3° Avaliação

Disciplina: Séries Temporais

Professor: Vinícius Lima

Aluno: Júlio Henri Maciel

Matrícula: 201807840011

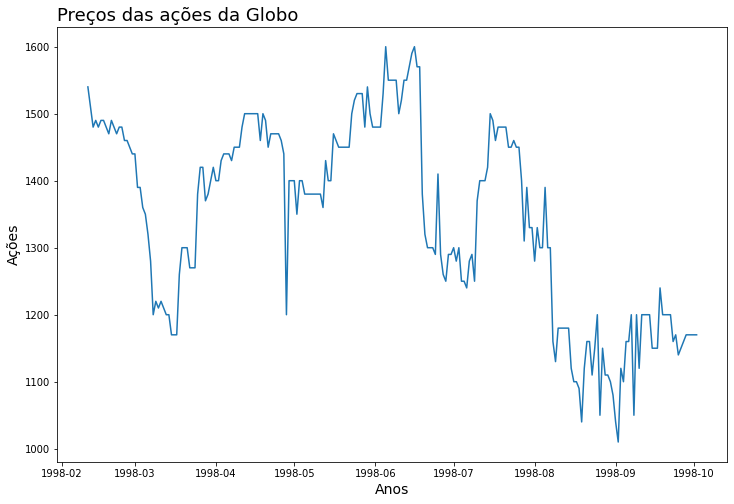
**Relatório Técnico**

1. **Pré-processamento**

Para este estudo foi escolhido um banco de dados com dois atributos e 234 registros. Um atributo se chama “Time” e contém as datas enquanto a segunda coluna apresenta os valores da ação “D-GLOBO”. Todos os valores nulos foram removidos anteriormente para tornar o pré-processamento mais simples e vale ressaltar que as observações do tempo vão de 11 de fevereiro de 1998 à 02 de outubro de 1998.

Para fazer a modelagem dos dados posteriormente foi necessário transformar a data em *DatetimeIndex* para que então os gráficos e estudos sejam realizados com facilidade. Dito isto, abaixo é exposto a Figura 01 onde mostra a série temporal.

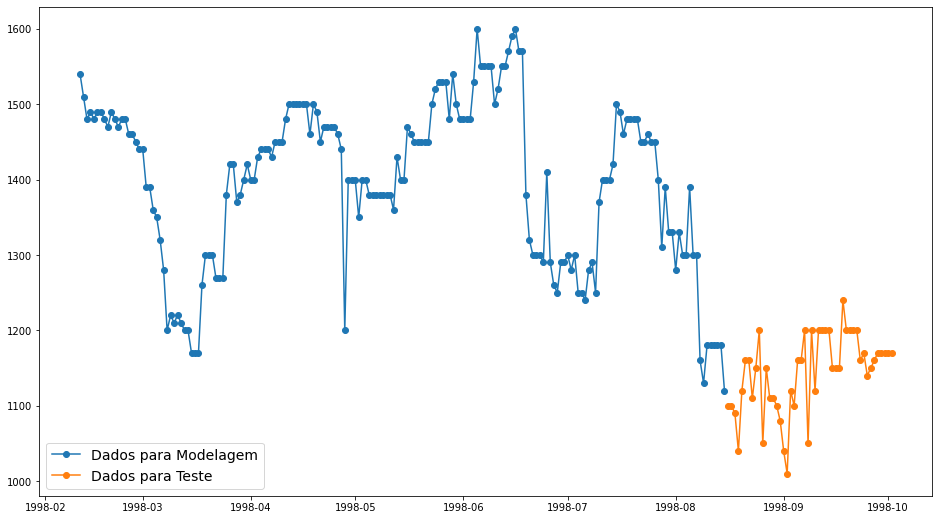
**Figura 01:** Preço das ações da Globo



Como pode-se observar, pode haver sazonalidade nos dados e tendência de decrescimento aparentemente. Para tanto, será de suma importância testes formais para considerar a série estacionária ou não.

No pré-processamento também foi efetuado a separação dos dados em dados de modelagem e dados de teste, o que foi decidido que de 11 de fevereiro de 1998 à 15 de agosto de 1998 seriam os dados para modelar e de 16 de agosto até o fim dos dados (02 de outubro de 1998) seriam para testar o modelo, como mostrado na Figura 02 a seguir.

