



Universidad de Sonora
División de Ciencias Exactas y Naturales
Departamento de Física
Física computacional
Pronóstico de Series de Tiempo
Actividad 6
Iveth R. Navarro

23 de febrero de 2021
Puerto Peñasco, Sonora, México

Resumen

La presente es una práctica de Física Computacional cuyo objetivo es el *Pronóstico de Series de Tiempo* siendo la información meteorológica de CONAGUA la estudiada, estación Puerto Peñasco, Sonora.

1. Introducción

Esta actividad tiene por propósito realizar pronósticos de series de tiempo, para lo que se requiere contar con un modelo para predecir valores futuros en base a los valores históricos pasados, como es el caso de la base de datos con la que hemos estado trabajando todo el parcial.

Algunas de las bibliotecas a utilizar son las siguientes: Utilizaremos las bibliotecas.

- `import statsmodels.api as sm`
- `from statsmodels.tsa.stattools import adfuller`
- `from statsmodels.tsa.arima_model import ARIMA`
- `from statsmodels.tsa.arima_model import ARIMAResults`
- `from statsmodels.tsa.stattools import acf, pacf`

A continuación se enlista los requisitos que requiere cumplir una serie de tiempo para crear un modelo de pronóstico.

1. Estacionalidad

Este concepto es importante en la teoría de Análisis de Series de Tiempo. Para aplicar un modelo, se requiere determinar si la serie de tiempo es estacionaria, es decir que sus propiedades estadísticas (promedio, varianza, covarianza) permanezcan constantes en el tiempo. Así podemos prever que las propiedades de la serie de tiempo se van a comportar de la misma forma en el futuro.

En Estadística existe la Prueba Aumentada de Dicker-Fuller (ADF), para rechazar o no la hipótesis nula de que una serie de tiempo es no-estacionaria. La biblioteca Statsmodels contiene una función para la prueba ADF,

2. Pronóstico de Series de Tiempo.

Una vez que se tiene una serie de tiempo estacionaria, y conocemos sus componentes, se construyen modelos de pronóstico a los cuales les podremos agregar posteriormente las tendencias y estacionalidad que poseían.

Usaremos métodos de modelación estadística llamado ARIMA (Auto Regressive Integrated Moving Average) para realizar pronósticos.

El método ARIMA es similar a una ecuación de regresión lineal donde la predicción depende de los parámetros (p,d,q) del modelo ARIMA.

- p: es el número de términos AR (Auto-Regresivos). Por ej. si p=3, el predictor de $y(t)$ serán los términos $y(t-1)$, $y(t-2)$, $y(t-3)$.

- q: es el orden de términos del promedio móvil MA (Moving-Average). Por ej. si q=3, los términos para estimar el predictor de $y(t)$ serán los términos $y(t-1)$, $y(t-2)$, $y(t-3)$.
- d: es el número de diferenciaciones para hacer la serie estacionaria.

Para determinar los valores de p y q, se utilizan métodos gráficos y son:

- La Función de Autocorrelación (ACF), la cual compara la correlación entre dos valores consecutivos pero con un retraso (lag).
- La Función de Autocorrelación Parcial (PACF), que mide el grado de asociación del valor $y(t)$ y $y(t-p)$.

Los valores de p y q son enteros, y se obtienen a partir de las gráficas de ACF y PACF. Cuando estas se crucen por primera vez con la línea de $+1.96 \sigma$, donde σ es la desviación estándar. El cruce de ACF nos da información sobre el valor de q y el cruce de PACF nos da la información del valor de p.

Se estudia cada caso: de autoregresión AR y el de promedios móviles MA y después los integramos en el método de ARIMA.

Utilizamos el Criterio de Información de Akaike (AIC), como indicador de cuál es el mejor modelo. Aquel modelo con el menor AIC será el mejor modelo.

2. Desarrollo

Esta actividad se puede dividir en dos partes:

1. Actividad 6.1

Tomar una serie de datos en un periodo sin huecos (Tmax, Tmin) de la estación que se está analizando, para probar si es una serie de tiempo estacionaria o hacer las transformaciones para convertirla en una. Ver el ejemplo mostrado en la Jupyter Notebook de esta actividad para realizar las acciones similares.

2. Actividad 6.2

Construir los modelos ARIMA para cada Temperatura (Tmax y Tmin), y contrastar la bondad del ajuste en cada caso, calculando el error de la raíz cuadrada promedio RMSE y sobreponiendo gráficamente ambas series, la observada y la construida con el método ARIMA.

