



## **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

PROYECTO: INPC

ALMUNOS:  
DEL VALLE LÓPEZ JOSÉ CARLOS  
GRANADOS ÁLVAREZ ALFONSO

PROFESOR:  
MARTHA ANGÉLICA MONTES FONSECA

AYUDANTE:  
HÉCTOR ÁLVAREZ OLMOS

FACULTAD DE CIENCIAS, UNAM

MÉXICO, D.F. DICIEMBRE 2016

# SERIES DE TIEMPO

## Resumen del proyecto

El siguiente proyecto es un análisis de las series de tiempo empleando el modelo clásico de descomposición, se pretende fragmentar nuestro proceso en una componente de tendencia, una componente estacional y un proceso de ruido blanco.

Nuestra serie de tiempo corresponde al INPC y trataremos de proyectarlo 12 meses a futuro.

Aunado a ello, aplicaremos nuestros conocimientos en la materia para realizar un análisis de los datos obtenidos.

Haremos uso de distintos softwares que nos ayudarán a resolver nuestro problema, es concreto: *R* y *R Studio*.

# Índice

<b>1. Introducción</b>	<b>4</b>
<b>2. Base de Datos</b>	<b>6</b>
2.1. ¿Qué información tenemos? . . . . .	6
2.2. Serie de tiempo en R . . . . .	6
<b>3. Explicación de los procedimientos realizados</b>	<b>8</b>
3.1. División de la serie . . . . .	8
3.2. Descomposición en componentes . . . . .	10
3.3. ACF Y PACF y estabilización . . . . .	10
<b>4. Eliminación de componentes de tendencia y estacionalidad</b>	<b>13</b>
<b>5. Ajuste del modelo y AIC</b>	<b>21</b>
5.1. Análisis de los residuales . . . . .	23
<b>6. Predicción</b>	<b>27</b>
<b>7. Conclusiones</b>	<b>29</b>

## 1. Introducción

El siguiente trabajo expone la serie de tiempo del Índice Nacional de Precios al Consumidor (INPC).

Un Índice de Precios al Consumidor (IPC) es un indicador cuya finalidad es estimar la evolución de los precios de una canasta de bienes y servicios que consume una familia promedio. En México lo que conocemos como inflación es el crecimiento porcentual del Índice Nacional de Precios al Consumidor (INPC).

El Índice Nacional de Precios al Consumidor (INPC) es un indicador cuya finalidad es estimar la evolución de los precios de una canasta de bienes y servicios que consume una familia urbana promedio de México.

Con el INPC se abarcan todos los aspectos del consumo familiar mexicano en las áreas urbanas. Por tanto, en su elaboración se consideran:

- Los bienes y servicios que las familias compran.
- La proporción del gasto que en promedio las familias mexicanas destinan a los distintos rubros del consumo.
- Las principales zonas urbanas donde viven los consumidores.
- Los puntos de venta donde las familias mexicanas realizan sus compras.
- Las marcas, presentaciones y modalidades que de cada bien y servicio, prefieren los consumidores.

La revisión de la representatividad de las ponderaciones respecto al índice general de cada bien o servicio se realiza:

- Para el INPC, cuando una nueva Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) indique que el patrón de consumo se ha modificado significativamente.

El INPC es un indicador que debe actualizarse al ocurrir cambios significativos en la estructura del gasto de los hogares. La transformación de ésta revela que con el desarrollo, los hogares destinan un menor gasto al vestido y a la alimentación.

El INPC se calcula como un índice de ponderación fija. En el caso del INPC, básicamente es un cociente de gastos de la siguiente forma:

$$I_{t/0} = (\text{Costo de la "Canasta A" con los precios del periodo actual}) / (\text{Costo de la "Canasta A" con los precios del periodo base})$$

Canasta A:

- Conjunto de bienes y servicios
- Es la misma combinación de bienes y servicios en el periodo base (periodo 0) y en el periodo actual (periodo t).

El Banco de México inició la elaboración de índices de precios en 1929, con el índice de precios de alimentos de la Ciudad de México. Este primer índice se com-

ponía de 16 conceptos genéricos. El INPC comenzó a elaborarse en 1969. A partir de entonces este indicador se ha actualizado cuatro veces.