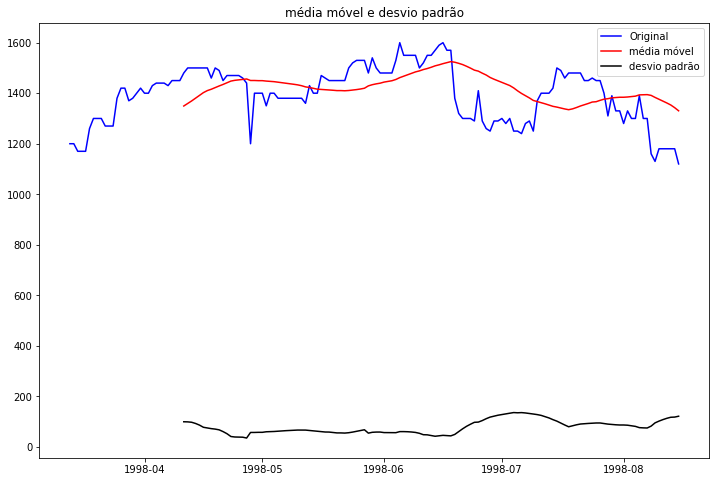
**Figura 02:** Dados de modelo X dados de teste



1. **Identificação de Modelos Candidatos**

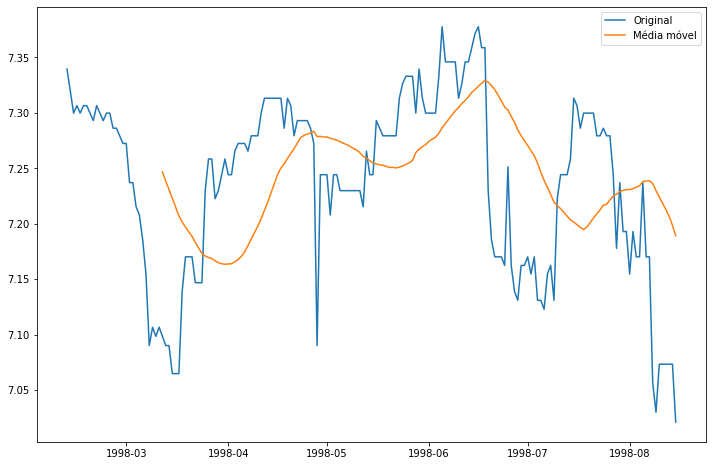
Primeiramente foi feito o gráfico da média móvel dos dados e o desvio padrão para se ter uma noção visual dessas estatísticas com o intuito de saber se os dados de modelagem são estacionárias ou não. O cálculo é feito através da soma do número de ações do mês e dividindo esse resultado por 30 (que são os dias totais de um mês), logo, essa média só começa a ser contabilizada a partir do segundo mês, dado que não se tem registros anteriores ao primeiro mês da série temporal.

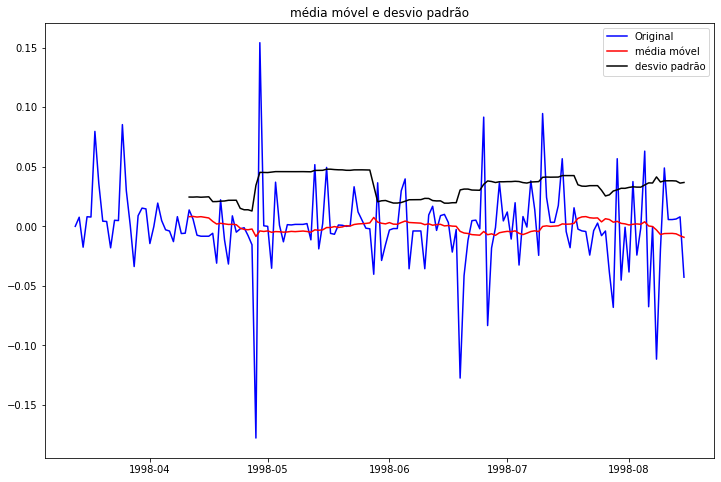
**Figura 03:** Estatísticas dos dados de Modelagem.



