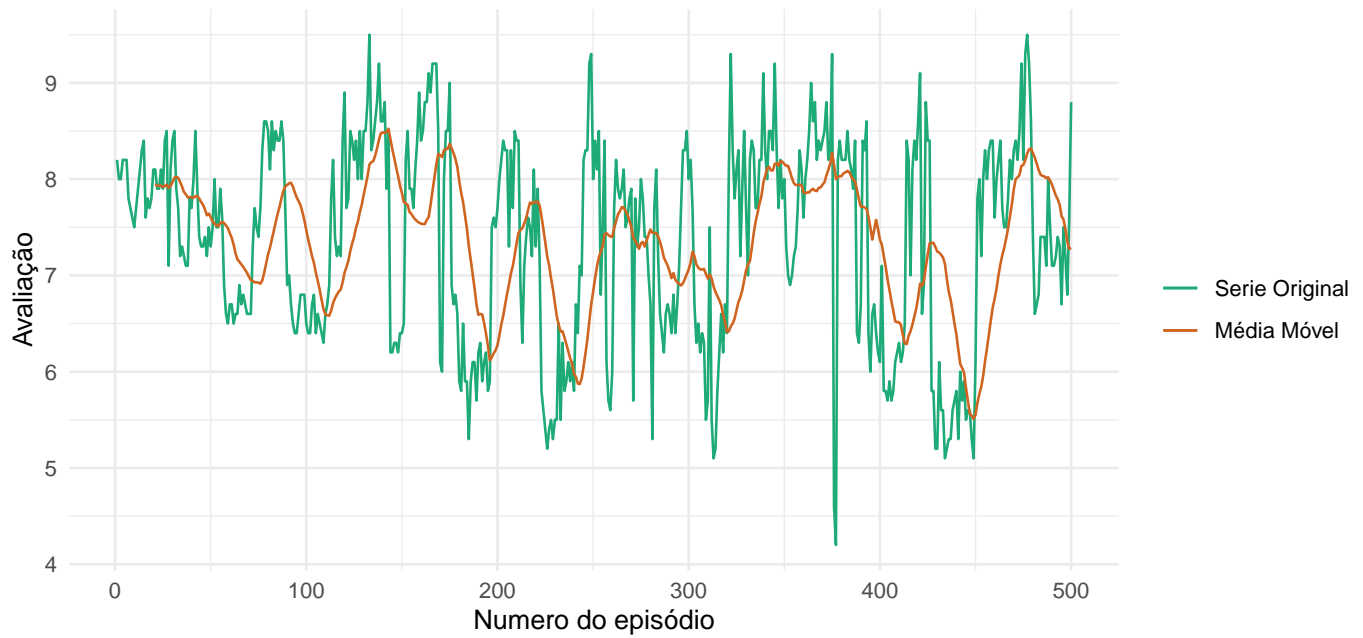
Conforme mostrado pela Tabela anterior podemos ver que pelos testes de Dickey-Fuller Aumentado, Phillips-Perron, KPSS, com índice de significância de 5%, temos que os p-valores dos testes ADF e PP são menores que 0.05 e do teste KPSS é maior que 0.05, então concluímos que não existe tendência estocástica na série temporal. Pois, nos testes ADF e PP, a hipótese nula é a existência de raiz unitária e no teste de KPSS, a hipótese nula é não existência de raiz unitária.

3.4 Resultados

Vimos visualmente que a série a partir de um certo começa ter uma descrescimento, assim portanto dando um vislumbre de tendência determinística, aplicando os testes obtemos mais uma evidência de existência de tendência na série. Em relação a sazonalidade os testes também confirmaram o que os gráficos mostraram: não existe evidência de sazonalidade na série.

