UNAP
F I N E S I

Series de Tiempo

V I Semestre
Grupo: B

Mtr. Alcides Ramos Calcina
FINESI

CONSTRUCCIÓN DE MODELOS ARIMA

1°. Identificación

Introducción

- En este capítulo se presenta la estrategia de construcción de modelos para series de tiempo desarrollada por Box-Jenkins (1970).

- Esta estrategia consta de cuatro etapas fundamentales, podría considerarse como una de las principales contribuciones de estos dos autores al análisis de series de tiempo.

- Se constituye como una técnica avanzada que hace uso de sofisticados recursos matemático-estadísticos.

- Existe una clara y consolidada guía de aplicación empírica de la misma que permite pasar con facilidad de las situaciones de laboratorio que crea la teoría, a la práctica.

- Los modelos ARIMA han demostrado una gran utilidad en la predicción a corto plazo de series de alta frecuencia. Ese es su campo natural de aplicación.

1. Estrategia de modelización

- Se han analizado las propiedades de los procesos ARIMA(p, d, q), tanto estacionarios $(d = 0)$ como no estacionarios ($d > 0$):

$$\phi_p(L)\Delta^d Y_t = \delta + \theta_q(L)a_t$$

- Si se conocen los parámetros del modelo teórico para el proceso Y_t, a partir de una realización concreta del ruido blanco, $a_1, a_2, \ldots, a_T$, y de valores iniciales para Y_t, a se genera la serie temporal $Y_1, Y_2, \ldots, Y_T$, que es una realización de tamaño T del proceso estocástico.

- La estructura es, por lo tanto,

$$\text{Proceso estocástico} \begin{cases} \phi(L)\Delta^d Y_t = \delta + \theta(L)a_t \\ a_t \sim N(0, \sigma^2) \end{cases}$$

$$\downarrow$$

Generación

$$\downarrow$$

Realización: $Y_1, Y_2, \ldots, Y_T$

1. Estrategia de modelización

- En la aplicación de la metodología Box-Jenkins el punto de partida es el contrario: se conocen los valores de la serie temporal $Y_1, Y_2, \ldots, Y_T$ y se trata de determinar la estructura ARIMA(p, d, q) que la ha podido generar.

Realización: $Y_1, Y_2, \ldots, Y_T$

$\downarrow$

Inferencia

$\downarrow$

$$\text{Proceso estocástico} \begin{cases} \phi(L)\Delta^d Y_t = \delta + \theta(L)a_t \\ a_t \sim N(0,\sigma^2) \end{cases}$$

- La construcción de modelos ARIMA se lleva a cabo de forma iterativa mediante un proceso en el que se pueden distinguir cuatro etapas:

1) Identificación

- Utilizando los datos y/o cualquier tipo de información disponible sobre cómo ha sido generada la serie, se intentará sugerir una subclase de modelos ARIMA(p, d, q) que merezca la pena ser investigada.

- El objetivo es determinar los órdenes p, d y q que parecen apropiados para reproducir las características de la serie bajo estudio y si se incluye o no la constante δ.

- En esta etapa es posible identificar más de un modelo candidato que pueda generar la serie.

2) Estimación

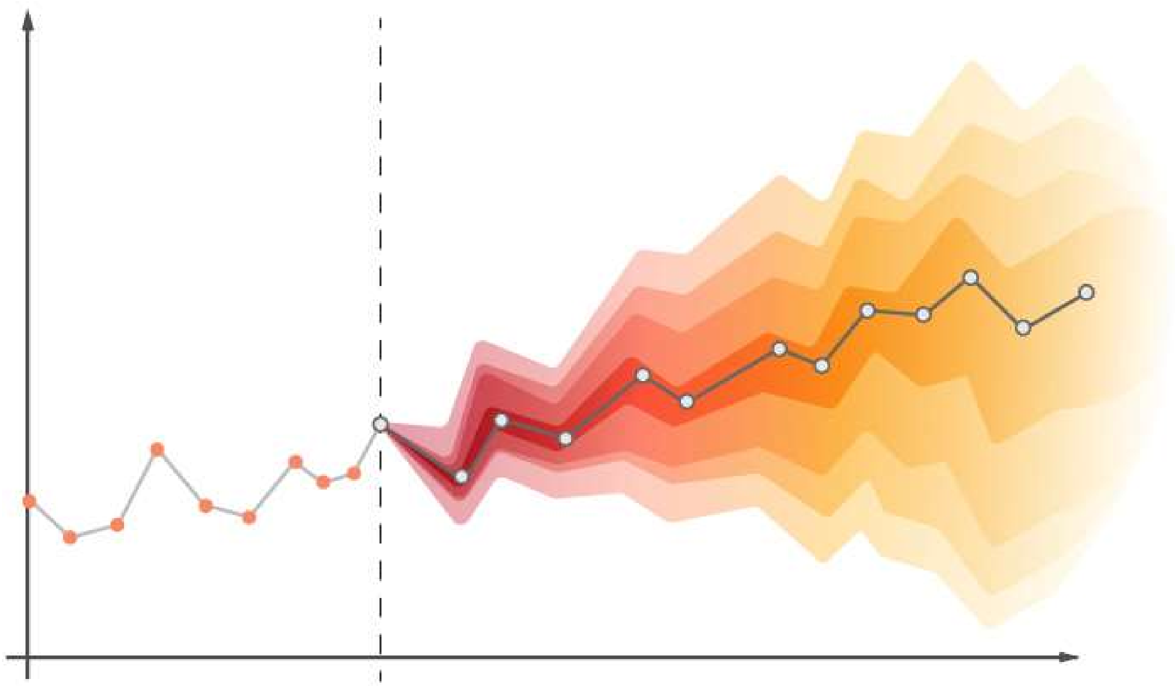
- La estimación de los parámetros involucrados en el modelo, a través de técnicas de estimación no lineal.
- Usando de forma eficiente los datos, se realiza inferencia sobre los parámetros condicionada a que el modelo investigado sea apropiado.
- Dado un determinado proceso propuesto, se trata de cuantificar los parámetros del mismo, $\phi_1, \ldots, \phi_p, \ \theta_1, \ldots, \theta_q, \ \sigma^2$ y, en su caso, δ.

3) Validación

- La validación del modelo proporciona un ajuste adecuado y que los supuestos básicos, implícitos en el modelo, se satisfacen; de no cumplirse los supuestos, se determinan las modificaciones necesarias y, de hecho, se repiten las etapas anteriores hasta que la verificación indique resultados aceptables.

4) Predicción

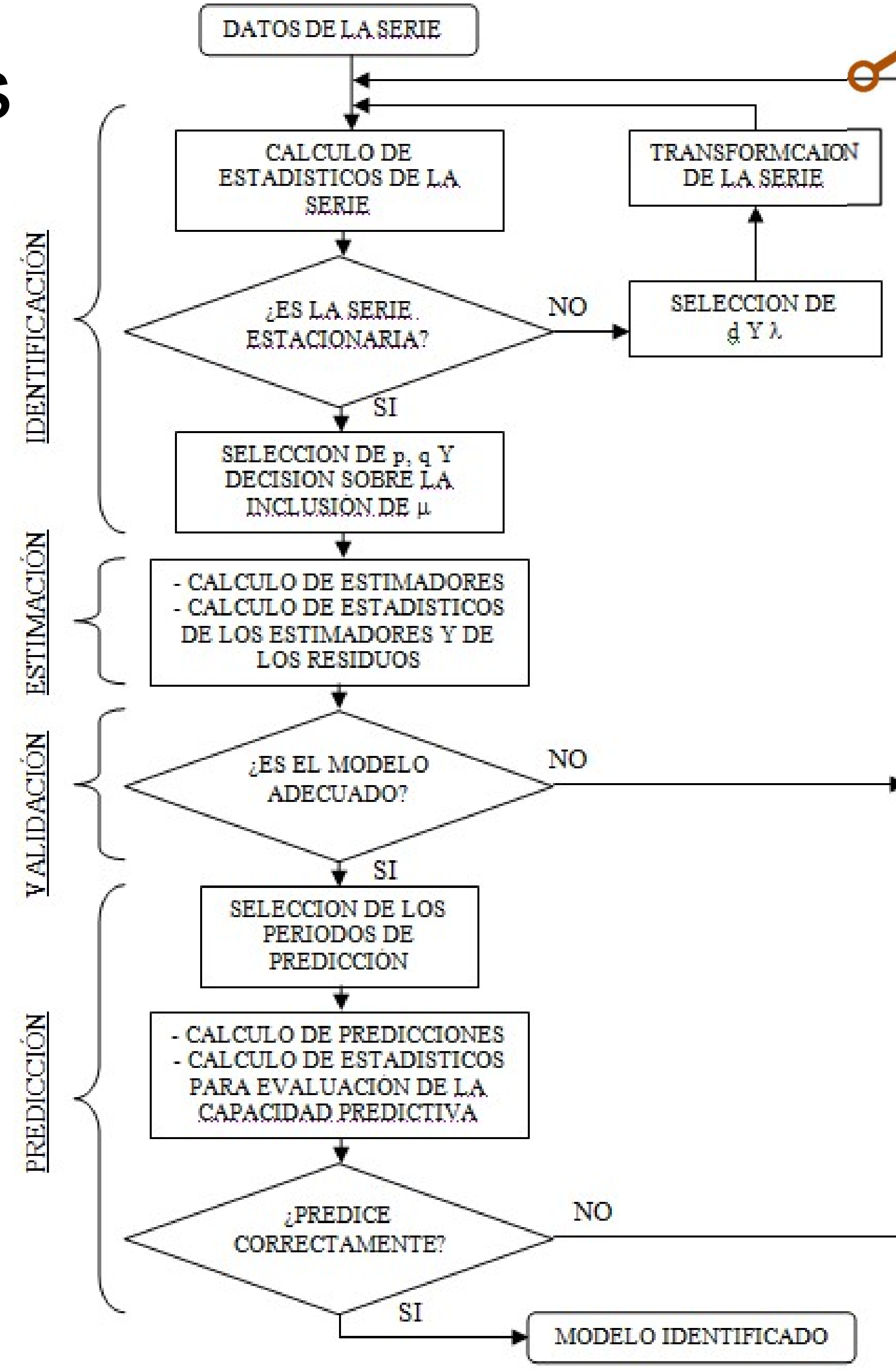
- Uso del modelo para los fines que el investigador haya tenido en mente al construirlo; dichos fines son por lo general de pronóstico, control, simulación o explicación del fenómeno en estudio.



