Series de tiempo

Unidad Educativa: Minería de datos
Docente: Mayra Cristina Berrones Reyes

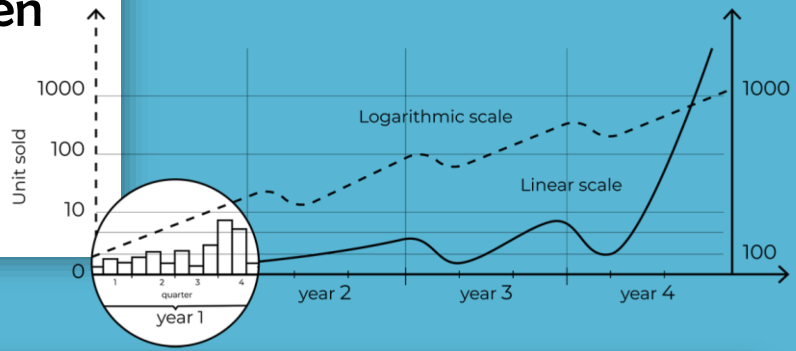
Equipo: 4

Sandra María Cavazos Huerta	1877283
Alberto Elizondo Villarreal	1869764
Marlene Michel Sepúlveda Bermúdez	1862646
Arturo Isaac Sánchez Tovar	1867031
Diana Villarreal Garza	1877498

Definición y Características

¿Qué es una serie de tiempo?

Es una secuencia de datos, observaciones o valores, medidos en determinados momentos y ordenados cronológicamente.



Se realiza un análisis sobre ellas para identificar patrones en los datos. Se busca entender qué sucede a medida que el tiempo va avanzando. Uno de los usos más habituales es su análisis para predicción y pronóstico.

Tipos de cambio o variación

- Tendencia o variación secular
- Fluctuación o variación cíclica
- Variación estacional
- Variación irregular

Definición y Características - Pronóstico

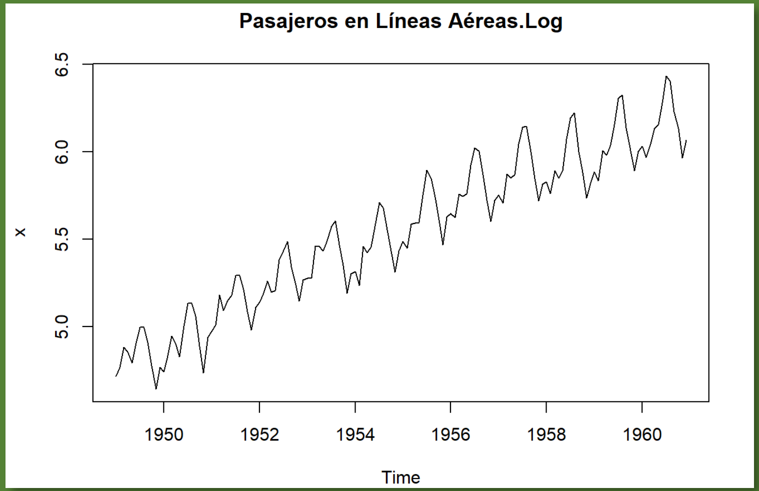
Los pronósticos son un método que se utiliza en el análisis de las series de tiempo para predecir una variable de respuesta.

Antes de generar pronósticos, se deben **ajustar varios modelos** candidatos a los datos para determinar cuál es el modelo más estable y exacto.

Los pronósticos se basan en patrones de datos existentes.