2.1. Resultados

2.1.1. Actividad 6.1

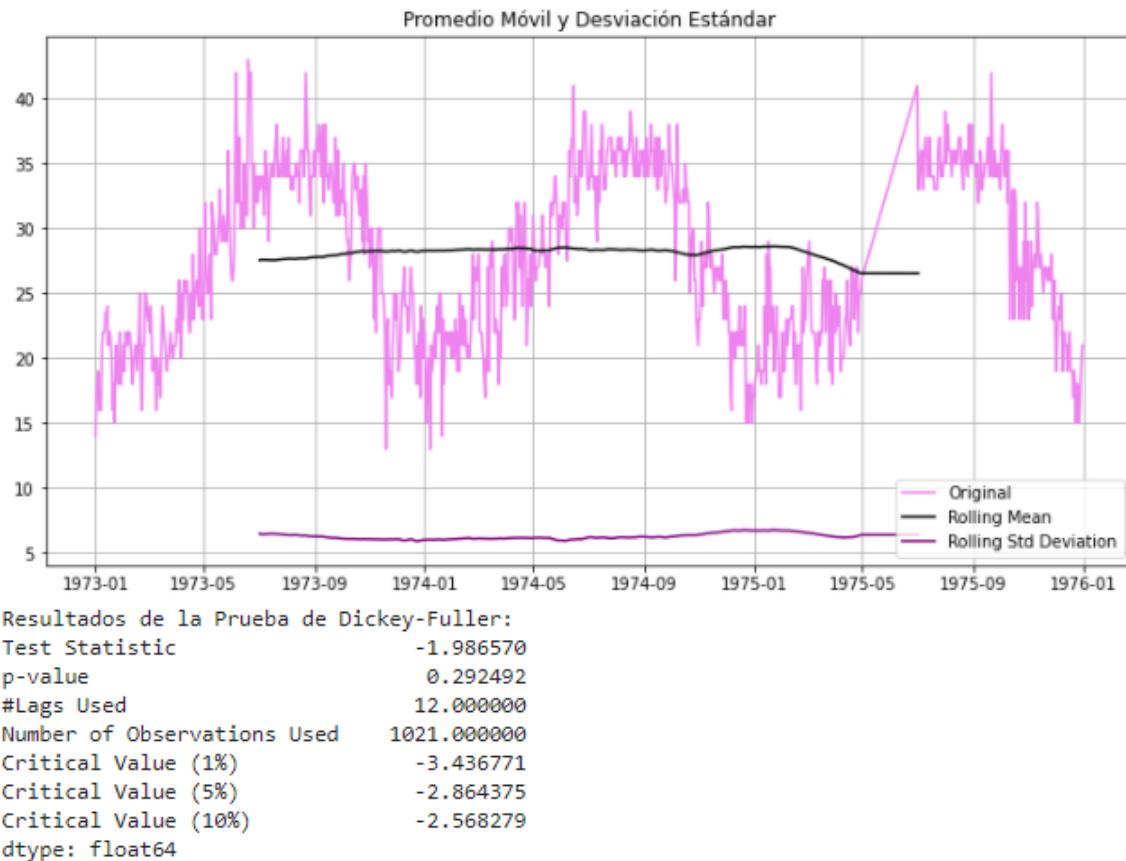


Figura 1

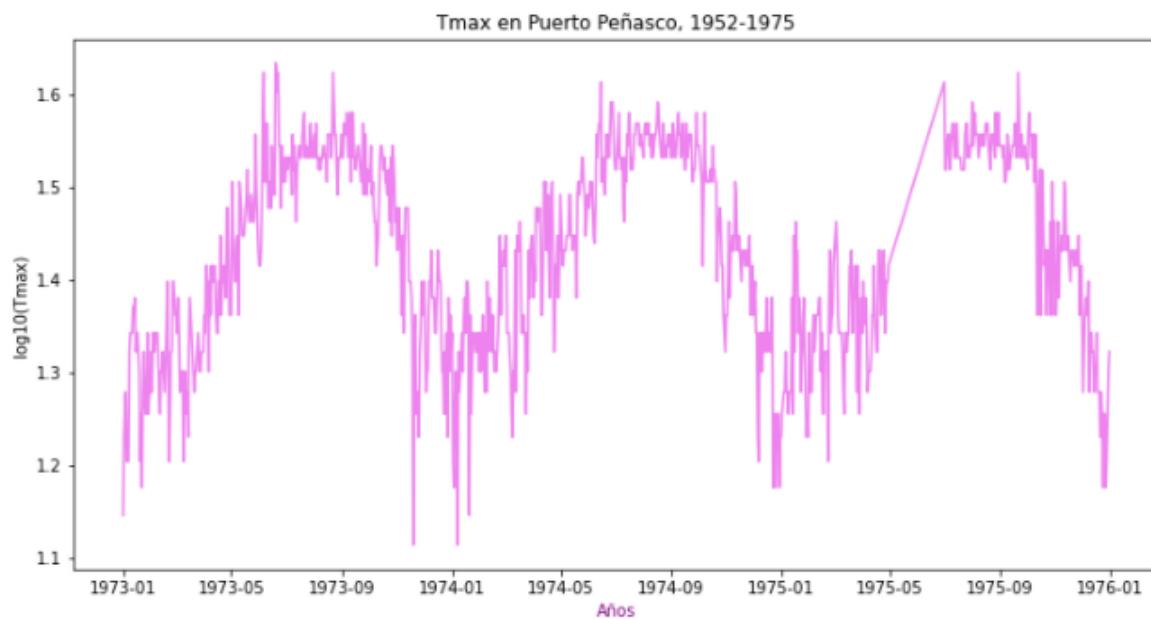
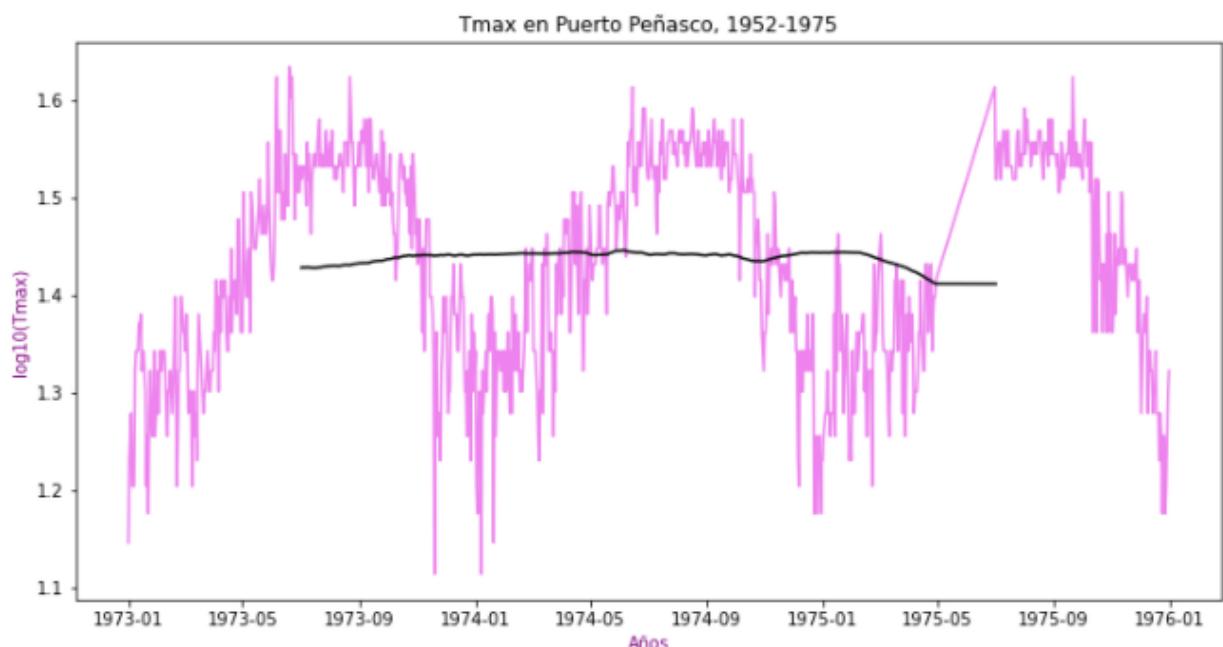
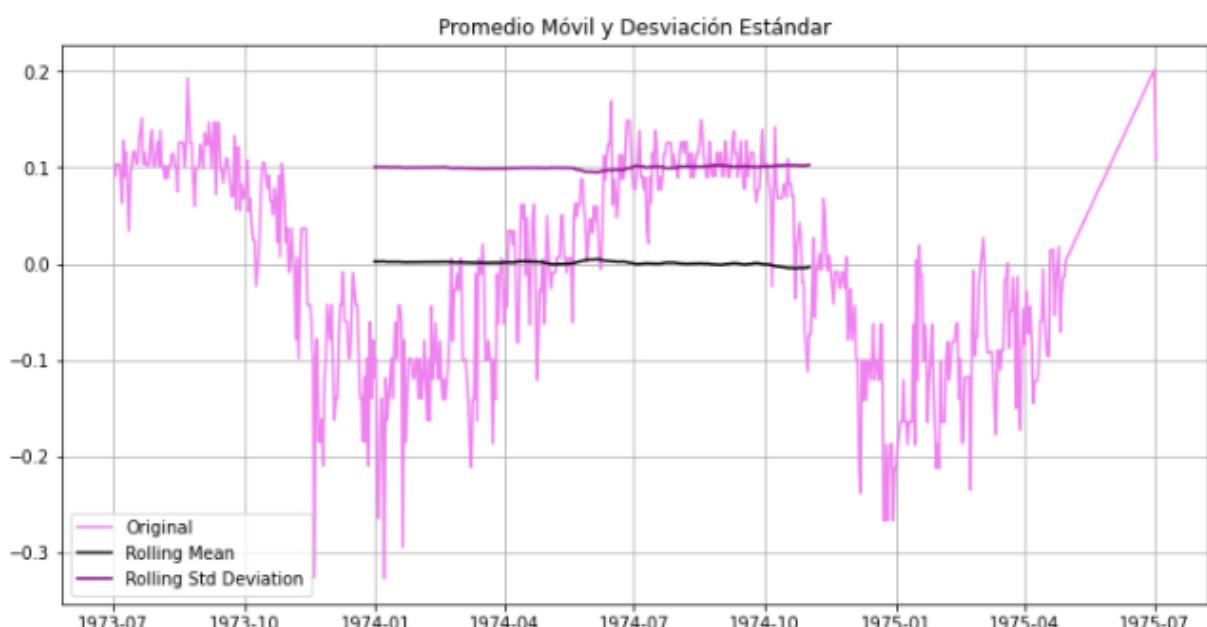