En sí, la construcción del modelo de acuerdo con la metodología de Box-Jenkins es un proceso iterativo que se puede representar gráficamente mediante el esquema de la siguiente figura.

Proceso iterativo de Box-Jenkins

George E. P. Box　　Gwilym M. Jenkins



- Ahora, se presenta con cierto detalle las etapas de identificación, estimación y validación mencionadas anteriormente; la cuarta etapa (predicción) está por lo regular íntimamente ligada con la obtención de pronósticos a partir del modelo y dicho tema será abordado más adelante, en el siguiente capítulo.

Ejemplo 1

Se tiene una serie clásica en la literatura con datos anuales de la cantidad de Pieles de lince (Yt) captadas cada año, en el periodo 1821 a 1930. Los datos se muestran en la Tabla 6.1.

Tabla 6.1
Cantidad de Pieles de Lince captadas por año, periodo 1821 – 1930.

Año	Y_t	Año	Y_t	Año	Y_t	Año	Y_t	Año	Y_t	Año	Y_t
1821	269	1840	409	1859	684	1878	299	1897	587	1916	674
1822	321	1841	151	1860	299	1879	201	1898	105	1917	81
1823	585	1842	45	1861	236	1880	229	1899	387	1918	80
1824	871	1843	68	1862	245	1881	469	1900	758	1919	108
1825	1475	1844	213	1863	552	1882	736	1901	1307	1920	229
1826	2821	1845	546	1864	1623	1883	2042	1902	3465	1921	1139
1827	3928	1846	1033	1865	3311	1884	2811	1903	6991	1922	2432
1828	5943	1847	2129	1866	6721	1885	4431	1904	6313	1923	3574
1829	4950	1848	2536	1867	4254	1886	2511	1905	3794	1924	2935
1830	2577	1849	957	1868	687	1887	389	1906	1836	1925	1537
1831	523	1850	361	1869	255	1888	73	1907	345	1926	529
1832	98	1851	377	1870	473	1889	39	1908	382	1927	485
1833	184	1852	225	1871	358	1890	49	1909	808	1928	662
1834	279	1853	360	1872	784	1891	59	1910	1388	1929	1000
1835	409	1854	731	1873	1594	1892	188	1911	2713	1930	1590
1836	2285	1855	1638	1874	1676	1893	377	1912	3800		
1837	2685	1856	2725	1875	2251	1894	1292	1913	3091		
1838	3409	1857	2871	1876	1426	1895	4031	1914	2985		
1839	1824	1858	2119	1877	756	1896	3495	1915	790		

Ejemplo 1

- **Librería e importación de datos**

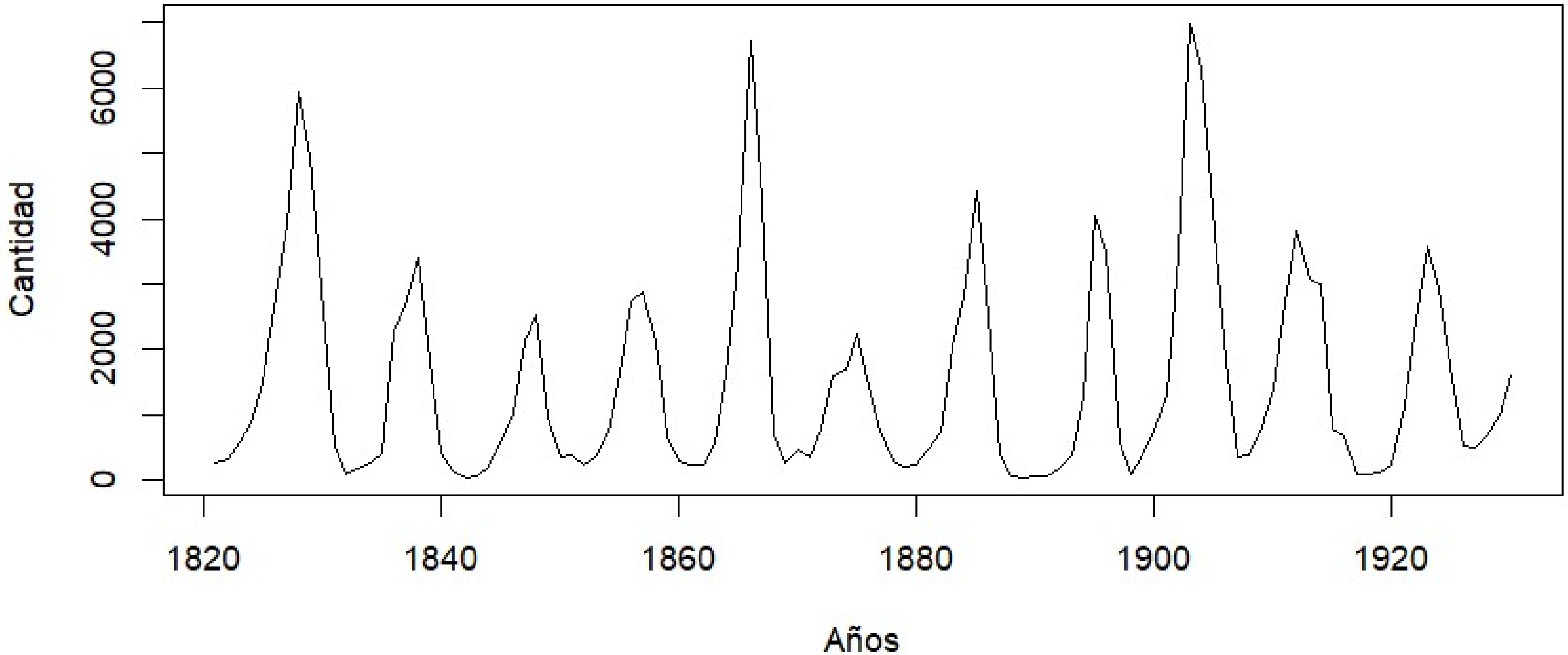
```
# Libreria necesaria
library(forecast) # Modelo ARIMA
library(tseries)  # Para series de tiempo
library(TSA)      # Para series de tiempo
library(urca)     # Para hacer el Test de Raiz Unitaria (detectar hay o no estacionariedad)
library(ggplot2)  # Para hacer gráficos
library(dplyr)    # Para la manipulación de datos (filtrar, seleccionar, agregar, transformar)
library(lmtest)   # Inferencia para coeficientes estimados
library(MASS)     # Transformacion de Box-Cox
library(nortest)  # Pruebas de normalidad
library(strucchange) # Cambio estructural - Test de Chow

#Importar datos
library(readxl)
Combustible <- read_excel("D:/.../Ejm_6_1-Pieles.xlsx")
View(Pieles)
```

Gráfica de la serie

```
Yt <- ts(Pieles$Y, start = 1821, frequency = 1)
plot(Yt, xlab="Años", ylab="Cantidad")
```