Após fazer o Teste de Dickey-Fuller Aumentado o resultado do p-valor foi 0.286, logo, há evidências o suficiente para dizer que a série não é estacionária, para tanto, foi optado por fazer o log da séria temporal e duas diferenciações no intuito de que a série passe a ser estacionária, abaixo é exposto os gráficos com os tratamentos citados acima.

**Figura 04:** Dados da série temporal após serem obtidos os logaritmos



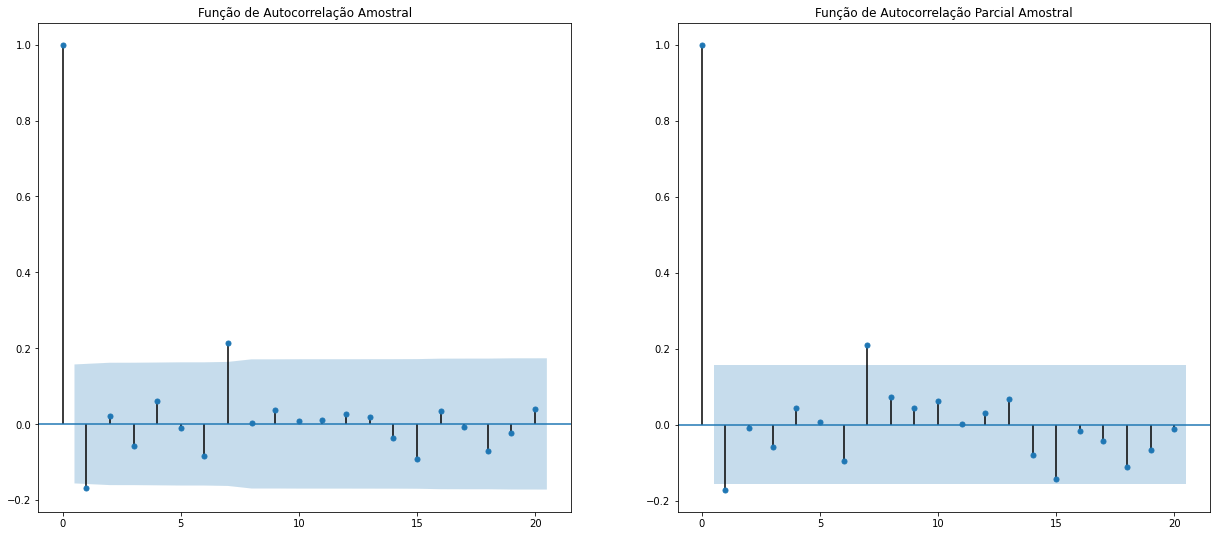
**Figura 05**: Gráfico da série após duas diferenciações 

Após tais tratamentos, aparentemente, os dados são estacionários dado que a série, o desvio padrão e a média móvel estão em torno de zero, e com o resultado do p-valor de 3.6885e-27, no Teste de Dickey Fuller Aumentado, pode-se concluir que a série temporal passou a ser estacionária.

Antes de realizar a modelagem dos dados, é necessário obter três informações de supra importância para o modelo de séries temporais. Estes são: P, é o parâmetro associado ao aspecto auto-regressivo do modelo, isto é, relacionado aos valores passados do evento em questão. D, é a quantidade de diferenciações feitas na série temporal, no caso, 2. E por fim, Q, se refere a média móvel do modelo.

Como dito anteriormente, já se tem a informação do valor de D, que é igual a 2, já os termos P e Q podem ser obtidos através da análise dos gráficos de Função de Autocorrelação Amostral e Função de Autocorrelação Parcial Amostral.

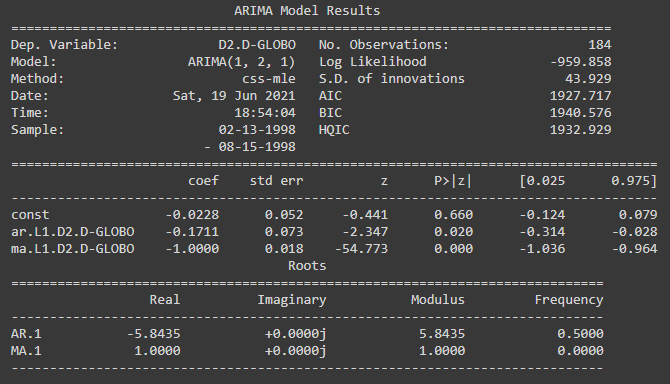
**Figura 06:** Função de Autocorrelação Amostral e Função de Autocorrelação Parcial Amostral.