Figura 2

**Figura 3****Resultados de la Prueba de Dickey-Fuller:**

```

Test Statistic          -1.876319
p-value                0.343282
#Lags Used            19.000000
Number of Observations Used 650.000000
Critical Value (1%)    -3.440450
Critical Value (5%)    -2.865997
Critical Value (10%)   -2.569143
dtype: float64

```

Figura 4

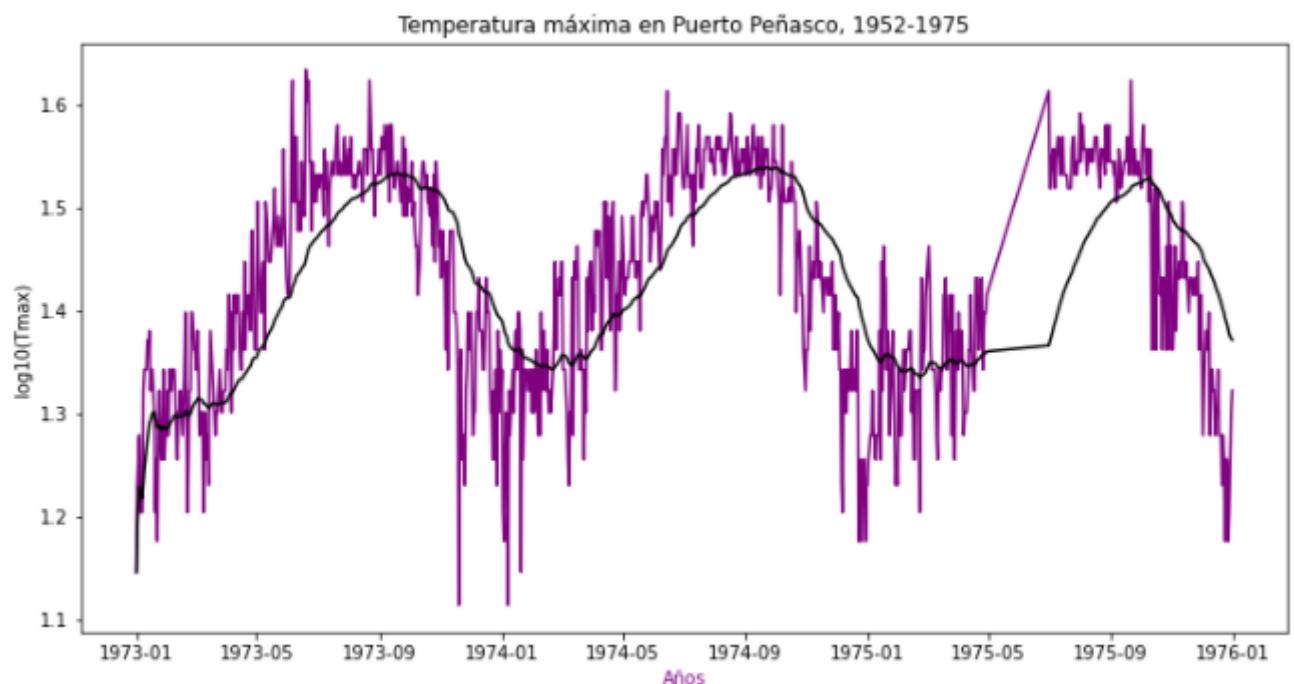
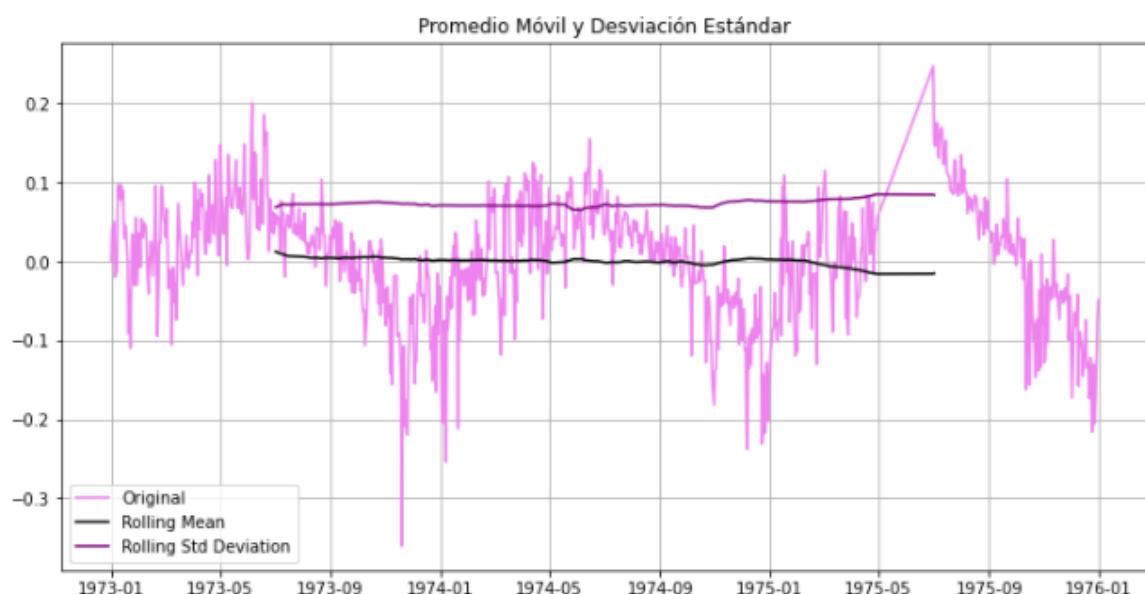