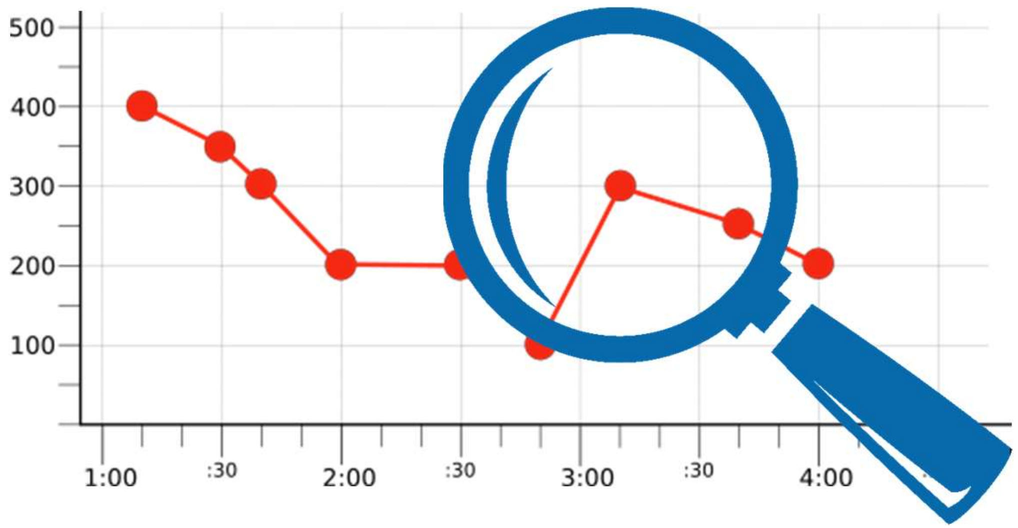
Ejemplo 1

Pieles de lince (Yt) 1821-1930.



Ahora realizaremos el proceso detallado de la construcción de modelos ARIMA apropiados para la serie Yt basados en la metodología de Box-Jenkins.

1) IDENTIFICACIÓN



- El objetivo de esta etapa de la modelización es seleccionar el modelo ARIMA(p,d,q) apropiado para la serie.
- La identificación del modelo se lleva a cabo en dos fases:

A) Análisis de estacionariedad

Estacionariedad en Varianza

Transformaciones estabilizadoras de varianza.

Estacionariedad en media

Número de diferencias d que hay que tomar para lograr que la serie sea estacionaria en media.

1) IDENTIFICACIÓN

B) Elección de los órdenes *p* y *q*

- Una vez obtenida la serie estacionaria, el objetivo es determinar el proceso estacionario ARMA(p, q) que la haya generado.

- Previo a desarrollar estas dos fases, es importante realizar el análisis de la tendencia y la estacionalidad.

- Los instrumentos que se van a utilizar en estas dos fases de la identificación del modelo son fundamentalmente los siguientes:

  ✓ Gráfico y correlogramas muestrales de la serie original.

  ✓ Gráfico y correlogramas muestrales de determinadas transformaciones de la serie: logaritmos, diferencias, etc.

  ✓ Contrastes de raíces unitarias.

Ejemplo 1

IDENTIFICACIÓN

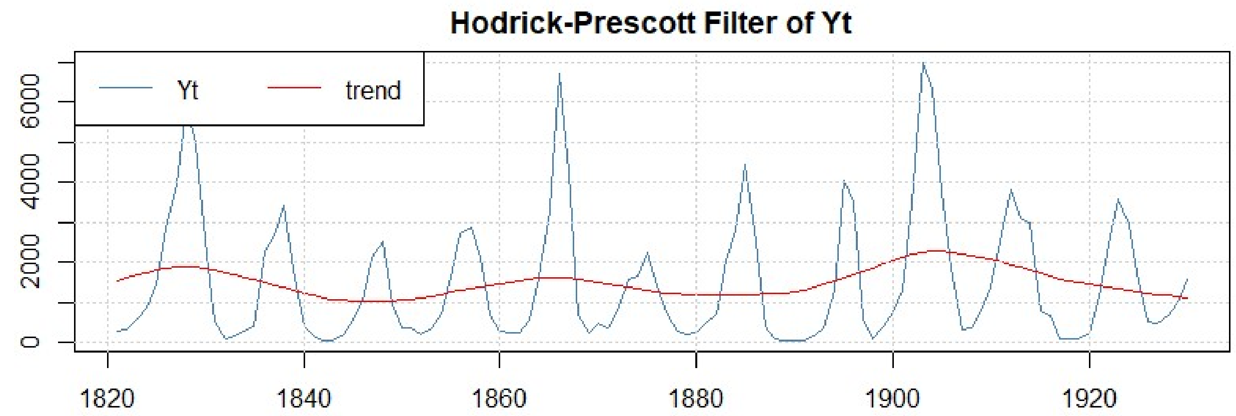
DESCOMPOSICIÓN DE LA SERIE.

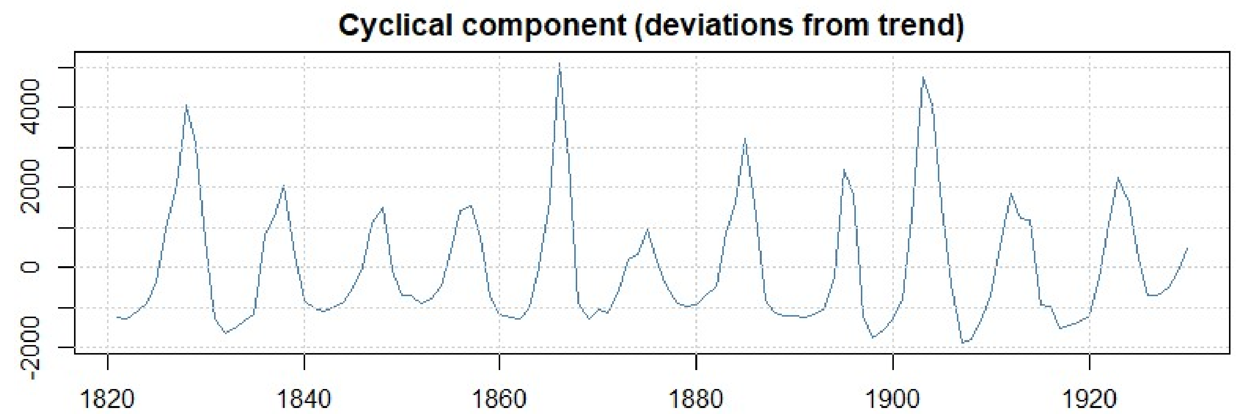
- ✓ Se tiene serie anual y no presenta periodicidad, por tanto, no se realiza la descomposición conjunta.
- ✓ Realizar de manera individual.

- Análisis de la tendencia
  Aplicando Filtro de Hodrick – Prescott

```
lambda_hp <- 1000
Yhp <- hpfilter(Yt, type="lambda",
freq=lambda_hp)plot(Yhp))
```

Muestra la evolución temporal de la tendencia, en la cual se puede detectar un comportamiento cíclico con periodo aproximado de 40 años.





- Análisis de la estacionalidad
  Se tiene una serie anual y corresponde a series con periodicidad menor a un año.

1.1. Análisis de Estacionariedad

a) Estacionariedad en Varianza

- Una serie será estacionaria en varianza cuando la variabilidad de la serie en torno a su media se mantenga constante a lo largo del tiempo.

- Si la serie no es estacionaria en varianza, se utilizan las transformaciones estabilizadoras de varianza, es decir, las transformaciones Box-Cox.

Transformación de Box-Cox

- La transformación potencia Box-Cox es una familia de transformaciones que inducen linealidad, homocedastidad (varianza estable) y normalidad.

- Las transformaciones más habituales se describen a continuación:

$$Y_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda} = \begin{cases} \frac{1}{Y_t^2} & \lambda = -2 \\ \frac{1}{Y_t} & \lambda = -1 \\ \frac{1}{\sqrt{Y_t}} & \lambda = -0.5 \\ \log(Y_t) & \lambda = 0 \\ \sqrt{Y_t} & \lambda = 0.5 \\ Y_t & \lambda = 1 \\ Y_t^2 & \lambda = 2 \end{cases}$$

siendo Y_t la variable a ser transformada y λ el parámetro de transformación.

1.1. Análisis de Estacionariedad

Identificación de λ

- La diferencia de un determinado orden es suficiente en muchos casos para obtener series estacionarias en media y varianza.
- En series económicas que se extienden a lo largo de un período dilatado de tiempo y que están afectados por una fuerte tendencia, suele ser necesario efectuar además alguna transformación instantánea del tipo Box - Cox, para obtener una serie estacionaria en varianza y que al mismo tiempo tenga una distribución normal.
- Este tipo de transformación quedará definido mediante el parámetro λ. Cuando:

$\lambda = 1$	$\Rightarrow$	no se transforma la serie.
$\lambda = 0$	$\Rightarrow$	se aplica logaritmos a la serie.
$0 < \lambda < 1$	$\Rightarrow$	se eleva la serie a un exponente fraccionario.