Há um decréscimo rápido tanto na função de autocorrelação quanto na função de autocorrelação parcial e se torna zero a partir do lag 2, com o lag 7 voltando a ser diferente de 0 nos dois gráficos. Isto é um forte indicativo de que poderia ser tanto um modelo ARMA(1,1) (onde o primeiro termo é P e o segundo é Q), quanto um MA(1), que seria um modelo mais simples.

1. **Estimação de Parâmetros**

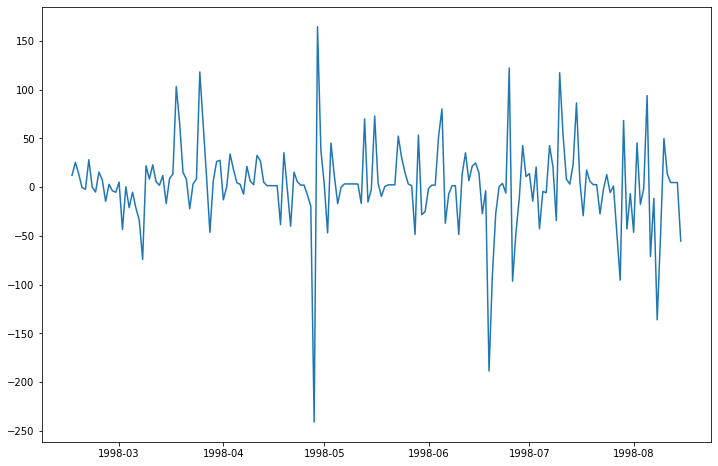
**Figura 07:** Sumário do modelo ARIMA(1,2,1)



Analisando a figura acima, pode-se constatar que o coeficiente MA no lag 1 deu igual a -0.99, enquanto o coeficiente do AR no lag 1 foi igual -0.171. Então a fórmula fica da seguinte maneira:

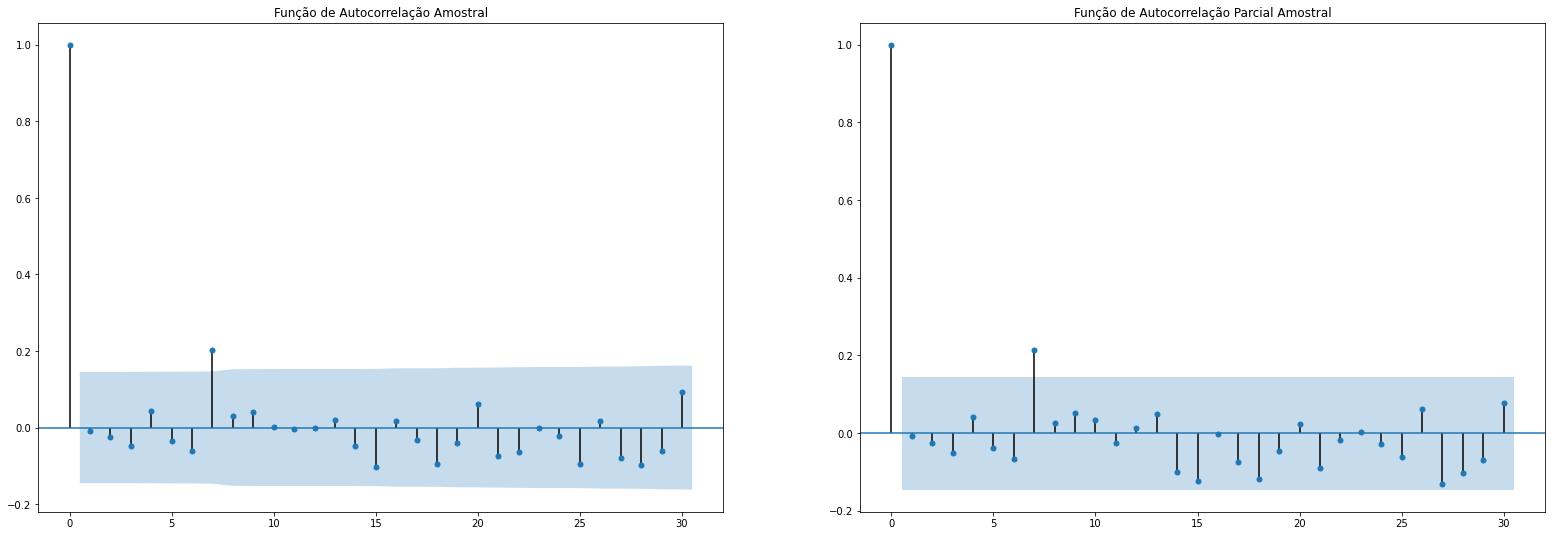
1. **Diagnóstico do Modelo**

**Figura 08:** Resíduos do modelo



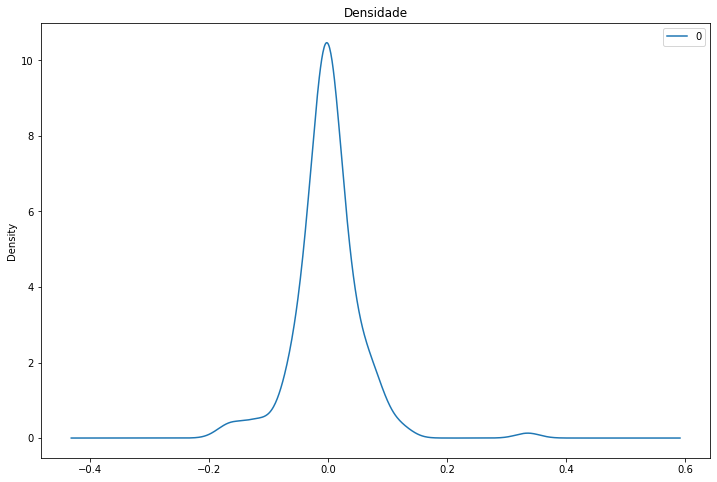
O gráfico dos resíduos parece estar de acordo com os pressupostos, porém no mês de maio parece ter grande variação desses resíduos, o que pode ser um problema para o modelo, então, abaixo foi apresentado o gráfico de Função de Autocorrelação e Autocorrelação Amostral para análise.

Através da análise da Figura 10, pode-se concluir que há um rápido e grande decréscimo nos dois gráficos e a partir do lag 1 todos os valores estão dentro do intervalo de confiança para a função de autocorrelação parcial (com exceção do lag 7), ou seja, são iguais a zero. A função de autocorrelação parcial se comporta do mesmo jeito que a função de autocorrelação.

**Figura 9:** Função de Autocorrelação e Função de Autocorrelação Parcial

Se faz necessário também a análise gráfica da densidade dos resíduos, então foi o feito o histograma com o objetivo de visualizar se os resíduos seguem uma distribuição normal.

**Figura 10:** Densidade dos resíduos do modelo



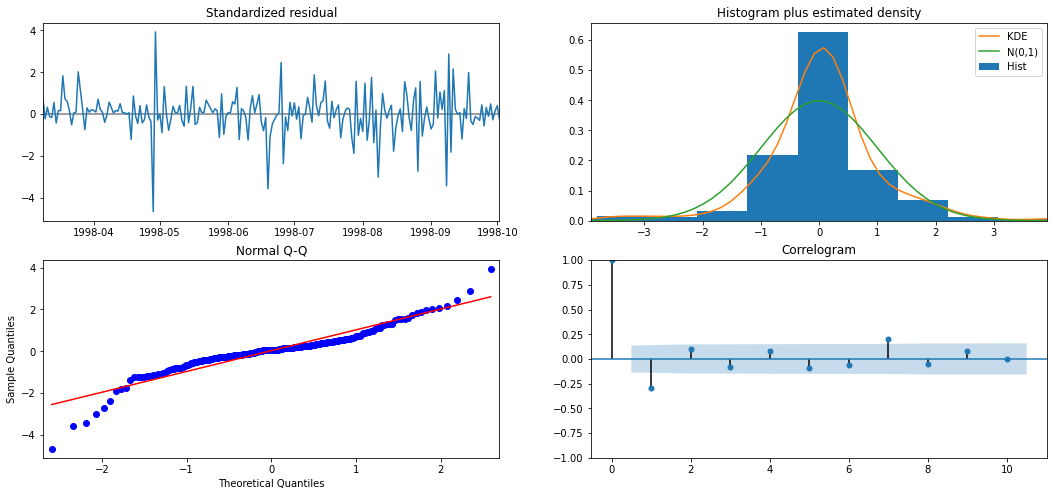
Como pode-se observar, aparentemente, os resíduos seguem uma distribuição normal com média igual a zero.