Figura 5



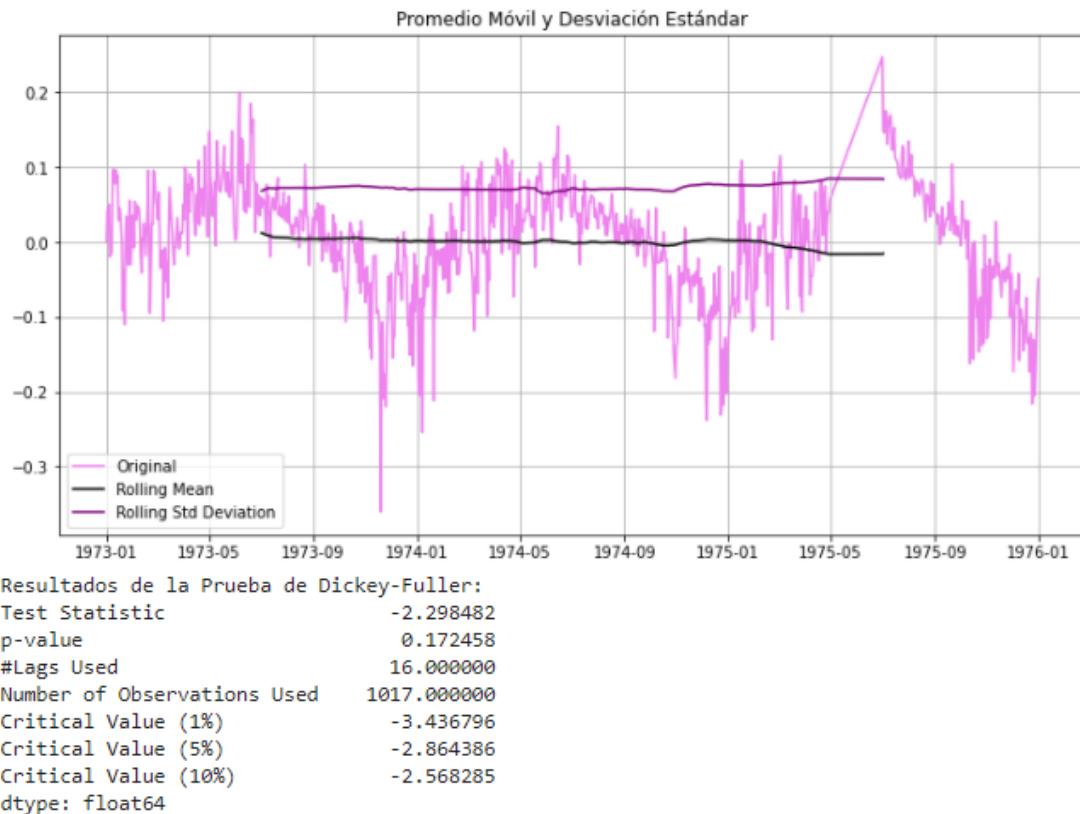
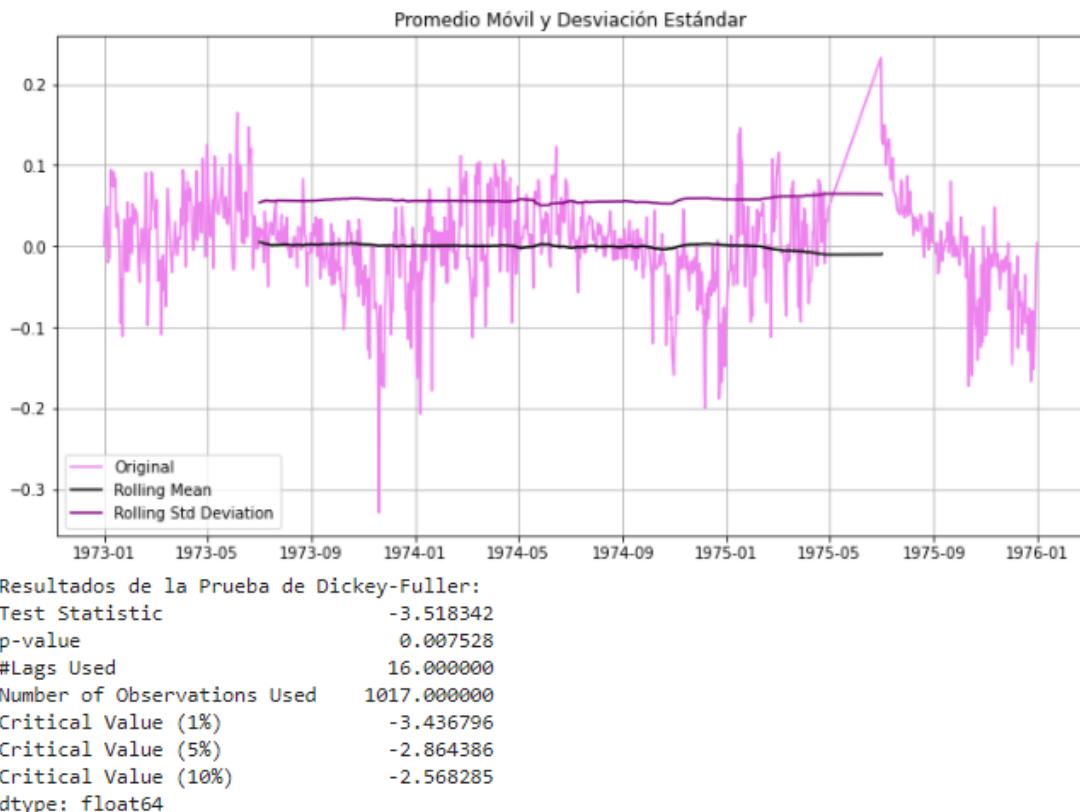
Resultados de la Prueba de Dickey-Fuller:

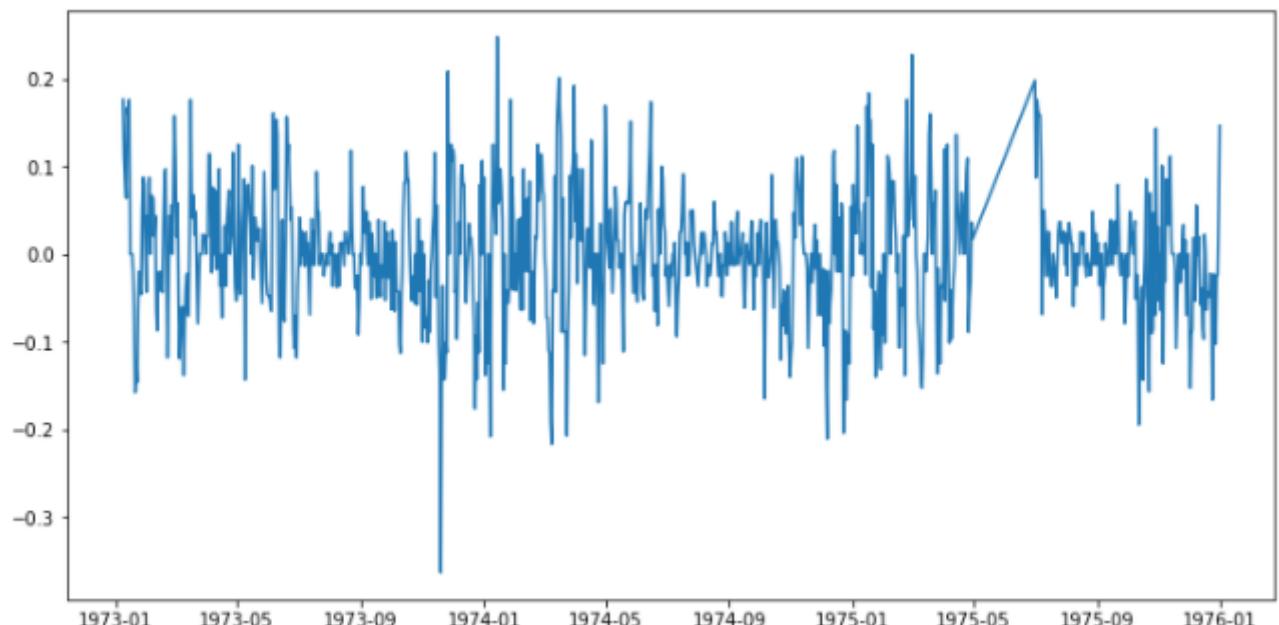
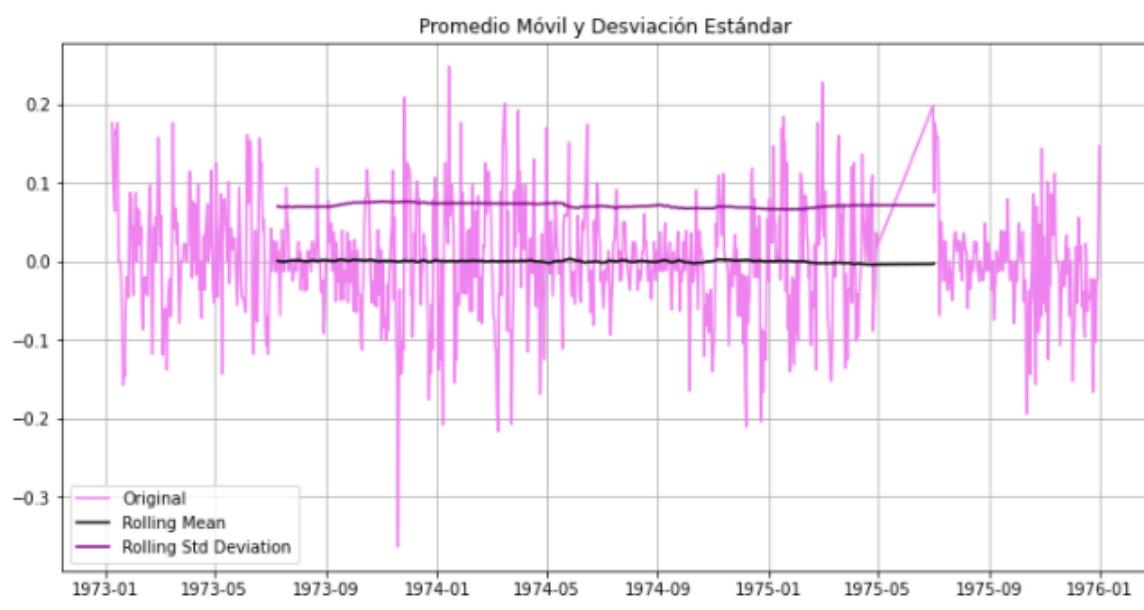
```

Test Statistic           -2.298482
p-value                 0.172458
#Lags Used              16.000000
Number of Observations Used 1017.000000
Critical Value (1%)      -3.436796
Critical Value (5%)       -2.864386
Critical Value (10%)      -2.568285
dtype: float64