- En la práctica, las alternativas que se utilizan son las dos primeras.

1.1. Análisis de Estacionariedad

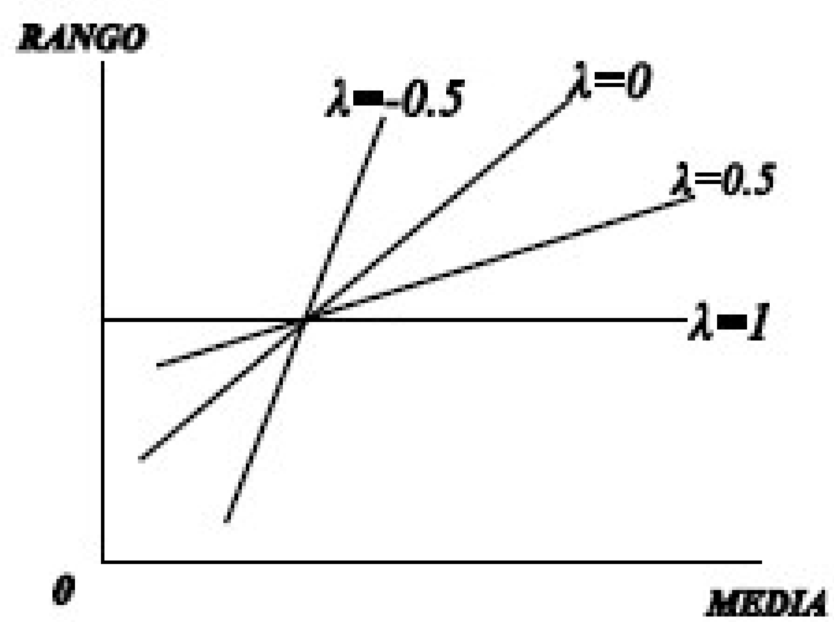
- Los instrumentos que usan para detectar si la serie es estacionaria o no en varianza son:

i) Ploteo de la serie

Es relativamente frecuente que a medida que aumenta el nivel, aumente también la dispersión, siendo aconsejable en ese caso tomar logaritmos neperianos, es decir, la transformación $\lambda = 0$.

ii) Gráfico Rango - Media.

La siguiente figura muestra la relación rango - media y tiene las alternativas siguientes:



Ejemplo 1

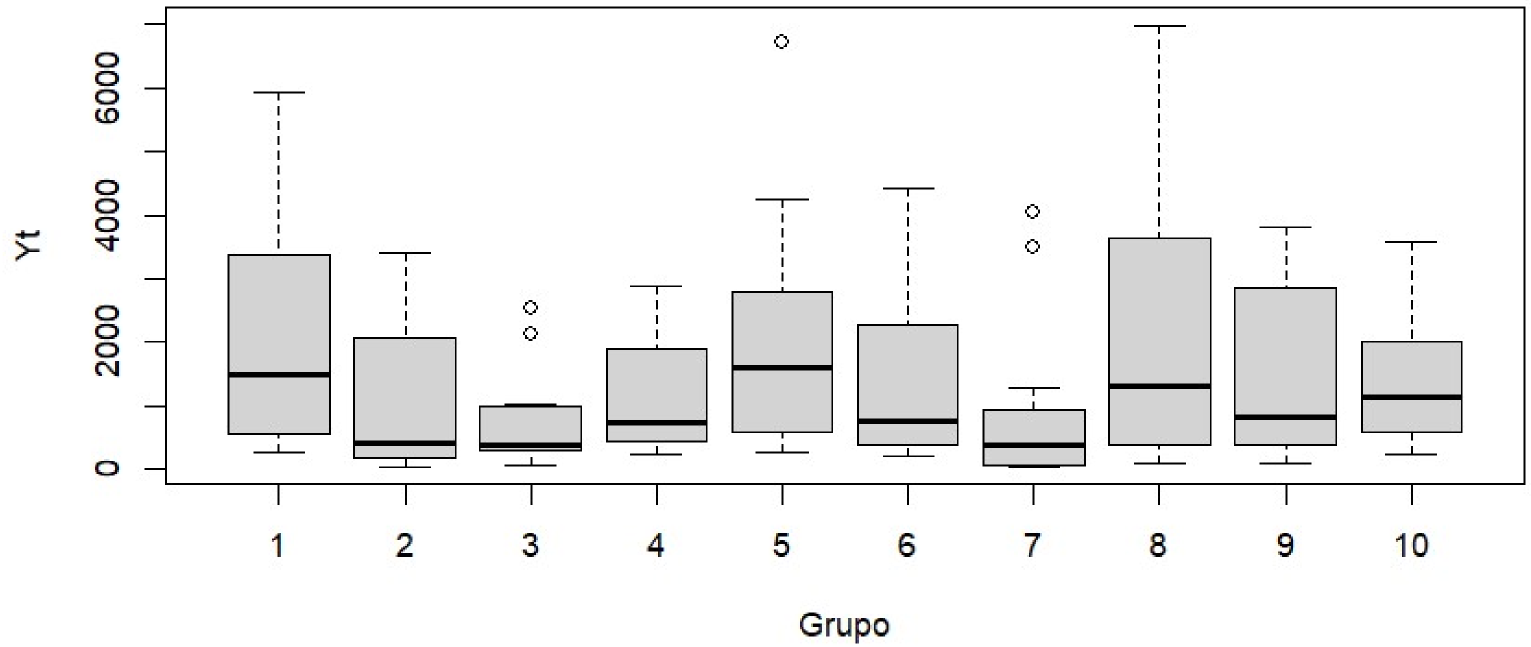
Estacionariedad en Varianza

Realizaremos un análisis gráfico a través de diagrama de cajas por grupos, de un total de 110 datos, generamos 10 grupos de 11.

```
Grupo <- c(rep(1:10,rep(11,10)))
boxplot(Pieles$Yt ~ Grupo, xlab = "Grupo", ylab="Yt")
```

En la figura se observa que, los grupos claramente presentan distinta variación, entonces hay indicios que la serie no sea estacionaria en varianza.

Por tanto, con el propósito de obtener la estacionariedad en varianza aplicamos la metodología de Box-Cox.



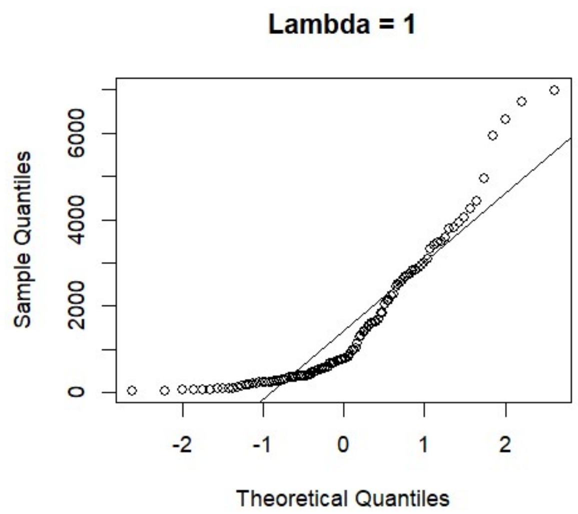
Ejemplo 1

Solo con fines explicativos, se muestran las dos formas de elegir el parámetro de lambda apropiado para la transformación de la serie.

i) Transformación de la serie para elegir el valor apropiado del parámetro λ.