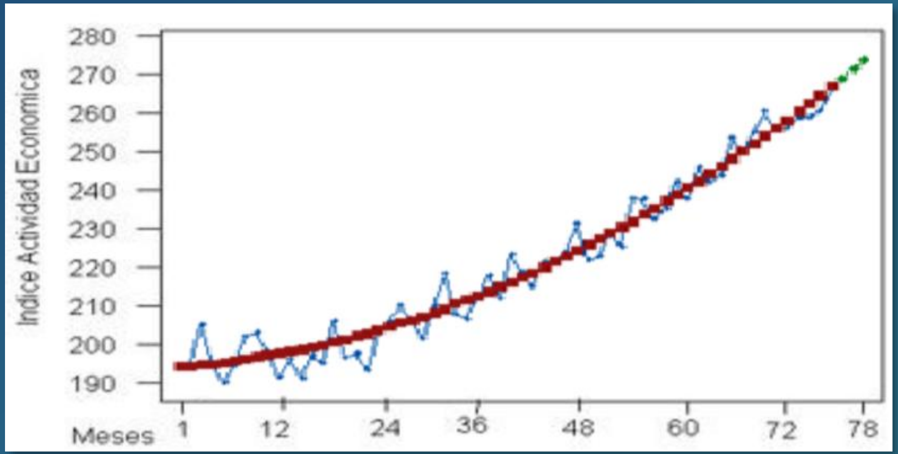
Definición y Características - Tendencia

La tendencia de una serie viene dada por el movimiento general a largo plazo de la serie. El método que se utiliza para obtener la línea recta o ecuación de mejor ajuste es el Método de Mínimos Cuadrados.



Tendencia lineal

Se aproxima a una línea recta que aumenta o disminuye a un ritmo constante.

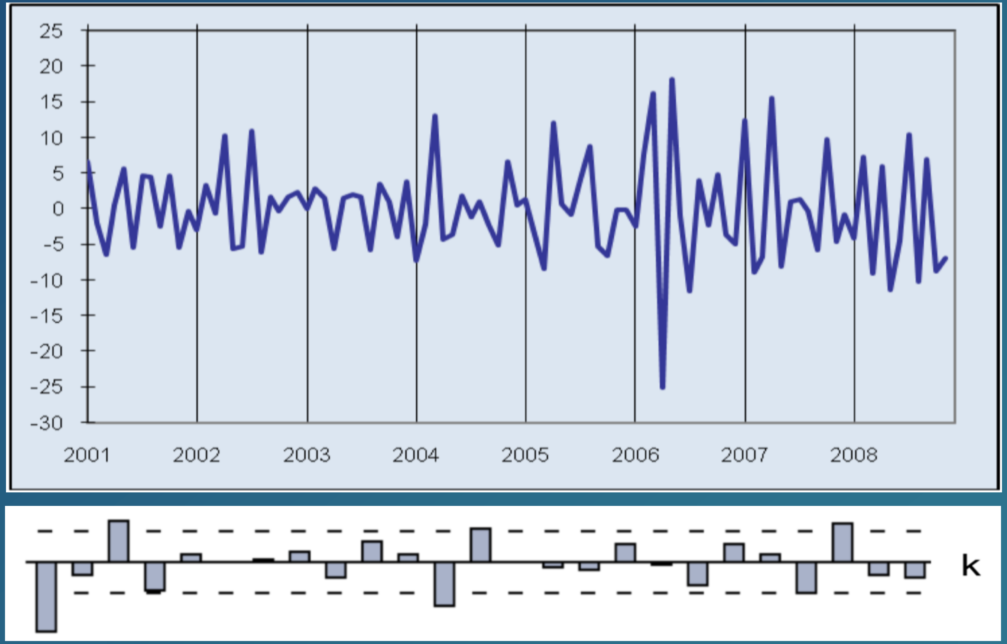


Tendencia no lineal

Comportamiento curvilíneo. La tendencia puede ser: polinomial, logarítmica, exponencial, potencial.