```

Figura 6

**Figura 7****Figura 8**

**Figura 9**

Resultados de la Prueba de Dickey-Fuller:

```

Test Statistic      -6.116969e+00
p-value            9.045594e-08
#Lags Used        2.200000e+01
Number of Observations Used 1.004000e+03
Critical Value (1%) -3.436880e+00
Critical Value (5%) -2.864423e+00
Critical Value (10%) -2.568305e+00
dtype: float64

```

Figura 10

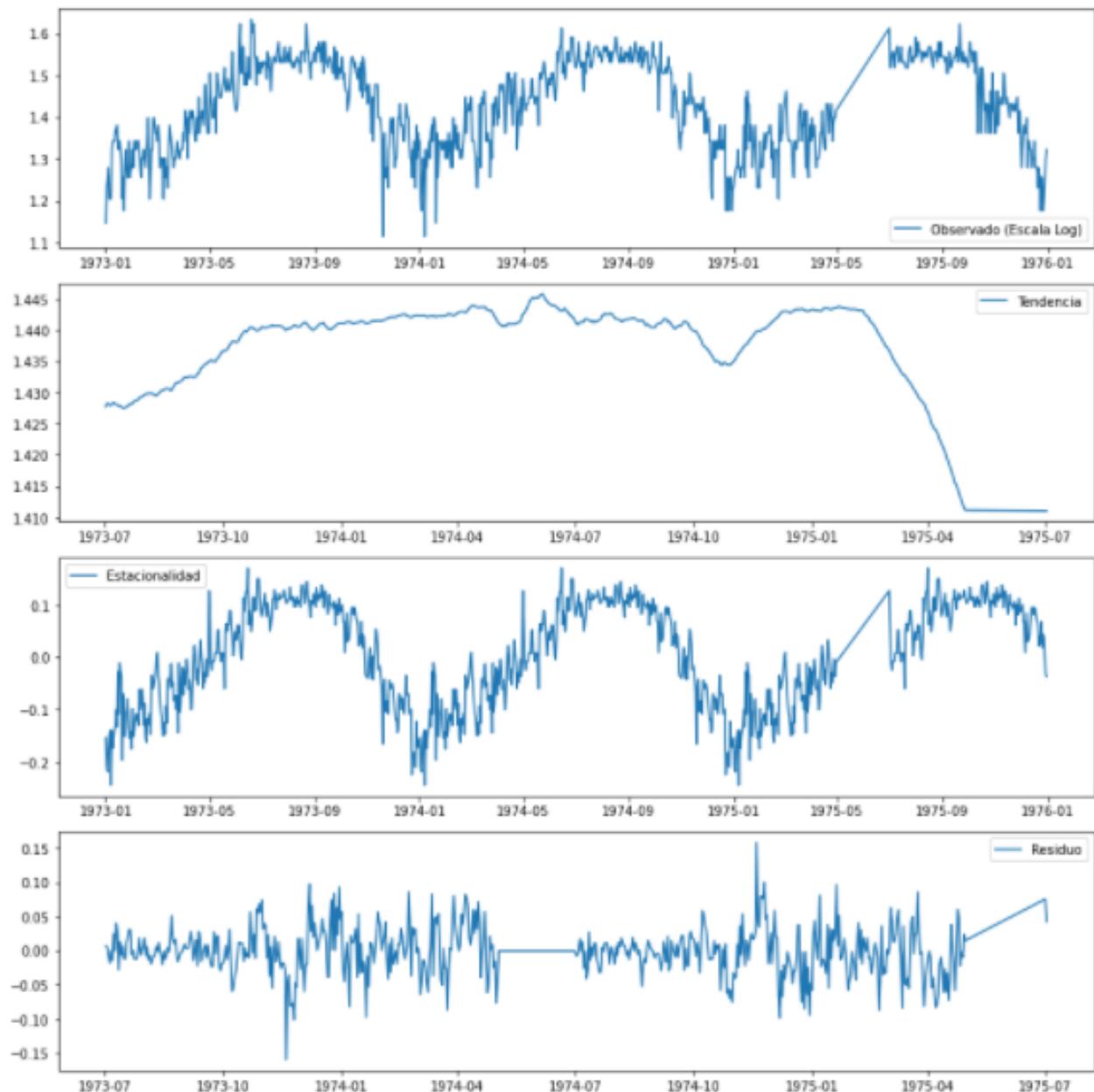
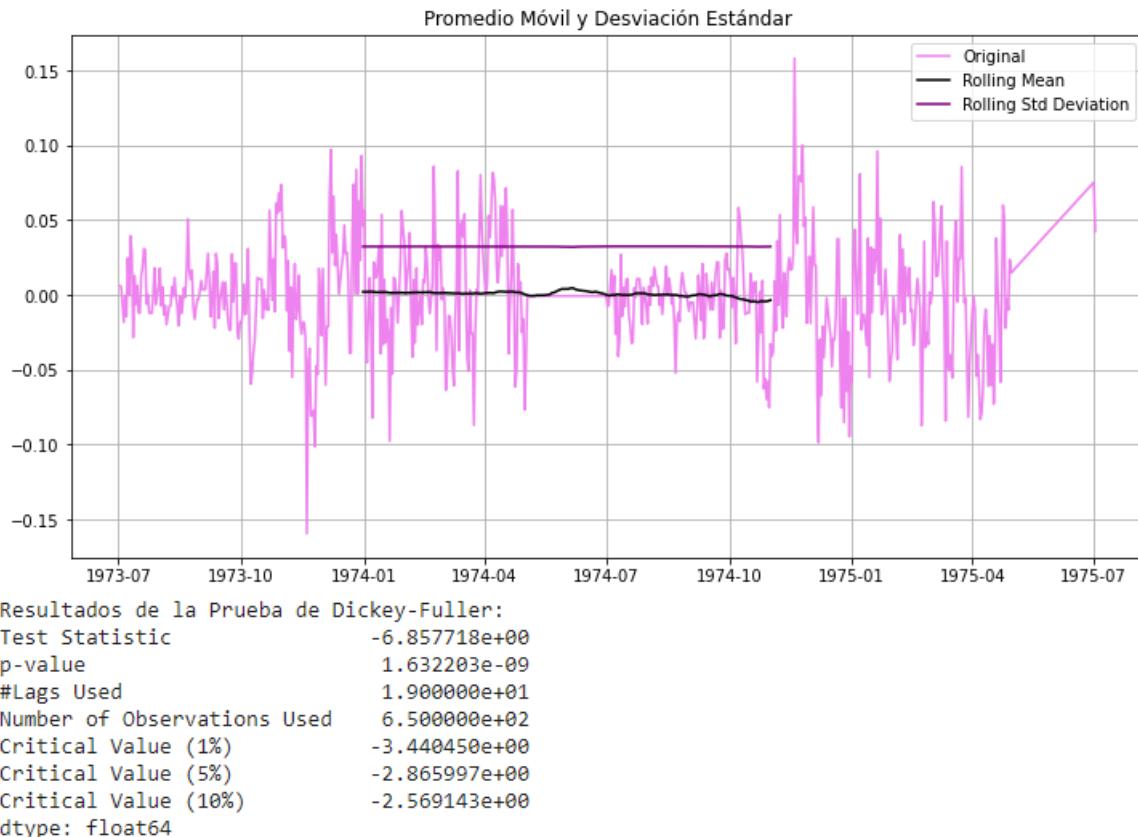
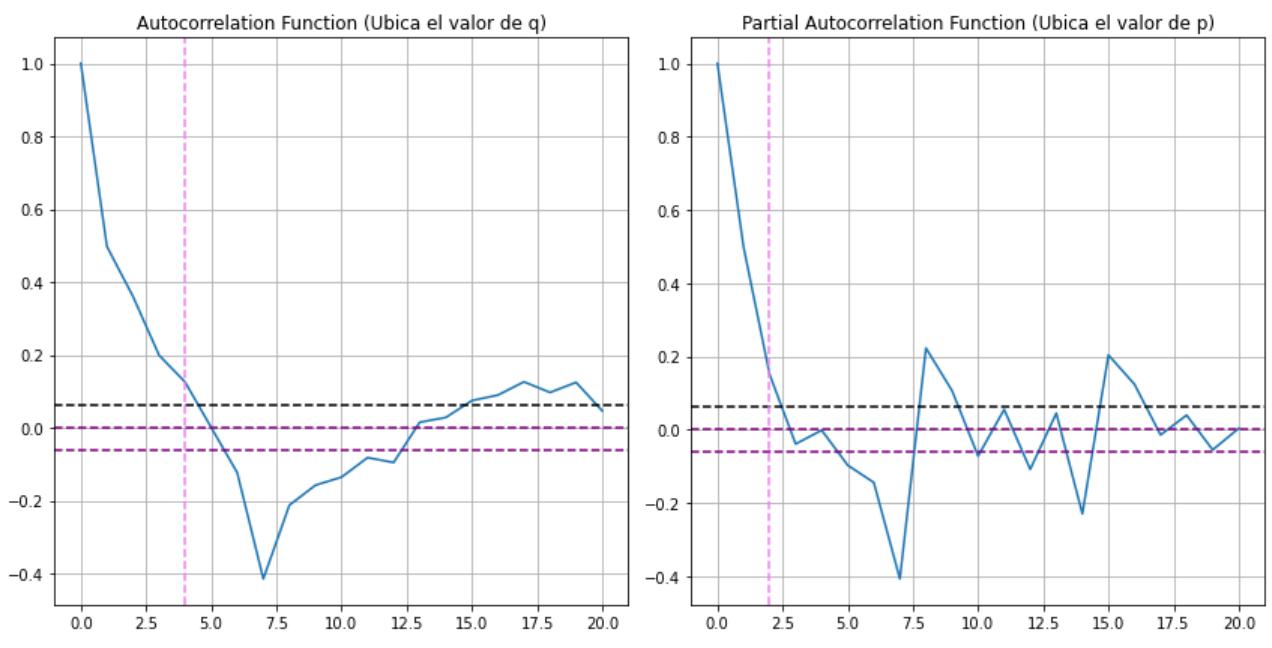