Para $\lambda = 1$, es la serie original Y

```
qqnorm(Yt, main = "Lambda = 1") # Yt: Original
qqline(Yt)
```

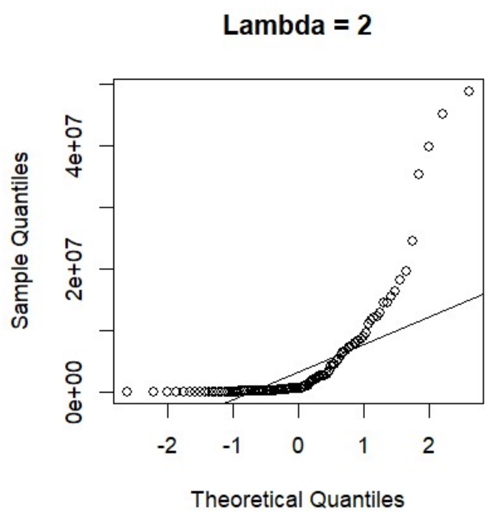


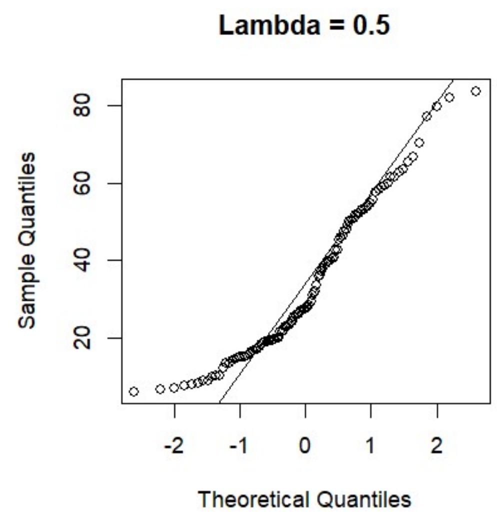
Para $\lambda = 2$

```
T1.Yt <- Yt^2
qqnorm(T1.Yt, main = "Lambda = 2")
qqline(T1.Yt
```

Para $\lambda = 0.5$

```
T3.Yt <- sqrt(Yt)
qqnorm(T3.Yt, main = "Lambda = 0.5")
qqline(T3.Yt
```





Ejemplo 1

Para $\lambda = 0$

```
T4.Yt <- log(Yt)
qqnorm(T4.Yt, main = "Lambda = 0")
qqline(T4.Yt)
```

Para $\lambda = -0.5$

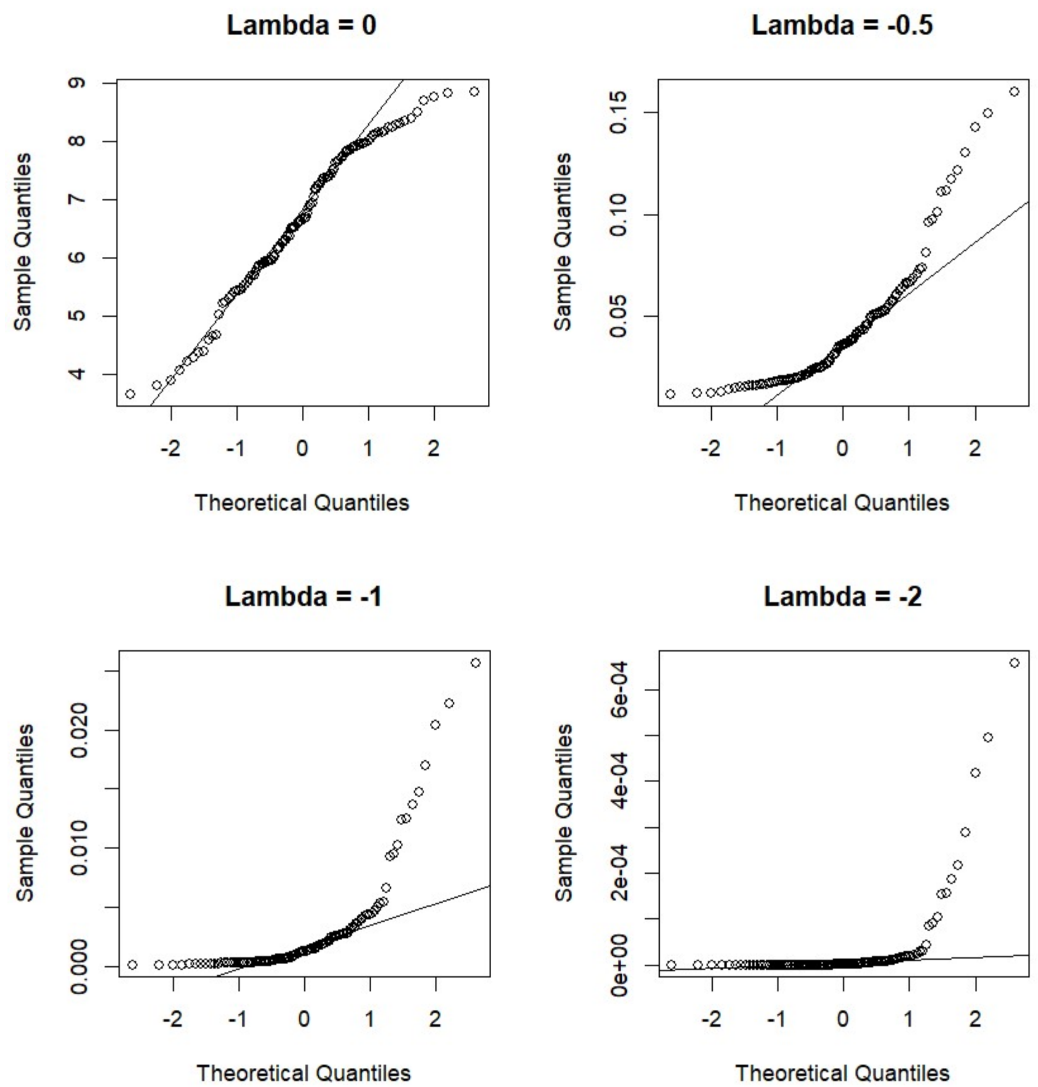
```
T5.Yt <- 1/sqrt(Yt)
qqnorm(T5.Yt, main = "Lambda = -0.5")
qqline(T5.Yt)
```

Para $\lambda = -1$

```
T6.Yt <- 1/Yt
qqnorm(T6.Yt, main = "Lambda = -1")
qqline(T6.Yt)
```

Para $\lambda = -2$

```
T7.Yt <- 1/(Yt^2)
qqnorm(T7.Yt, main = "Lambda = -2")
qqline(T7.Yt)
```



Se observa el ajuste a la distribución normal de la serie original Yt y en las demás figuras se tiene todas las demás transformaciones. La serie transformada que presenta mejor ajuste a la distribución normal es la serie T4.Yt = log(Yt) que se logra con $\lambda = 0$, por lo cual una posible transformación para estabilizar aproximadamente la varianza de Y es log(Yt).

Ejemplo 1

ii) Aproximación directa del valor de lambda optimo para realizar la transformación.

En R, podemos hacer uso de la función **boxcox** de la librería **MASS** para estimar el parámetro de transformación λ.

```
b = BoxCox.ar(Yt)
```

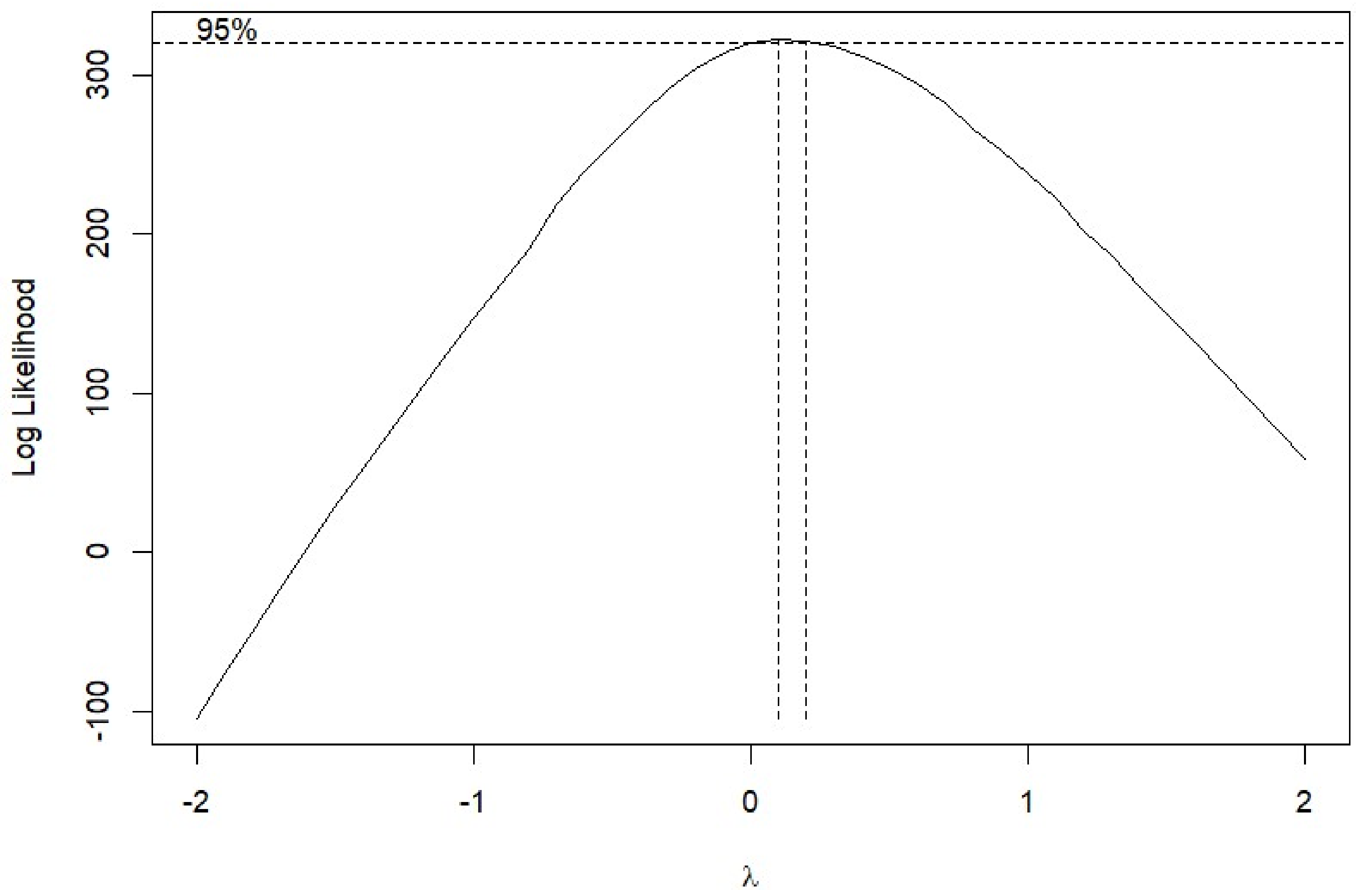
Valor estimado de λ.

```
lambda <- b$mle
round(lambda,1)
```

```
[1] 0.1
```