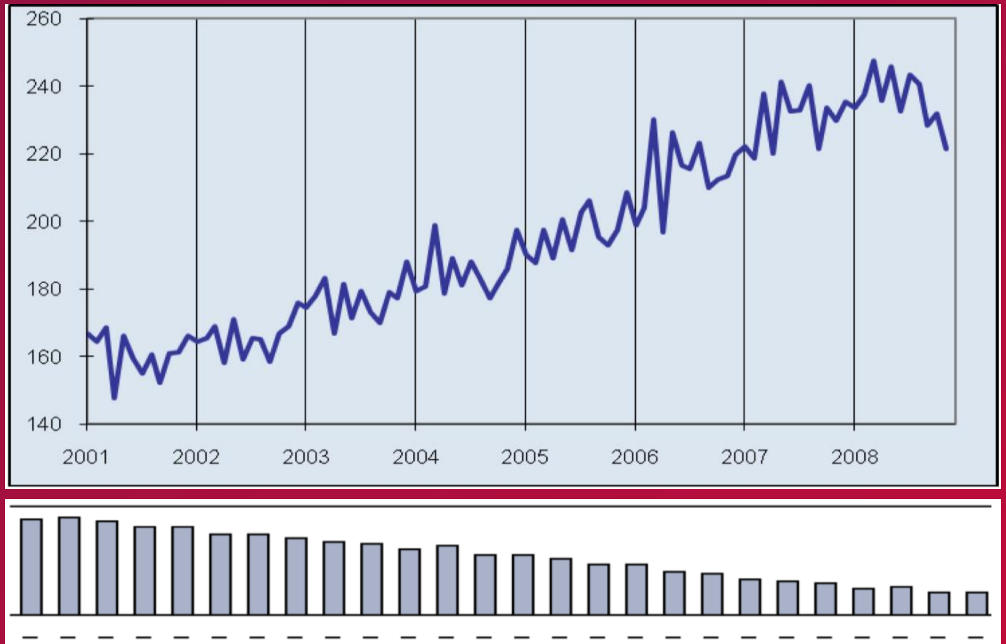
Definición y Características - Clasificación

Serie estacionaria



– Media constante
– Varianza constante
– Función de autocorrelación decae rápidamente cuando aumenta k.

Serie no estacionaria



– Media no es constante
– Varianza no es constante
– Función de autocorrelación decae lentamente

Modelos de series estacionarias

MODELO AR(p)

Los modelos **autoregresivos** se basan en la idea de que **el valor actual de la serie**, X_t , puede **explicarse** en función de **p valores pasados** $X_{t-1}, X_{t-2}, X_{t-3}, \ldots,$ X_{t-p}

Expresión matemática:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \emptyset_3 X_{t-3} + \cdots + \emptyset_p X_{t-p} + e_t$$

MODELO MA(q)

En el método de **promedio móvil** se especifica un número constante de datos y se puede **calcular una media** (o promedio) para las observaciones más recientes.

Modelo "determinados por una fuente externa".

Expresión matemática:

$$X_t = e_t + \beta_1 e_{t-1} + \cdots + \beta_q e_{t-q}$$

MODELO ARMA(p,q)

Es muy probable que una serie de tiempo, X_t, tenga **características** de **AR** y de **MA** a la vez y, por consiguiente, sea **ARMA** .

Así, X_t sigue un proceso ARMA(p,q), en este habrá p términos autoregresivos y q términos de media móvil.

Expresión matemática:

$$Xt = \emptyset_1 X_{t-1} + \cdots + \emptyset_p X_{t-p} + \beta_1 e_{t-1} + \cdots + \beta_{1q} e_{t-q} + c$$

Criterio de Desempeño (AIC)

AIC proporciona

Un medio para la selección del modelo, considera:

- La bondad de ajuste
- Complejidad

AIC calcula

El 'criterio de información' de Akaike para uno o varios objetos de modelo ajustados para obtener un valor de log verosimilitud:
-2*log-verosimilitud + k*npar

AIC sirve para

- Seleccionar el mejor modelo.
- Medir la distancia de cada modelo bajo comparación de datos verdaderos.

AIC obtiene

Una estimación de la distancia relativa esperada entre cada modelo estimado y los verdaderos mecanismos que han generado los datos observados.