### **Etapas del Índice Nacional de Precios al Consumidor**

<i>Periodo Base</i>	<i>Vigencia</i>	<i>NCC</i>	<i>NPS</i>
1968	1969.01–1978.12	7	5,100
1978	1979.01–1981.12	16	7,100
1980	1982.01–1995.0	35	9,700
1994	1995.03–200.06	46	43,400
2ª Qna. Junio 2002	2002.07–2010.12	46	58,200
2ª Qna. Diciembre 2010	2010.12–	46	83,500

NCC=Número de ciudades de cotización

NPS=Número de productos y servicios específicos cotizados al comienzo de la utilización de la base

### **¿Qué nos dice el INPC?**

<i>Periodo</i>	<i>Índice</i>
2ª Qna. Junio 2002 Periodo Base (0)	100
2ª Qna. Enero 2010 Periodo Actual (t)	140.360

Significa que los precios de los bienes y servicios que consumen los mexicanos se incrementaron en promedio 40.36 % de la 2ª Qna. de Junio de 2002 a la 2ª Qna. de Enero de 2010.

## 2. Base de Datos

### 2.1. ¿Qué información tenemos?

Para el desarrollo de este proyecto originalmente se consiguió la serie histórica del INPC del año 1973 hasta 2016, obteniendo 527 observaciones.

Gracias a estos datos nos enfocaremos en realizar nuestras predicciones. Para obtener más información sobre nuestra base, haremos uso de R. El código se hallará en conjunto con los datos arrojados por el software.

Antes que nada debemos cargar las librerías que nos serán de utilidad para las series de tiempo.

```
library(descomponer)
library(timeSeries)
library(forecast)
library(nortest)
```

Primero cargamos la base.

```
seriecompleta<-read.table(choose.files(), header=T)
```

### 2.2. Serie de tiempo en R

Ahora toca convertir nuestra base en una serie de tiempo, sabemos que nuestros datos son registrados mensualmente y que históricamente se repite cada 12 meses, es decir, hay una frecuencia de 12 y colocamos como inicio enero de 1973.

```
sercomp.ts<-ts(seriecompleta["INPC"], start = 1973, frequency = 12)
```

Con el comando summary examinaremos la información básica de nuestros datos.

```
summary(sercomp.ts)
```

Min.	:	0.01896
1st Qu.	:	0.39763
Median	:	19.64038
Mean	:	41.64581
3rd Qu.	:	79.56838
Max.	:	121.00700

Si deseamos ver los promedios por periodo usaremos el siguiente comando:

```
for(i in 1:12){
mp<-mean(sercomp.ts[seq(i,length(sercomp.ts),12)])
print(mp)}
```

Enero	40.91545	Julio	41.85248
Febrero	41.15289	Agosto	42.0479
Marzo	41.40855	Septiembre	42.34319
Abril	41.5481	Octubre	42.60931
Mayo	41.52933	Noviembre	41.15595
Junio	41.67787	Diciembre	41.4941

Ya definida nuestra serie, echaremos un vistazo a nuestra serie de tiempo [3.1.1].