1. **Teste de Modelos**

Efetuando a comparação de diversos modelos SARIMAX para saber qual é o mais adequado, foi concluído que o melhor modelo para o estudo é o SARIMAX (0,1,0) x (1,1,1,12) a partir do valor AIC de cada opção de modelo.

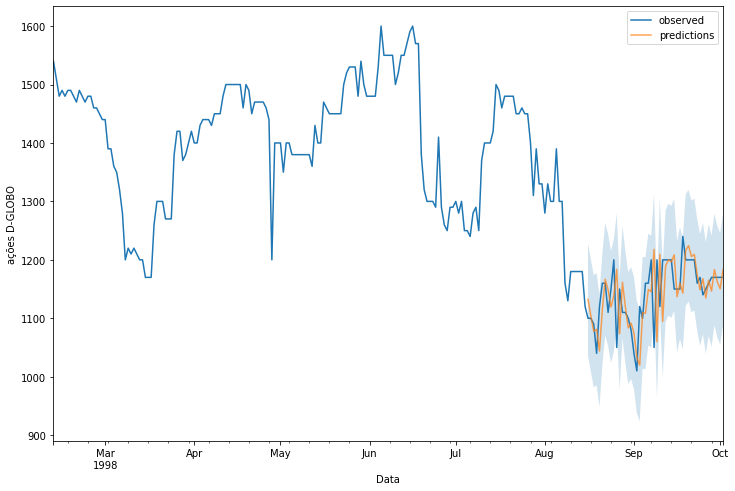
O sumário do novo modelo resultou no coeficiente de AR igual a -0.634, coeficiente MA igual a -1 e o sigma sendo igual a 2251. Em relação a Figura 12, que mostra gráficos importantes para análise e verificação do modelo perante os pressupostos da validação de um modelo de séries temporais, pode-se constatar que os resíduos e a série em si estão de acordo.

**Figura 11:** Diagnóstico do novo modelo



Então, a predição dos dados foi efetuada baseado nos dados de modelagem/treino e será testado e avaliado nos dados de teste, para assim se ter noção do quão bom o modelo é para ajustar os dados.

**Figura 12:** Valores preditos X valores reais



Visualmente, pode-se dizer que os valores previstos aparentam estar próximos aos valores reais, para calcular a acurácia do modelo, foi escolhido o método raiz do erro quadrático médio (RMSE, em inglês) pois o resultado da acurácia está na mesma dimensão da variável analisada, no caso, as ações da Globo. O Cálculo se dá da seguinte maneira:

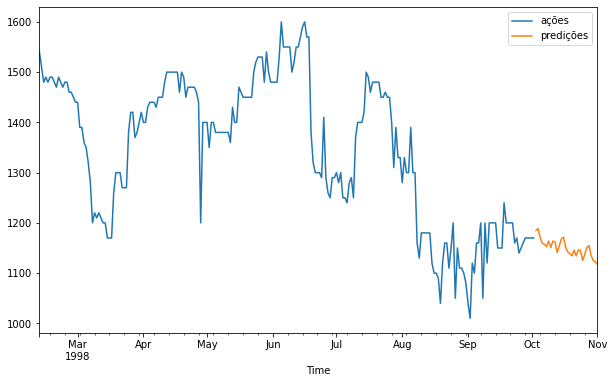


O resultado do RMSE foi de 6.2056, o que é uma acurácia relativamente boa para o estudo em questão.

1. **Previsão**

Por fim, realizou-se a previsão das ações da Globo para o próximo mês, isto é, baseado no treino do modelo nos dados e no teste feito, espera-se a raiz do erro quadrático médio em torno de 6.2056 para as predições futuras.

**Figura 13:** Predições das ações da Globo para o mês de outubro.

****