Se buscan modelos con AIC pequeños.

npa= número de parámetros en el modelo ajustado.

k = 2 para el AIC usual.
k = log (n) (n es el número de observaciones) para el llamado BIC.

Aplicaciones

Áreas

- Economía

- Demografía

- Medicina

- Meteorología

Aplicaciones en economía

Estudio de:

- Precios de un artículo.
- Precio del dólar.
- Precios de acciones.

Aplicaciones en demografía

Al estudiar las tasa de natalidad o mortalidad.

Aplicaciones en medicina

A través de estudios de:

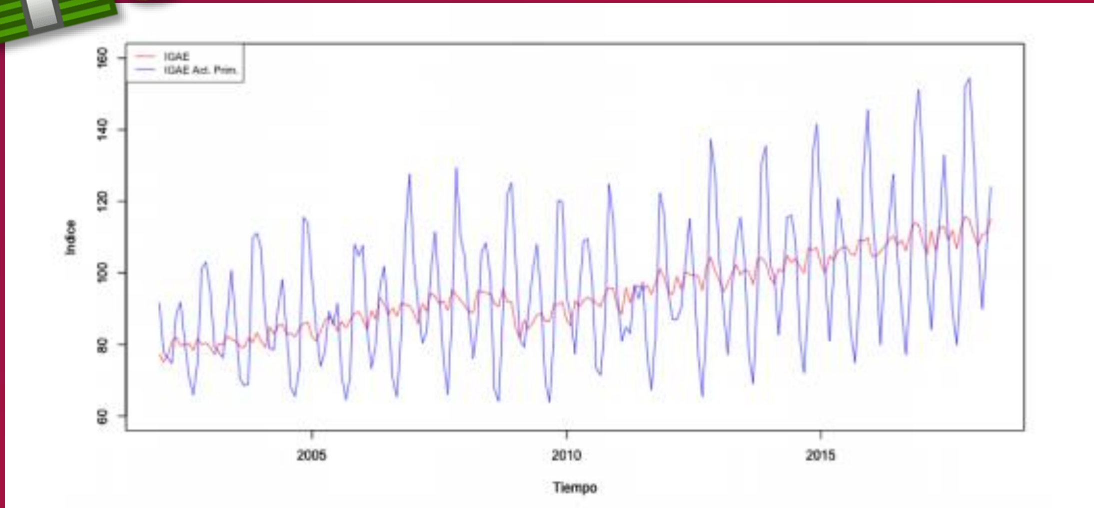
- Electrocardiograma.
- Electroencefalograma.

Aplicaciones en meteorología

Al analizar la:

- Temperatura máxima diaria.
- Velocidad del viento (energía eólica).
- Energía solar.

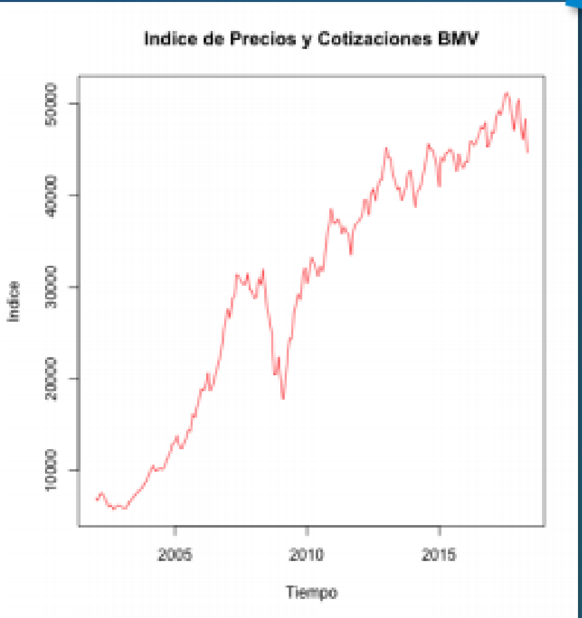
Aplicaciones



Indicador Global de la Actividad Económica (IGAE)

El IGAE en su versión global y en su versión para las actividades primarias, muestra la evolución de la actividad económica del país.