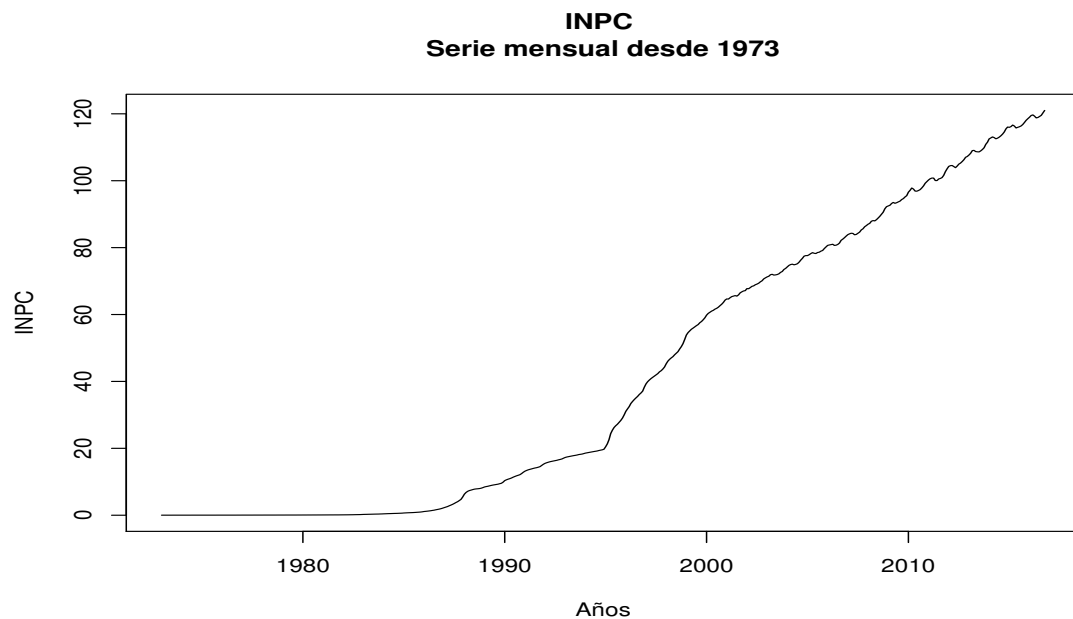


Figura 1: [3.1.1]

A partir de la información anterior, no podemos hacer una buena inferencia con base a la media ni a la mediana, pues los máximos de los últimos años son exponencialmente más grandes con respecto a los primeros, de lo que podemos destacar que existe una componente de tendencia, que podemos corroborar observando la gráfica.

Si observamos los promedios por periodo no nos dicen mucho sobre la información actual y debido a que el rango de información es muy grande empezando en 1973, es muy posible que sólo utilicemos los últimos datos para encontrar una buena predicción.

Debido a este corto análisis, acotaremos nuestra serie a partir de 1995, esto en vista de que la información inicial es muy diferente a como se comporta actualmente y por más razones que se explicarán en la siguiente sección, por lo que nuestra serie nueva será:

```
inpc = read.csv(choose.files(), header = T)
serie = ts(inpc["INPC"], start = 1995, frequency = 12)
```



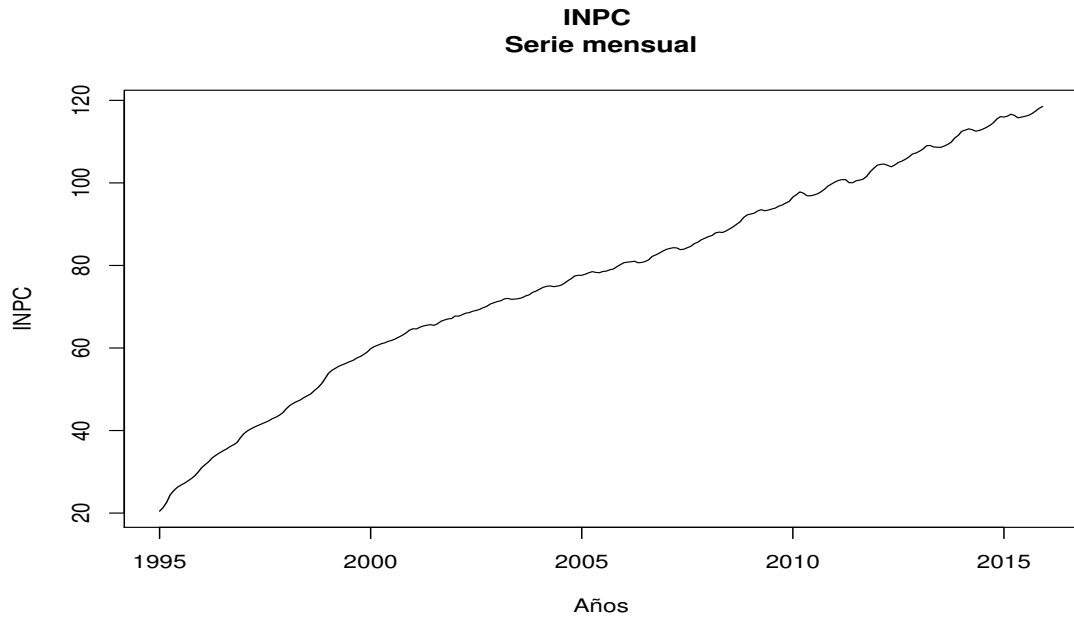


Figura 2: [3.1.1]