Por lo tanto, debemos hacer la siguiente transformación para buscar conseguir una aproximación a la normalidad.

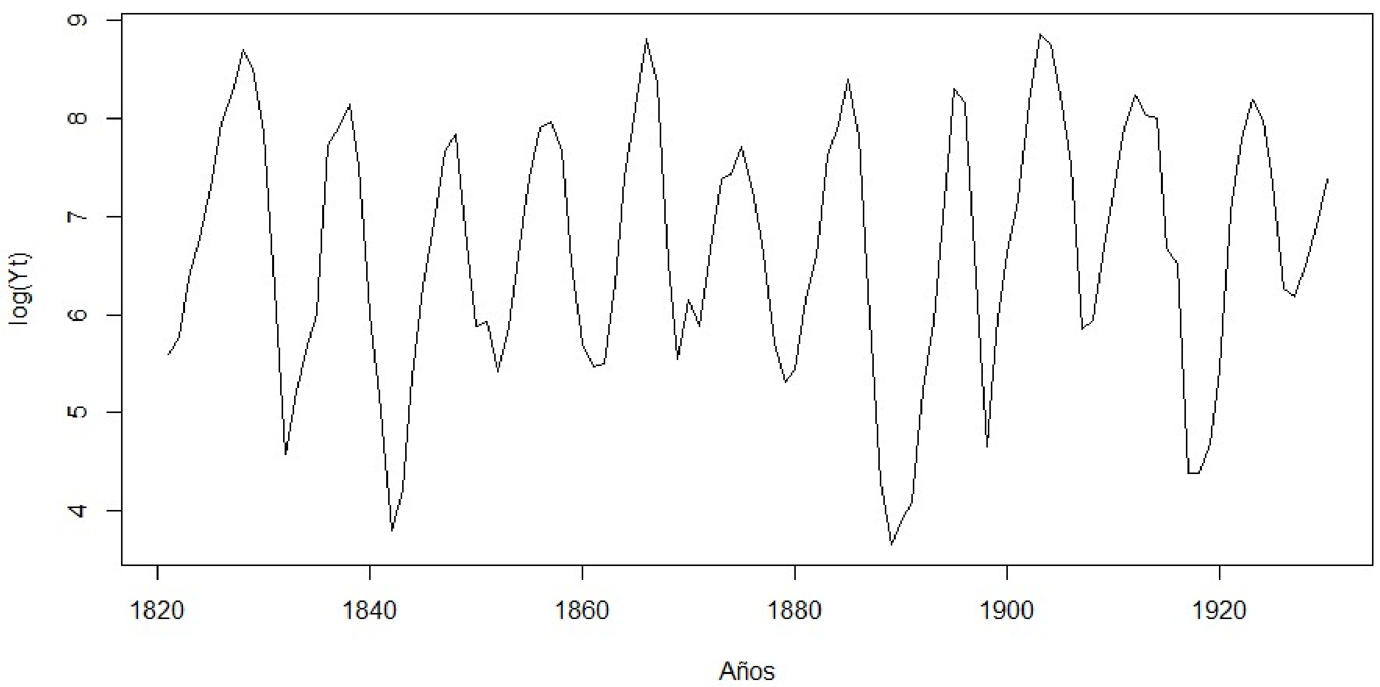
$$Y_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda} = \log(Y_t), \quad \text{para} \quad \lambda = 0$$



Ejemplo 1

Transformación de la serie Yt

```
T.Yt <- log(Yt)
plot(T.Yt, xlab="Años", ylab="Log(Yt)")
```



Ahora tomando logaritmos a la serie (la transformación Box-Cox más utilizada para las series económicas) se observa que la amplitud de estos ciclos se ha homogeneizado y que la variabilidad de la serie es mucho más estable en el tiempo como se muestra en la figura. Se puede concluir que la serie T.Yt = Log(Yt) es ESTACIONARIA EN VARIANZA.

1.1. Análisis de Estacionariedad

b) Estacionariedad en Media

- Como se ha visto en el capítulo 5, si la serie no es estacionaria en media se puede lograr la estacionariedad transformándola tomando diferencias.

$$Z_t = \Delta^d Y_t$$

- Sabemos que para las series económicas los valores de *d* más habituales son $d = 0, 1, 2$ y para decidir cuál es el más apropiado para la serie bajo estudio utilizaremos los siguientes instrumentos:

i. Gráfico de la serie original y las transformaciones correspondientes, para observar si se cumple o no la condición de estacionariedad de oscilar en torno a un nivel constante.

ii. Correlograma estimado de la serie original y de las transformaciones correspondientes, para comprobar si decrece rápidamente hacia cero o no.

iii. Contrastes de raíces unitarias. Estas se pueden realizar a través de la prueba de Dickey-Fuller.