Se puede conocer si se espera que la economía crezca o decrezca, así como patrones de cambio.

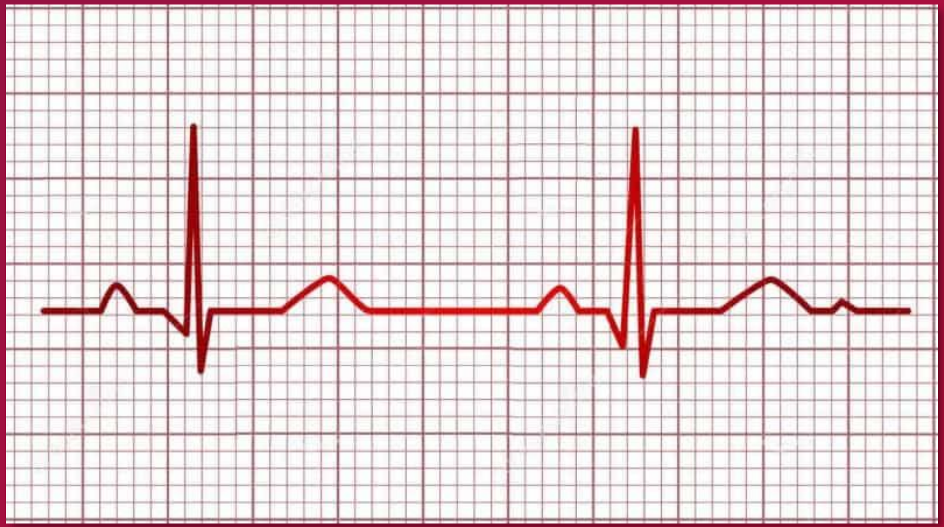


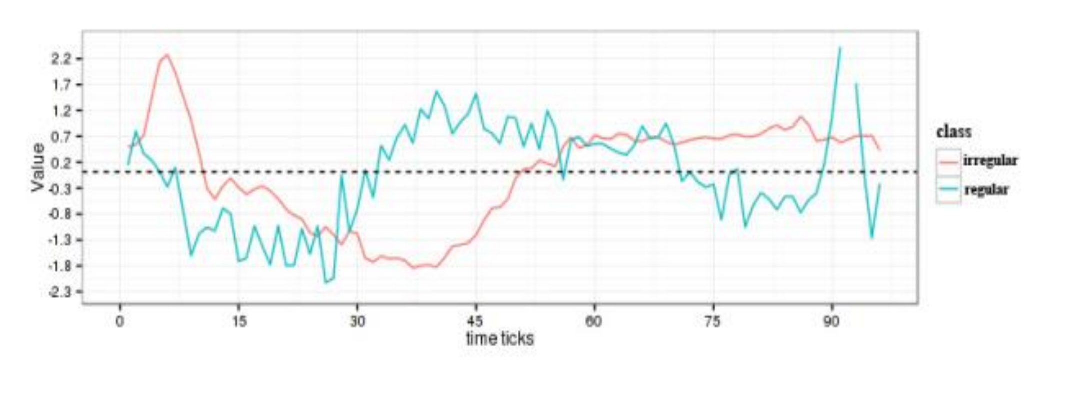
Índice de Precios y Cotizaciones de la BMV (IPC)

EL IPC refleja el rendimiento del capital promedio invertido en las empresas que cotizan en la BMV.

Se puede conocer la tendencia del índice y a su vez el comportamiento de los rendimientos a largo plazo.

Aplicaciones





Electrocardiogramas

Cada dato almacenado en un ECG es una serie de tiempo registrada por un electrodo durante cada pulsación del corazón.