Figura 11

**Figura 12**

2.1.2. Actividad 6.2

**Figura 13**

```

ARIMA Model Results
=====
Dep. Variable: D.Tmax   No. Observations: 1033
Model: ARIMA(4, 1, 0)   Log Likelihood 1771.874
Method: css-mle   S.D. of innovations 0.044
Date: Thu, 25 Feb 2021 AIC -3531.748
Time: 03:39:51   BIC -3502.107
Sample: 1   HQIC -3520.500
=====

      coef    std err        z     P>|z|    [ 0.025    0.975]
-----
const    9.62e-05    0.001    0.138    0.890    -0.001    0.001
ar.L1.D.Tmax -0.4592    0.031   -14.744    0.000    -0.520    -0.398
ar.L2.D.Tmax -0.2353    0.034   -6.956    0.000    -0.302    -0.169
ar.L3.D.Tmax -0.1913    0.034   -5.650    0.000    -0.258    -0.125
ar.L4.D.Tmax -0.0568    0.031   -1.796    0.073    -0.117    0.005
Roots
=====
      Real      Imaginary      Modulus      Frequency
-----
AR.1    0.5179    -1.6996j    1.7768    -0.2029
AR.2    0.5179    +1.6996j    1.7768     0.2029
AR.3   -2.2253    -0.8373j    2.3776    -0.4427
AR.4   -2.2253    +0.8373j    2.3776     0.4427
-----
```

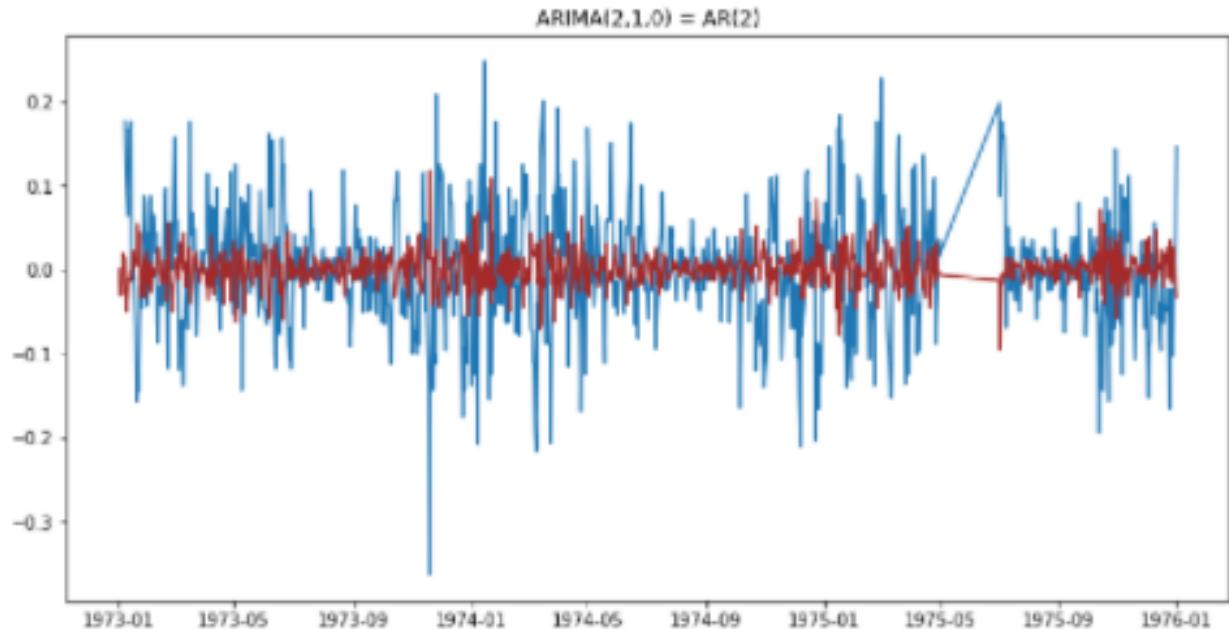


Figura 14

ARIMA Model Results

```
=====
Dep. Variable: D.Tmax   No. Observations: 1033
Model: ARIMA(0, 1, 2)   Log Likelihood 1775.231
Method: css-mle   S.D. of innovations 0.043
Date: Thu, 25 Feb 2021 AIC -3542.462
Time: 03:39:51   BIC -3522.701
Sample: 1   HQIC -3534.964
=====
```

	coef	std err	z	P> z	[0.025	0.975]
const	7.415e-05	0.001	0.134	0.893	-0.001	0.001
ma.L1.D.Tmax	-0.4933	0.030	-16.302	0.000	-0.553	-0.434
ma.L2.D.Tmax	-0.0989	0.033	-3.038	0.002	-0.163	-0.035