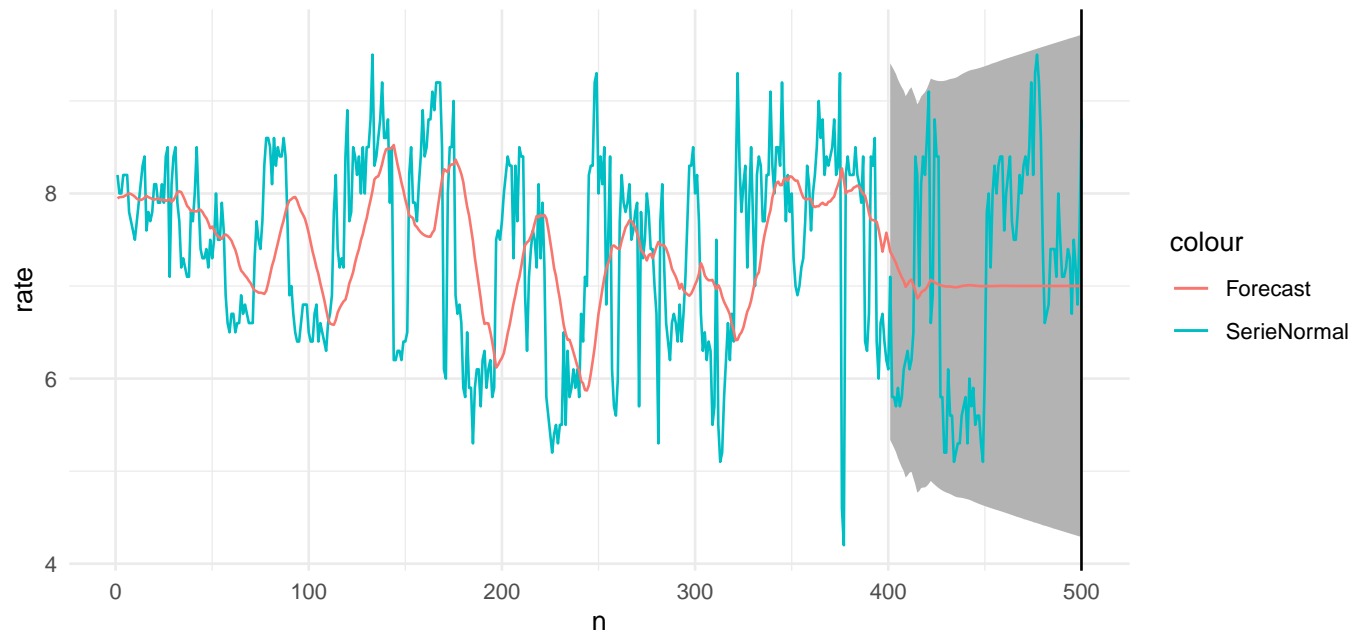
4 Modelagem

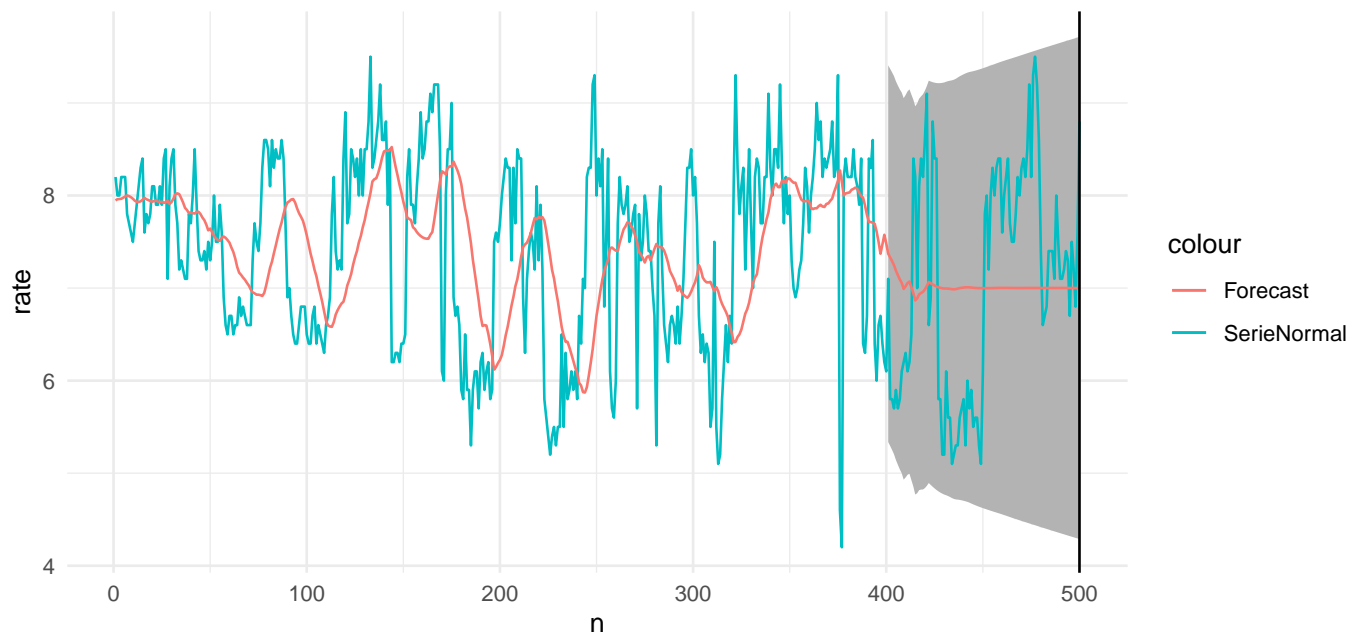