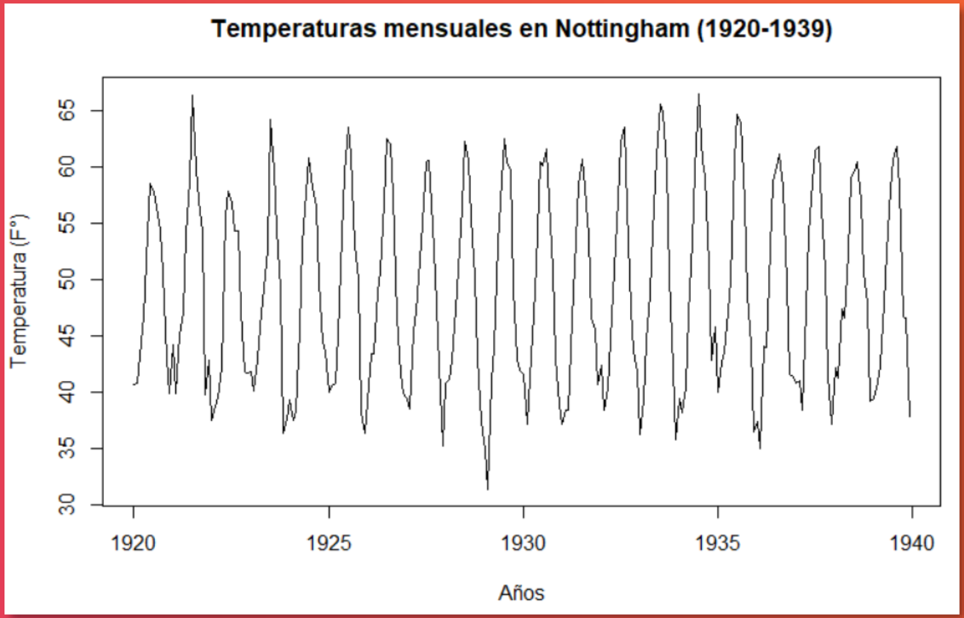
Se puede estudiar con el fin de conocer si un conjunto de datos es clasificado en normal o anormal.

Ejemplo-Gráfica

Ejemplo

Ejemplo hecho en R

Se trata de la Serie de Tiempo de "**nottem**": Temperaturas promedio del castillo de Nottingham en Fahrenheit desde 1920 a 1939.




```
#Usando la Serie de Tiempo de "Nottem"----
#Serie de Tiempo----
data("nottem")
help(nottem)
#Serie de Tiempo de temperaturas promedio del
#Castillo de Nottingham en Fahrenheit por 20 años.

#Grafica de Serie de Tiempo----
plot(nottem,
     main = "Temperaturas mensuales en Nottingham (1920-1939)",
     ylab = "Temperatura (F°)",
     xlab = "Años")
```

Ejemplo-Tendencias

```
#TENDENCIA DEL MODELO LINEAL----
reg_lineal = lm(nottem~time(nottem))
reg_lineal
#Modelo Lineal con B0 = -92.72334 y B1 = 0.07345
#Checamos con p-valor si esta regresion es confiable
summary(reg_lineal)
#H0: No tendencia
#Ha: Tiene Tendencia
#R H0 si p-valor < 0.05
> summary(reg_lineal)

Call:
lm(formula = nottem ~ time(nottem))

Residuals:
    Min      1Q  Median      3Q     Max
-17.675  -7.575  -1.620   8.028  17.882

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -92.72334  185.13281  -0.501    0.617
time(nottem)   0.07345    0.09593   0.766    0.445

Residual standard error: 8.58 on 238 degrees of freedom
Multiple R-squared:  0.002458,  Adjusted R-squared:  -0.001734
F-statistic: 0.5864 on 1 and 238 DF,  p-value: 0.4446
```

Conclusión

Ya que el p-valor es de 0.4446

P-valor > 0.05
No se rechaza H0 con
R2 Ajustada de -0.001734

Por lo tanto:
No hay evidencia para mostrar que se tiene tendencia lineal.