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	1.5470	+0.0000j	1.5470	0.0000
MA.2	-6.5332	+0.0000j	6.5332	0.5000

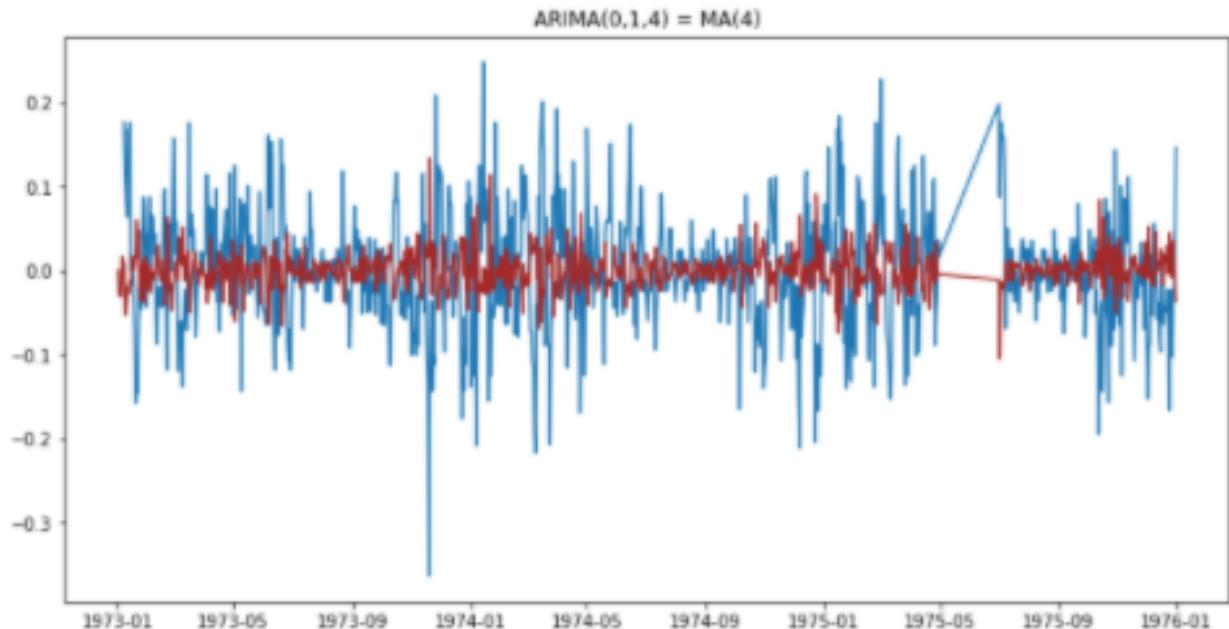


Figura 15

ARIMA Model Results

```
=====
Dep. Variable: D.Tmax   No. Observations: 1033
Model: ARIMA(4, 1, 2)   Log Likelihood 1787.447
Method: css-mle   S.D. of innovations 0.043
Date: Thu, 25 Feb 2021 AIC -3558.894
Time: 03:39:53   BIC -3519.372
Sample: 1   HQIC -3543.896
=====
```

	coef	std err	z	P> z	[0.025	0.975]
const	4.123e-05	0.000	0.112	0.911	-0.001	0.001
ar.L1.D.Tmax	-0.1119	0.226	-0.494	0.621	-0.556	0.332
ar.L2.D.Tmax	0.3278	0.092	3.574	0.000	0.148	0.508
ar.L3.D.Tmax	0.0523	0.049	1.058	0.290	-0.045	0.149
ar.L4.D.Tmax	0.0578	0.036	1.599	0.110	-0.013	0.129
ma.L1.D.Tmax	-0.3787	0.225	-1.683	0.093	-0.820	0.062
ma.L2.D.Tmax	-0.4357	0.190	-2.289	0.022	-0.809	-0.063

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.4803	-0.0000j	1.4803	-0.0000
AR.2	-1.4859	-0.0000j	1.4859	-0.5000
AR.3	-0.4501	-2.7684j	2.8048	-0.2756
AR.4	-0.4501	+2.7684j	2.8048	0.2756
MA.1	1.1415	+0.0000j	1.1415	0.0000
MA.2	-2.0107	+0.0000j	2.0107	0.5000