4.1 Médias Móveis Simples



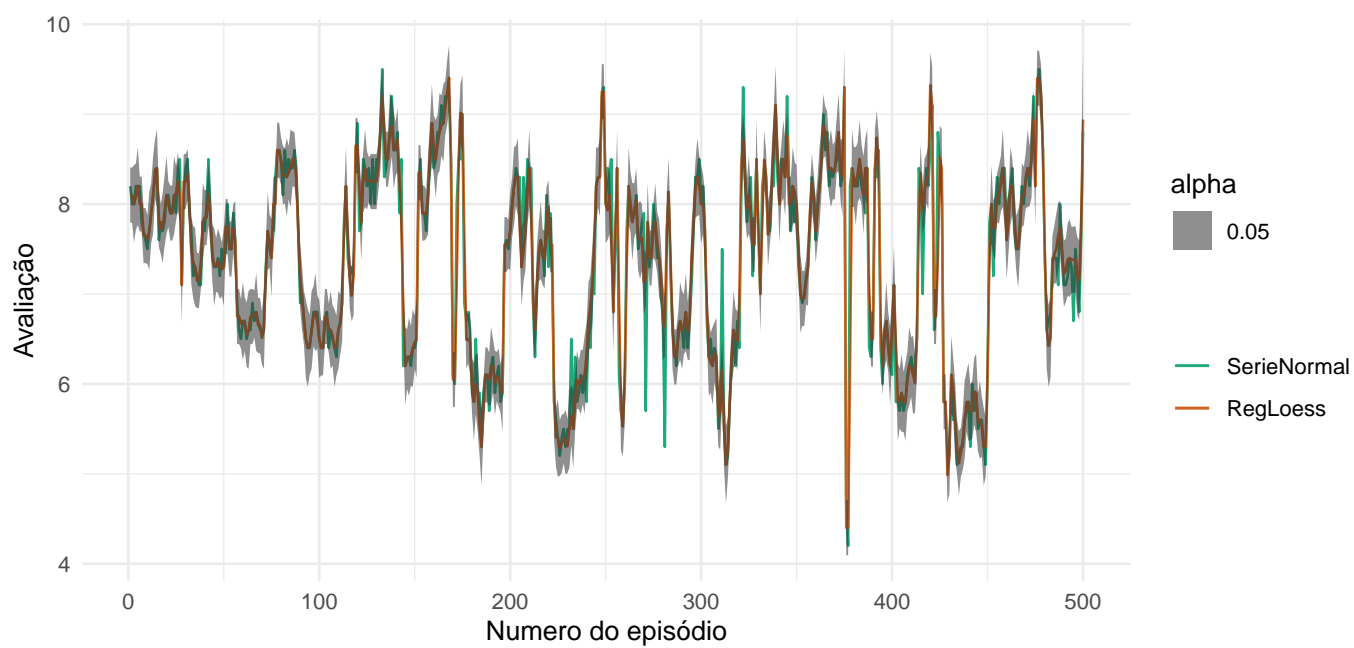
4.1.1 Previsão

Utilizando o pacote smooth.