Ejemplo-Tendencias

```r
#MODELO CUBICO----
tiempo<-as.numeric(time(nottem))
modelo_cub<-lm(nottem~poly(tiempo,degree=3))

summary(modelo_cub)
#H0: No tendencia
#Ha: Tiene Tendencia
#R H0 si p-valor < 0.05
> summary(modelo_cub)

Call:
lm(formula = nottem ~ poly(tiempo, degree = 3))

Residuals:
    Min      1Q  Median      3Q     Max 
-17.430  -7.660  -1.627   8.145  17.559 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 49.040      0.556  88.207   <2e-16 ***
poly(tiempo, degree = 3)1    6.570      8.613   0.763    0.446    
poly(tiempo, degree = 3)2    2.788      8.613   0.324    0.746    
poly(tiempo, degree = 3)3   -2.208      8.613  -0.256    0.798    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.613 on 236 degrees of freedom
Multiple R-squared:  0.003178,	Adjusted R-squared:  -0.009494 
F-statistic: 0.2508 on 3 and 236 DF,  p-value: 0.8607
```

Conclusión

Ya que el p-valor es de 0.8607

p-valor > 0.05
No se rechaza H0 con
R2 Ajustada de -0.009494

Por lo tanto:
No hay evidencia para mostrar que se tiene tendencia cúbica.

Ejemplo-Varianza

```
65 ▾ #VARIANZA----
66   adf.test(nottem, alternative="stationary")
67   #H0 No Estacionaria en Varianza
68   #Ha Estacionaria en Varianza
69   #R H0 si p-valor < 0.05
> adf.test(nottem, alternative="stationary")

        Augmented Dickey-Fuller Test

data:  nottem
Dickey-Fuller = -12.998, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Conclusión

Ya que el p-valor es de 0.01

p-valor < 0.05
Se rechaza H0 con

Por lo tanto:
La Serie de Tiempo es estacionaria en varianza.

Conclusión

Ya que no se tiene tendencia en media y es estacionaria en varianza, la serie de tiempo se clasifica como estacionaria.

Ejemplo-ARMA