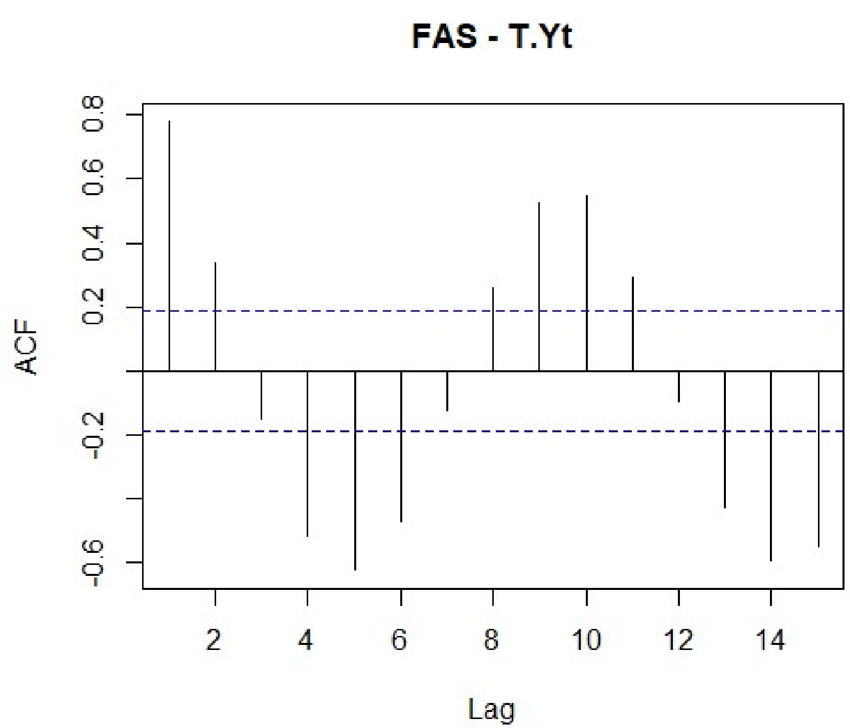
Ejemplo 1

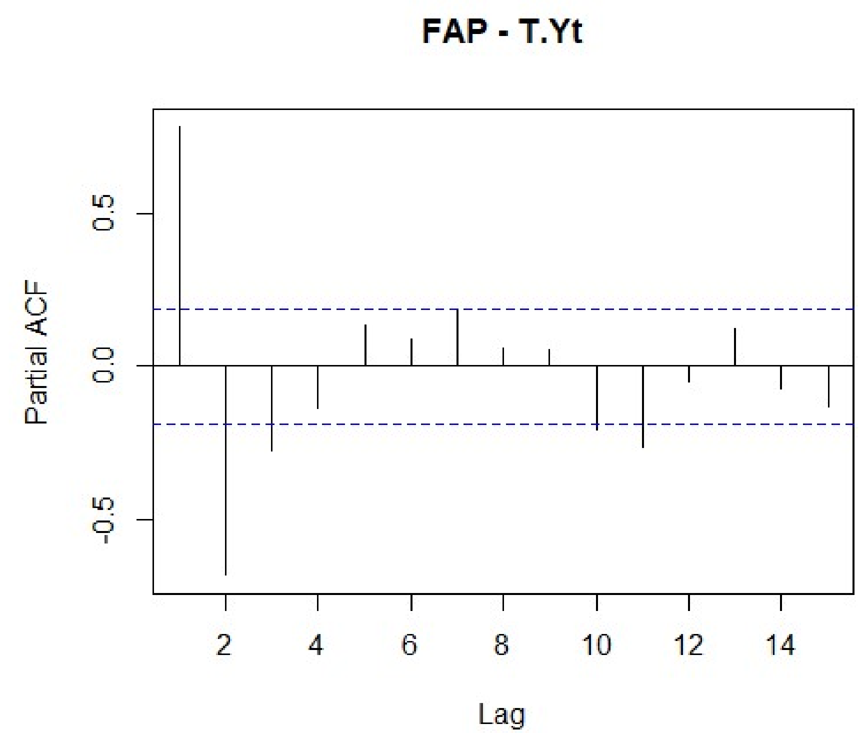
Estacionariedad en Media

Correlogramas del FAS y FAP de la serie transformada T.Yt = log(Yt)

```
FAS <- acf(T.Yt, lag.max = 15, main="FAS - T.Yt", level = 0.95)
FAP <- pacf(T.Yt, lag.max = 15, main="FAP - T.Yt", level = 0.95)
```

Los coeficientes de autocorrelación simple (FAS) decrecen hacia cero rápidamente de forma oscilatoria, además, el primer coeficiente de autocorrelación parcial (FAP) es menor a 0.9; por tanto, esta serie de pieles de lince presenta altos indicios de ser estacionaria.