4.2 Regressão LOESS



4.3 Suavização Exponencial

4.3.1 Com outros pacotes

```
## Holt-Winters exponential smoothing with trend and multiplicative seasonal component.
##
## Call:
## HoltWinters(x = treino, seasonal = "multiplicative")
##
```



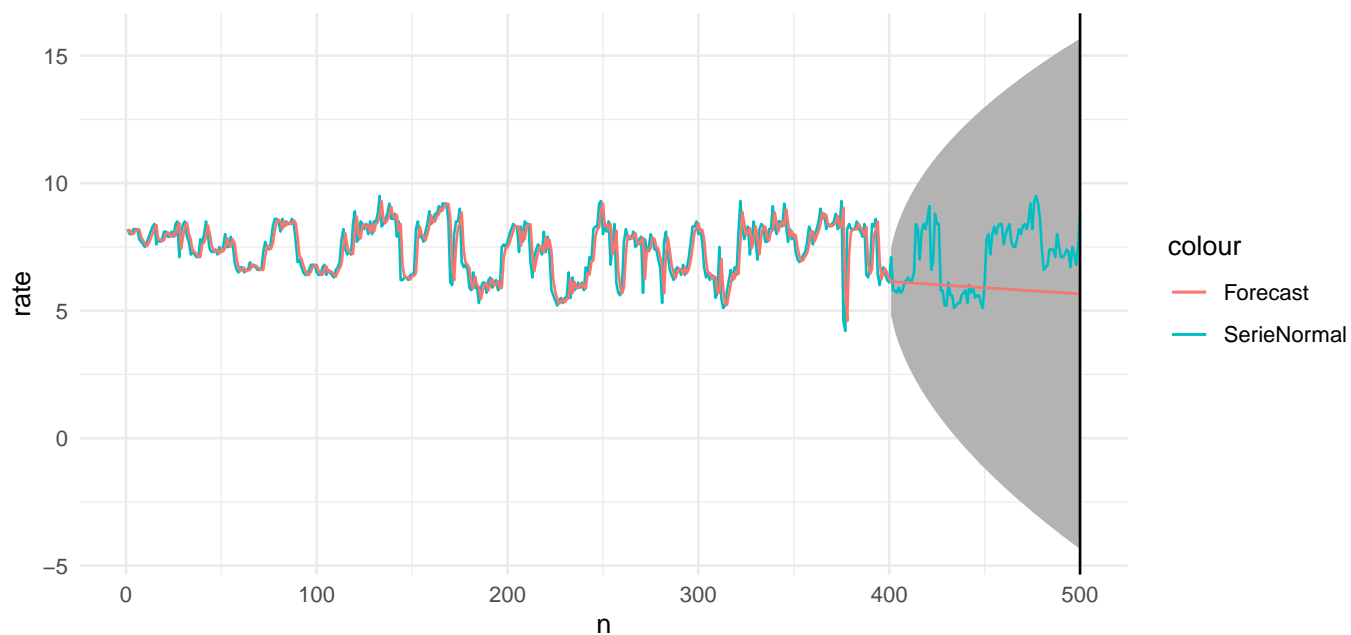
```

## Smoothing parameters:
## alpha: 0.747
## beta : 0
## gamma: 0.515
##
## Coefficients:
##      [,1]
## a      6.23957
## b     -0.00792
## s1     0.97274
## s2     0.94760
## s3     0.97191
## s4     0.97599
## s5     1.04107
## s6     1.06997
## s7     1.06873
## s8     1.03901
## s9     1.04321
## s10    1.00204
## s11    1.01854
## s12    1.01820
## s13    0.98331
## s14    0.98048
## s15    1.04988
## s16    1.03390
## s17    1.05076
## s18    1.05857
## s19    1.06820
## s20    1.04287
## s21    0.99206
## s22    0.95780
## s23    0.92976
## s24    0.92432
## s25    0.93824
## s26    0.92864
## s27    0.87862
## s28    0.79702