```
> aic.arma
[1] 1112.724
> orden.arma
[1]  9  0 10
```

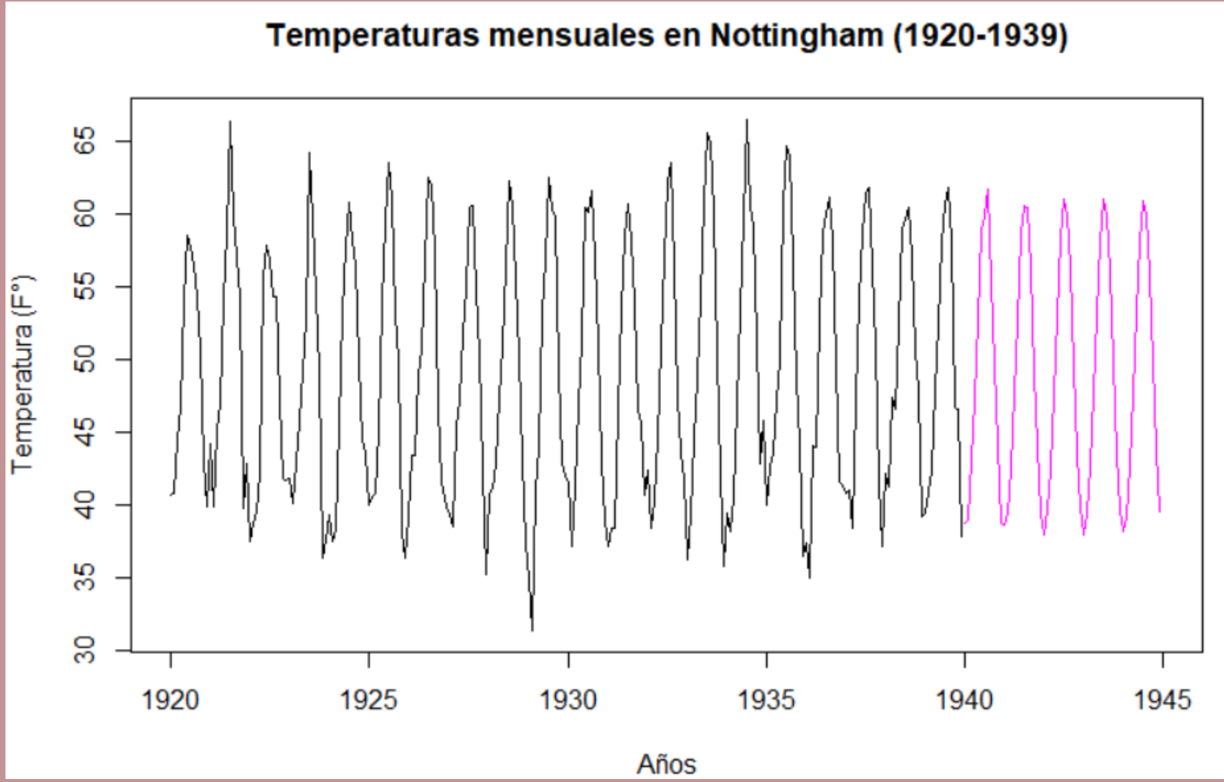
```r
#MODELO ARMA----
#Ciclo doble para el mejor orden, tomando en cuenta el AIC
aic.arma <- 100000
orden.arma <- c(0,0,0)
for(j in 0:10){
  for (i in 0:10){
    try(aic.xx <- AIC(arima(nottem,order = c(j,0,i),method ="ML")))
    if (aic.arma > aic.xx){
      aic.arma <- aic.xx
      orden.arma <- c(j,0,i)
    }
  }
}
aic.arma
orden.arma
```

Ejemplo-Predicción

```r
#GRÁFICA PREDICCIÓN----
plot(nottem,
     main = "Temperaturas mensuales en Nottingham (1920-1939)",
     ylab = "Temperatura (F°)",
     xlab = "Años",
     xlim = c(1920,1945))

#Predicción ARMA
pred.arma <- predict(arima(nottem,orden.arma),n.ahead = 60)$pred
lines(pred.arma,col = "magenta")
```

Gráfico



Preguntas

Responde a lo siguiente:

1. ¿Por qué es importante el análisis de una serie de tiempo?
2. ¿Cuáles son las características de una serie estacionaria?
3. ¿En qué consiste el modelo ARMA(p,q)?
4. ¿Qué obtiene el criterio de desempeño AIC?
5. ¿Cuál es una aplicación del análisis de series de tiempo?

Referencias

- Briega, L. R. E. (2016, 26 septiembre). *Series de tiempo con Python*. Github. https://relopezbriega.github.io/blog/2016/09/26/series-de-tiempo-con-python/
- González Castellanos, M., & Soto-Valero, C. (2013). *Minería de datos para series temporales* (1.ª ed., Vol. 1). Editorial Feijó - CDICT UCLV. https://doi.org/10.13140/RG.2.1.2571.9841
- Hernández, S. (2015). *Análisis de Series de Tiempo [Diapositivas].* sepal. https://www.cepal.org/sites/default/files/courses/files/01_1_conociendo_una_serie_de_tiempo.pdf
- John Villavicencio, (2021). Introducción a Series de Tiempo. http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D
- Ríos, G. (2008). *Series de Tiempo* (1.ª ed., Vol. 1). Facultad De Ciencias Físicas y Matemáticas. https://www.u-cursos.cl/ingenieria/2010/1/CC52A/1/material_docente/bajar?id_material=296003
- *Series de tiempo.* (n.a.). estadistica.mat.uson. Recuperado 5 de septiembre de 2021, de http://www.estadistica.mat.uson.mx/Material/seriesdetiempo.pdf