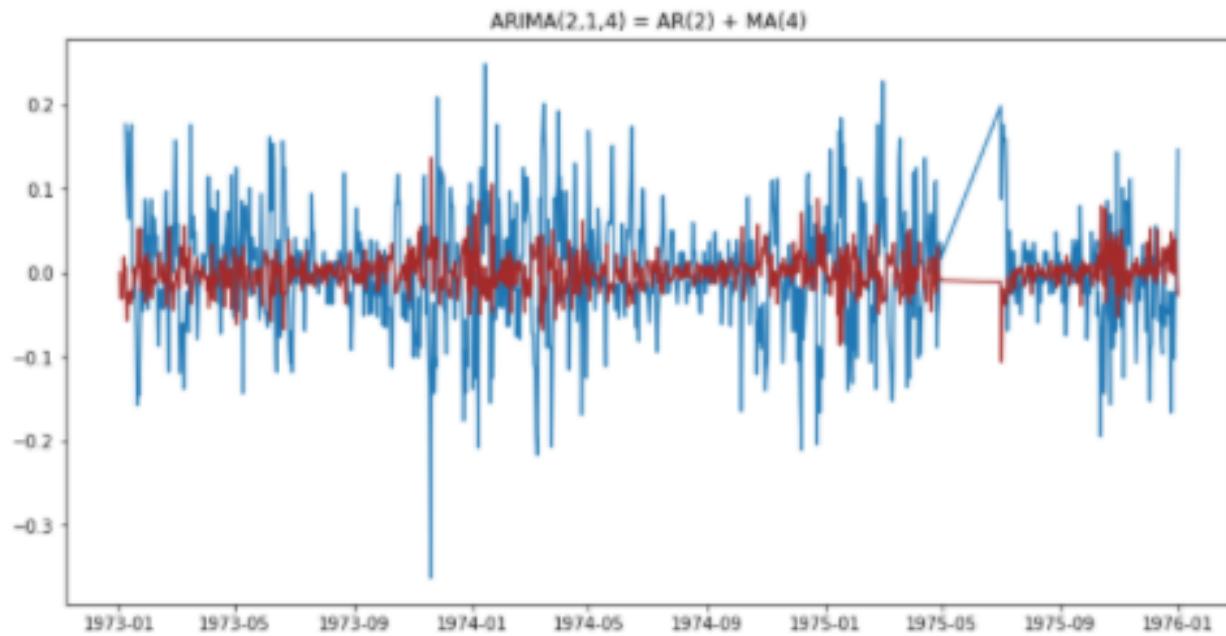
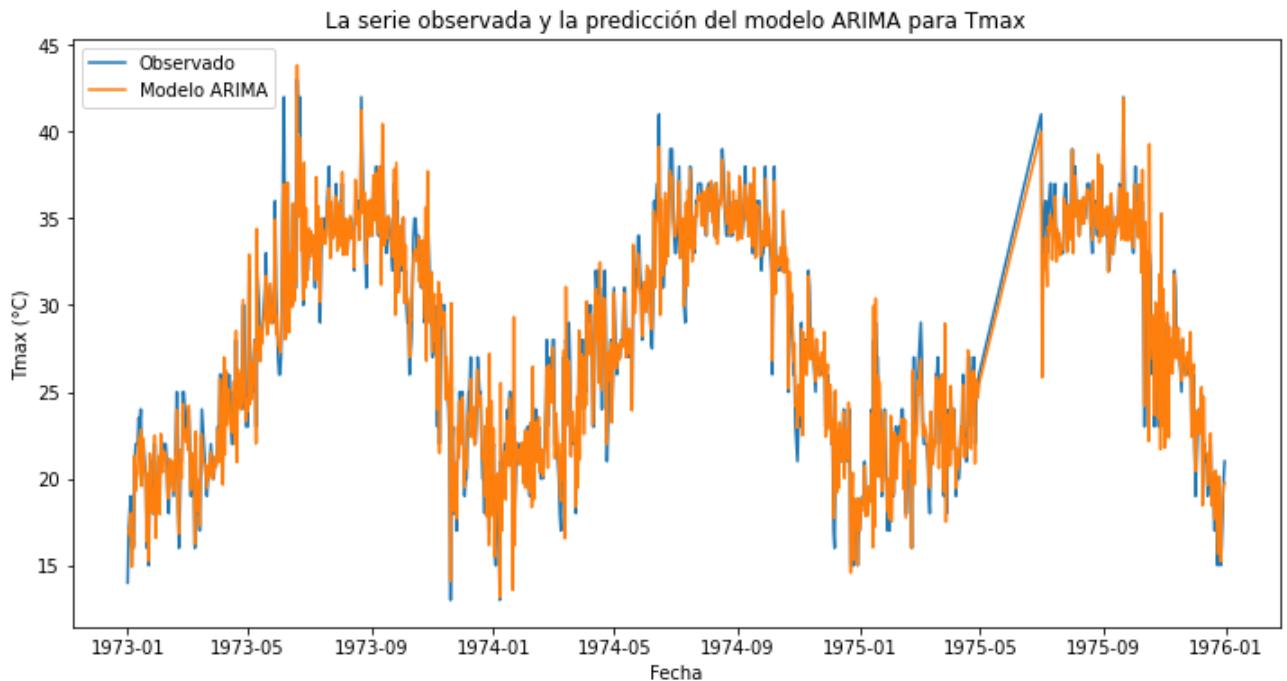
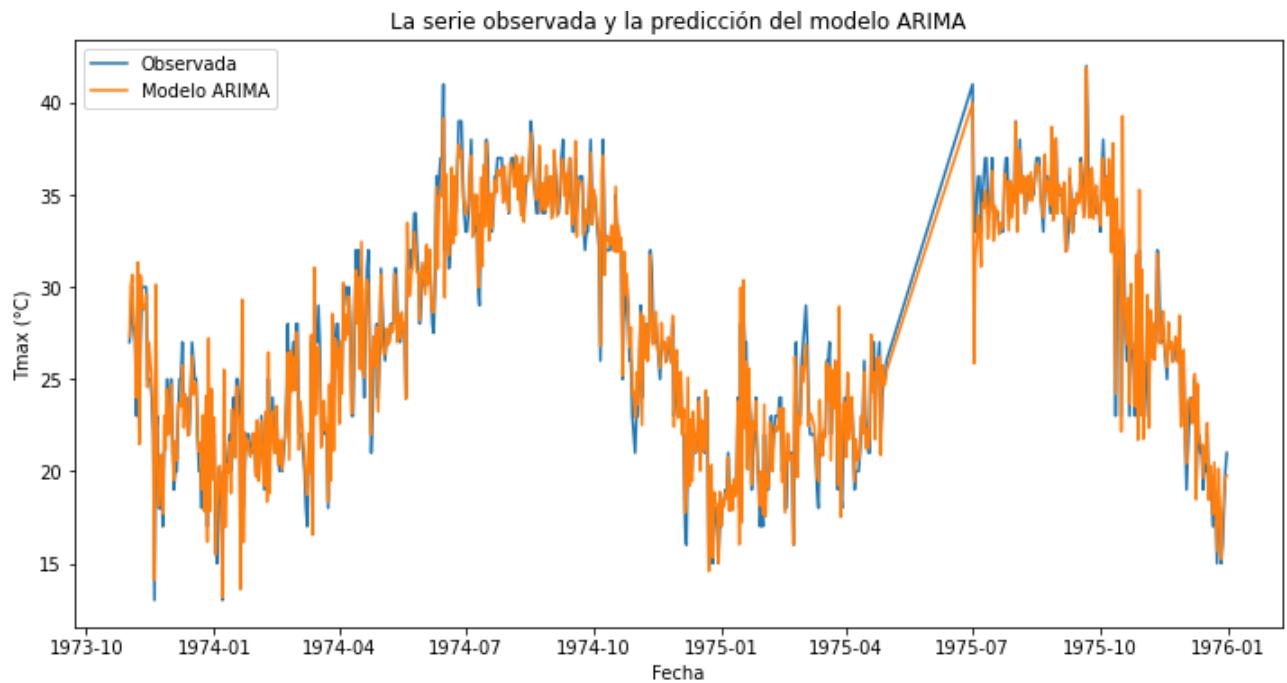


Figura 16

**Figura 17****Figura 18**

3. Conclusión

En la Actividad 6.1, me resultó aún más difícil que en la actividad pasada encontrar un periodo de tiempo sin huecos, pues la estación cuenta con un rango amplio, pero muchos valores faltantes

entre años. Fue tan extremo el caso, que me vi orillada a hacer el análisis de solo 2 años.

En el caso de esta estación, desde un principio podemos notar una linealidad en las gráficas, sin embargo el valor de p fue obtenido hasta después de haber aplicado escalas exponenciales.

Por medio del análisis *halflife*, por otra parte, se pudo observar que mientras más corto es, más lineal se ven el promedio y la desviación.

Luego, en la Actividad 2, con ayuda de la gráfica de la **Figura 13**, podemos observar que el valor aproximado de q es aproximadamente 4, mientras que el valor de p es aproximadamente 2. Con esta información y un valor de d igual a 1, se pudieron desarrollar los tres modelos solicitados.

Una vez realizados los modelos, obtenemos que es el último modelo de ARIMA el que es más conveniente al tener el valor de AIC menor.

La tendencia de temperaturas parece bajar en las temporadas del año de 1975 de manera muy extrema, hasta llegar a un punto en que parece constante. Esto puede ser por que la cantidad de datos analizada es muy pequeña.

Esta actividad me pareció muy entretenida, pero a la vez algo confusa, ya que siento que no terminé de comprender en su totalidad los conceptos, además de que se me complicó mucho la elección de periodo de tiempo, por la falta de datos. Personalmente, considero que necesito practicarlo mucho más. El nivel de complejidad que le asignaría es alto, pero es debido a mi propia ignorancia y no porque le falte algo a la actividad, ya que las instrucciones me parecieron claras.