## s29    0.86606
## s30    0.93552
## s31    0.88901
## s32    0.92278
## s33    0.91405
## s34    0.89616
## s35    0.90375
## s36    0.93849
## s37    0.95931
## s38    1.03572
## s39    1.03496
## s40    1.00495
## s41    1.05577
## s42    1.15630
## s43    1.17270
## s44    1.05047
## s45    1.03678
## s46    1.01539

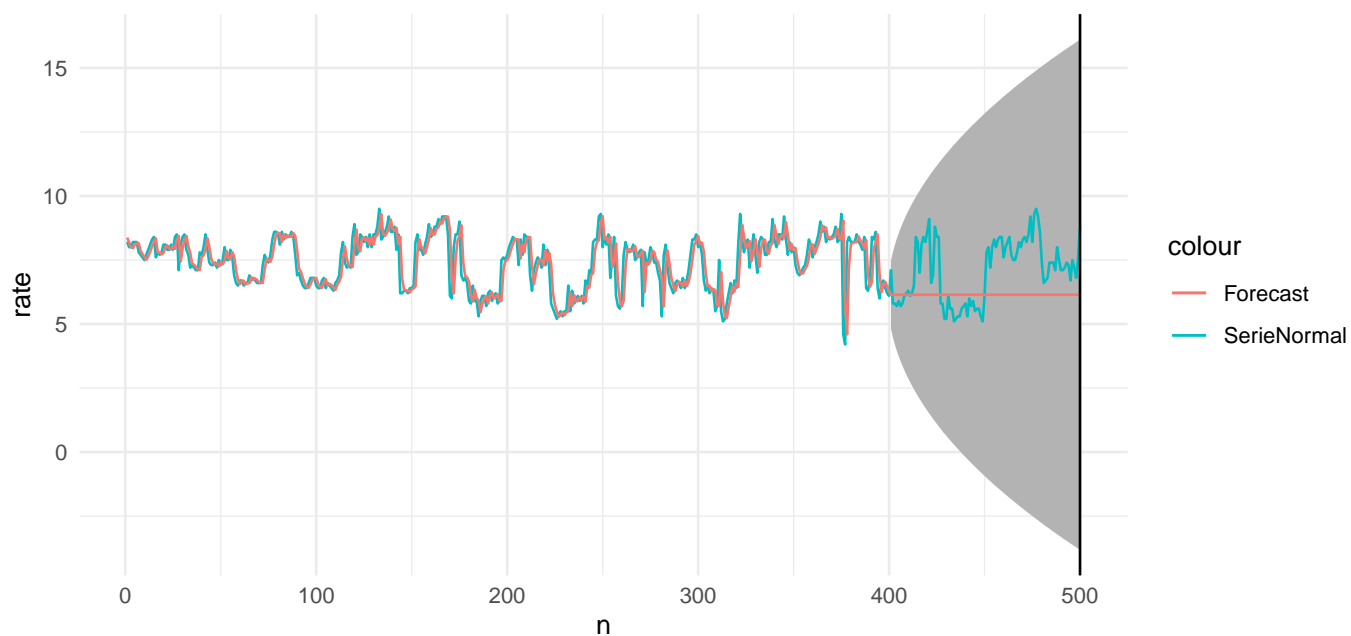
```

```
## s47 1.05935
## s48 1.07748
## s49 1.06448
## s50 1.00373
## s51 0.98643
## s52 0.97885
```

4.3.1.1 Holt não Amortecido



4.3.1.2 Holt com Amortecimento



4.4 Calculando Métricas de Acurácia dos Modelos não-ModelTime

Tabela 5: Métricas nos dados de teste

	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
SMA	1.08	15.9	2.36	15.5	1.23	0.060
Holt	1.44	18.6	3.12	21.0	1.74	0.154
Holt amortecido	1.28	16.7	2.78	18.4	1.53	0.062

[1] 23.8

Referências

- [1] Morettin PA, Toloi CM. Análise de séries temporais: Modelos lineares univariados. Editora Blucher; 2018.