Ejemplo 1

Verificación con la prueba de Raíz unitaria de Dickey-Fuller Aumentada.

```
TY_adf <- ur.df(T.Yt, type="drift", lags = 1)
summary(TY_adf)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################
Test regression drift
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5816 on 105 degrees of freedom
Multiple R-squared:  0.5402,      Adjusted R-squared:  0.5314
F-statistic: 61.68 on 2 and 105 DF,  p-value: < 2.2e-16

Value of test-statistic is: -7.984 31.8841

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

De los resultados, el valor de $\tau_{calculado}$ = -7.984 < $\tau_{crítico}$ = -2.88, por tanto, se acepta la hipótesis nula (Ho) de la existencia de raíz unitaria, es decir, que la serie T.Yt es ESTACIONARIA en media y varianza.

Por consiguiente, no requiere diferenciar la serie T.Yt.

1.2. Identificación del modelo estacionario

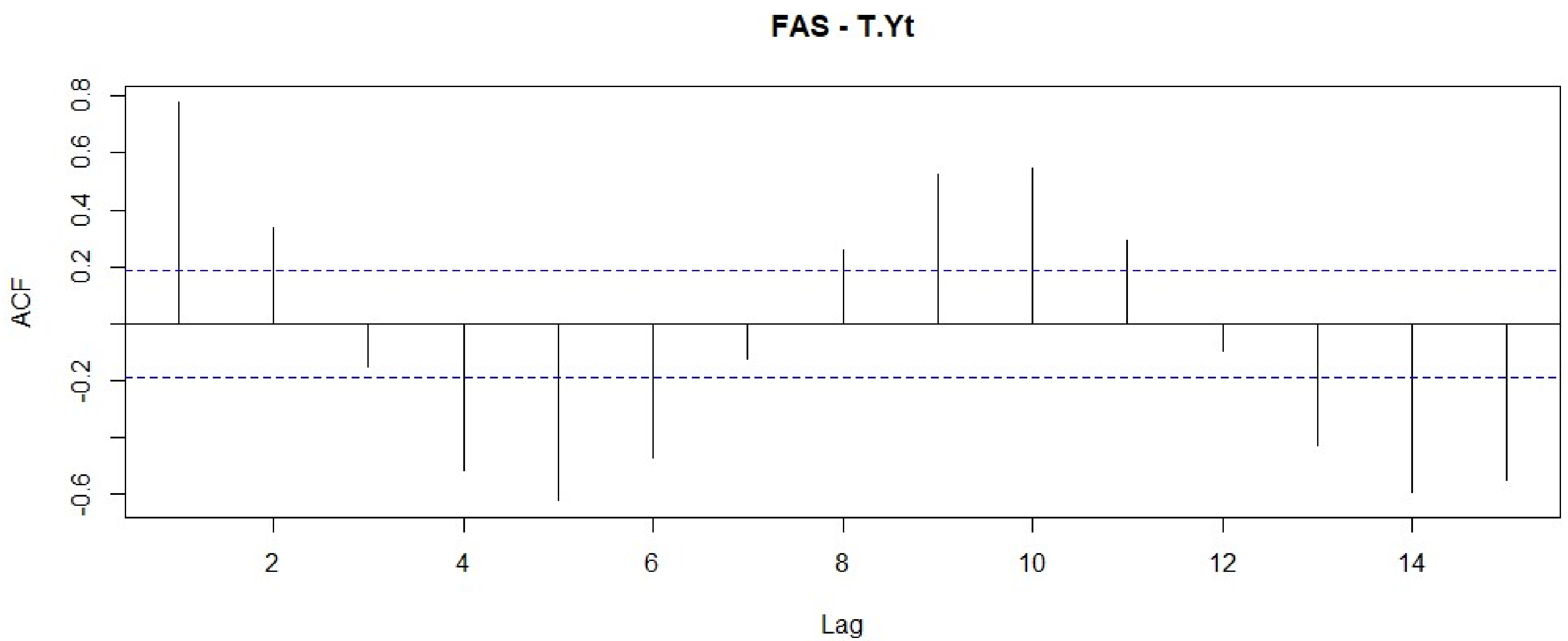
a) Identificación de las órdenes *p* y *q*

- Características básicas de las FAS y FAP teóricas.

Parámetros	FAS	FAP
AR(p)	Decrecimiento rápido de tipo exponencial, sinusoidal o mezcla de ambos tipos.	Los *p* primeros coeficientes son distintos de cero.
MA(q)	Los *q* primeros coeficientes son distintos de cero.	Decrecimiento rápido de tipo exponencial, sinusoidal.
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.

1.2. Identificación del modelo estacionario

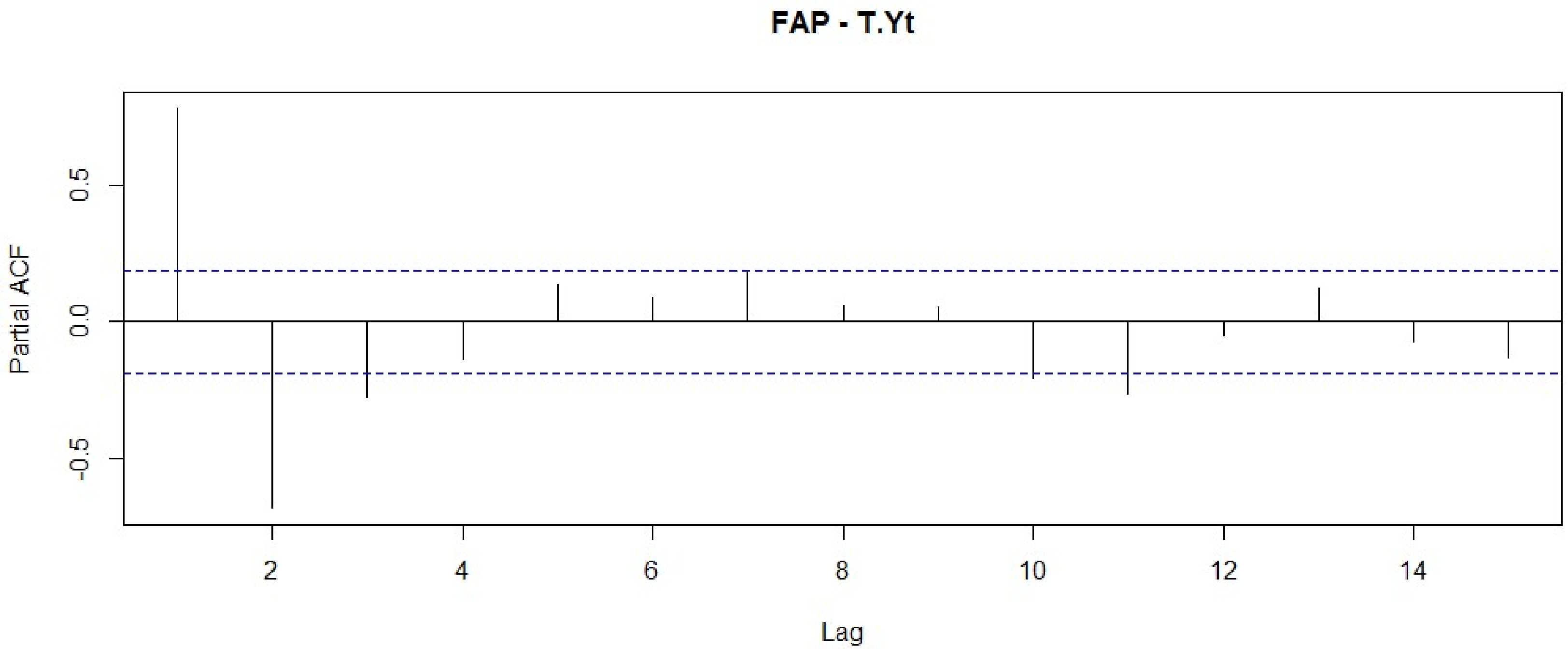
FAP para parte de Media Móvil (MA)



- La FAS estimada muestra una estructura infinita decreciendo en forma de onda seno-coseno amortiguada que sugiere un modelo ARMA(p, q) con parte autorregresiva de orden ≥ 2.

1.2. Identificación del modelo estacionario

FAS para parte Autorregresiva (AR)



- Los tres primeros coeficientes son significativamente distintos de cero y el resto no. Según esta interpretación, la FAP estaría truncada en el retardo 3 y el modelo adecuado sería un **AR(2)** y **AR(3)**..

- Modelos propuestos
  - a) log(Yt) ~ ARIMA(3, 0, 0)
  - b) log(Yt) ~ ARIMA(0, 0, 2)
  - c) log(Yt) ~ ARIMA(2, 0, 1)

1.2. Identificación del modelo estacionario

b) Inclusión del termino independiente (δ) o intercepto

- La media de un proceso ARMA(*p*, q) estacionario está directamente relacionada con la constante "δ". Si esta constante es cero, entonces la media del proceso es cero.

- Planteamiento de hipótesis.

H0: $E(Zt) = 0$ (Sin intercepto)
H1: $E(Zt) \neq 0$ (Con intercepto)

- El estadístico de contraste es:

$$t = \frac{\overline{Z}}{\hat{\sigma}_{\overline{z}}} \sim t_{\left(T^{*}-1\right)g.l.}$$

donde:

$T^{*} = T - d$ es la longitud de la serie estacionaria *Zt*.
T es el tamaño de la muestra.
$\hat{\sigma}_{\overline{z}}$ es el estimado de la varianza de la media $\overline{Z}$ que viene dado por:

$$\hat{\sigma}_{\overline{z}} = \frac{C_0}{T^{*}}\left(1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \cdots + 2\hat{\rho}_k\right)$$

1.2. Identificación del modelo estacionario

$$C_0 = \frac{\sum \left(Z_t - \bar{Z}\right)^2}{T-1}$$ es la varianza muestral de la serie estacionaria .

$$1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \cdots + 2\hat{\rho}_k$$ representan las k primeras autocorrelaciones muestrales significativas de Zt.

En ocasiones, para calcular esta varianza, se aplica la siguiente aproximación: $\hat{\sigma}_{\bar{z}} = \frac{C_0}{T^*}$

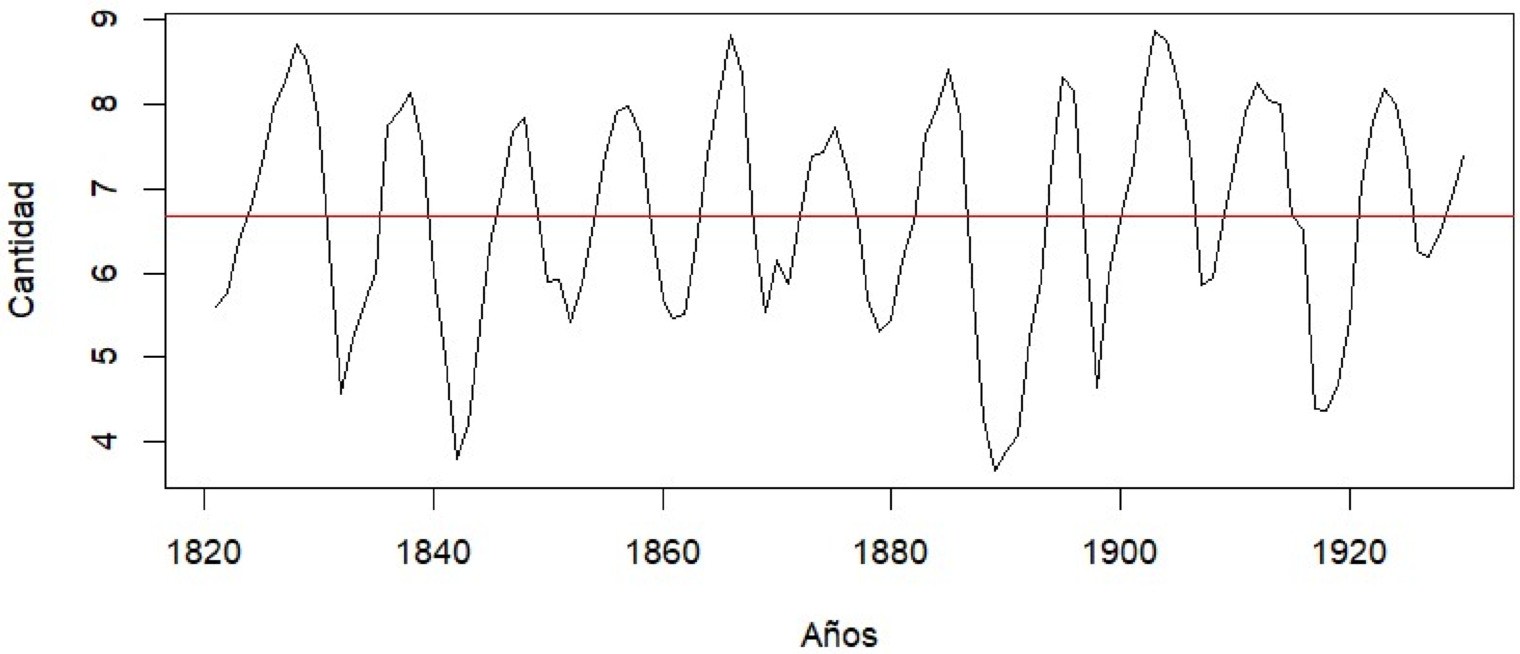
Se rechazará la H_0: $E(Zt) = 0$ a un nivel de significación α y, por lo tanto, se incluirá el parámetro δ en el modelo si:

$$t > t_{\alpha/2}\left(T^* - 1\right) g.l.$$

Ejemplo 1

Determinamos la inclusión del término constante, haremos un análisis gráfico para ver entorno a que valor oscilan las variaciones.

```
plot(T.Yt, xlab="Años", ylab="Cantidad")
abline(h = mean(T.Yt), col = "red")
```



La figura muestra la serie estacionaria log(Yt) esta parece oscilar en torno a una media diferente de cero. Para confirmar realizaremos la prueba estadística para contrastar si la media diferente de cero.

Ejemplo 1

Prueba estadística.

```
Z <- mean(T.Yt)
Co <- var(T.Yt)
Tn <- length(Yt)
Ta <- Tn - 1
Sigma <- Co/Ta
t <- Z/Sigma
tt <- qt(1-0.05/2,Ta-1)
pruebaT <- c(t, tt)
names(pruebaT) <- c("t-calculado","t-critico")
pruebaT
```

```
t-calculado   t-critico
 446.086795    1.982173
```

Se acepta la hipótesis alterna H_1: $E(Zt) \neq 0$ para un α del 5% y se debe incluir el término constante en el modelo.

Ejemplo 1

Resumiendo, se proponen los dos modelos siguientes para la serie estacionaria Zt = T.Yt = Log(Yt), en caso que fuesen:

- AR(3) ≅ ARIMA(3, 0, 0): $$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t$$
- MA(2) ≅ ARIMA(0, 0, 2): $$Z_t = \delta + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$
- ARMA(2,1) ≅ ARIMA(2, 0, 1): $$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t + \theta_1 a_{t-1}$$

GRACIAS

https://aulavirtual2.unap.edu.pe/