### 3. Explicación de los procedimientos realizados

#### 3.1. División de la serie

La primera acción que se realizó fue omitir la información previa al año 1995 por las siguientes razones, el comportamiento de la serie durante el periodo de 1973 a finales de 1988 prácticamente es constante y muy cercana a cero, posteriormente empieza a tener un crecimiento, pero observamos una especie de pico en 1995 y a partir de ahí adopta cierta tendencia creciente; las posibles razones por las que se presenta el comportamiento anterior es que en México se presentó una crisis a finales de 1994 conocida como “el error de diciembre” acompañada de una devaluación del peso, como se sabe las crisis afectan directamente la dinámica de la población en muchos aspectos y relacionado a lo mencionado al inicio del trabajo, dijimos que el índice se debe actualizar al ocurrir cambios significativos en la estructura del gasto de los hogares. Para 1995 entro en vigor el INPC tomando como año base al 1994, pero con una gran diferencia respecto al anterior ya que para el nuevo índice se consideraron 43,400 productos y servicios que cotizaban para éste, representando un crecimiento de más de 400 % en los productos considerados para el cálculo del mismo, y ese número se ha ido incrementando a través del tiempo. Por todo lo anterior se considerará la serie del INPC a partir del año 1995 hasta el 2015.

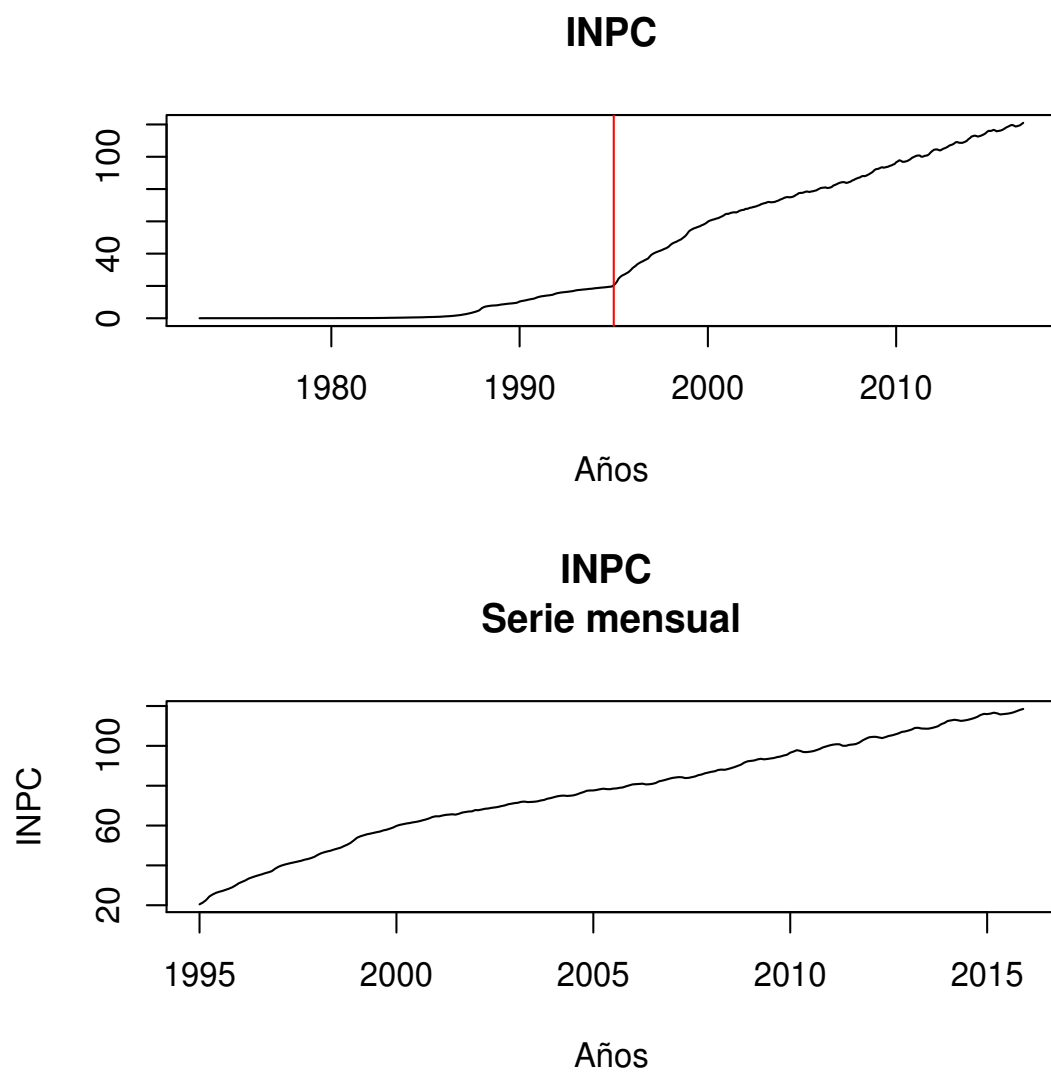


Figura 3: [3.1.1]

### 3.2. Descomposición en componentes

Nuestra serie de tiempo se compone de tres distintas componentes, una estacional, una tendencia y un proceso de ruido blanco o parte aleatoria. La forma de visualizar dentro de *R* cada una de las componentes es a través del comando:

```
plot(decompose(serie))
```

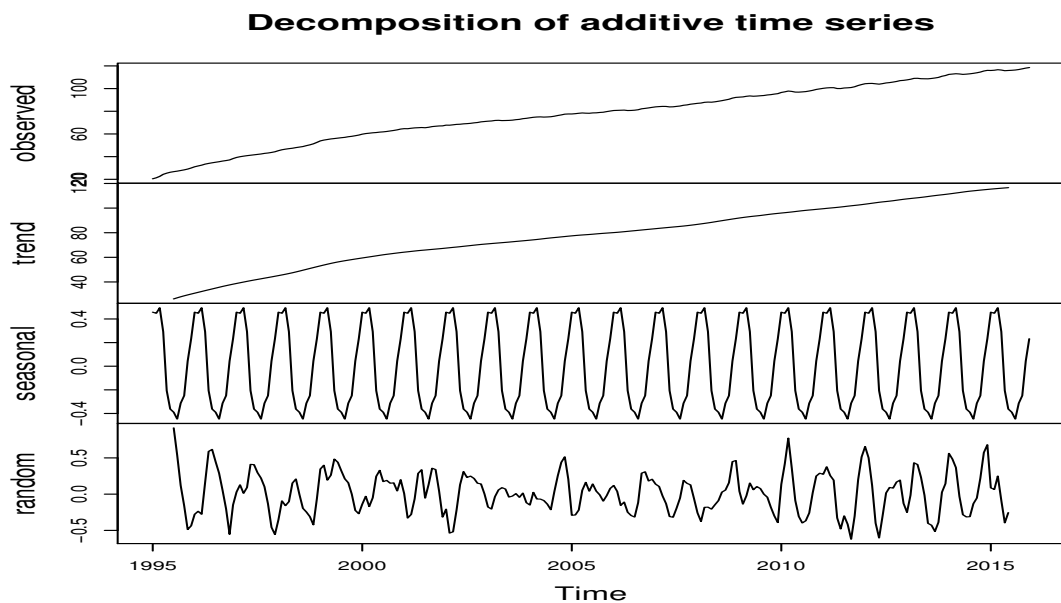


Figura 4: [3.2.1]

Lo que nos arroja la figura[3.2.1], donde *observed* es nuestra serie original, *seasonal* es nuestra componente estacional, *trend* nuestra componente de tendencia y *random* nuestra parte aleatoria.

Más adelante eliminaremos las componentes de tendencia y estacionalidad con el operador *diferencia*.

### 3.3. ACF Y PACF y estabilización

Dado que nuestra serie tiene componentes de tendencia y estacionalidad, podemos decir de entrada que se ajusta a un modelo  $ARMA(p,q)$  o  $ARIMA(p,d,q)$  estacional si empleamos diferencias y realizar el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), para determinar un modelo adecuado para la serie que sea útil para realizar un pronóstico.

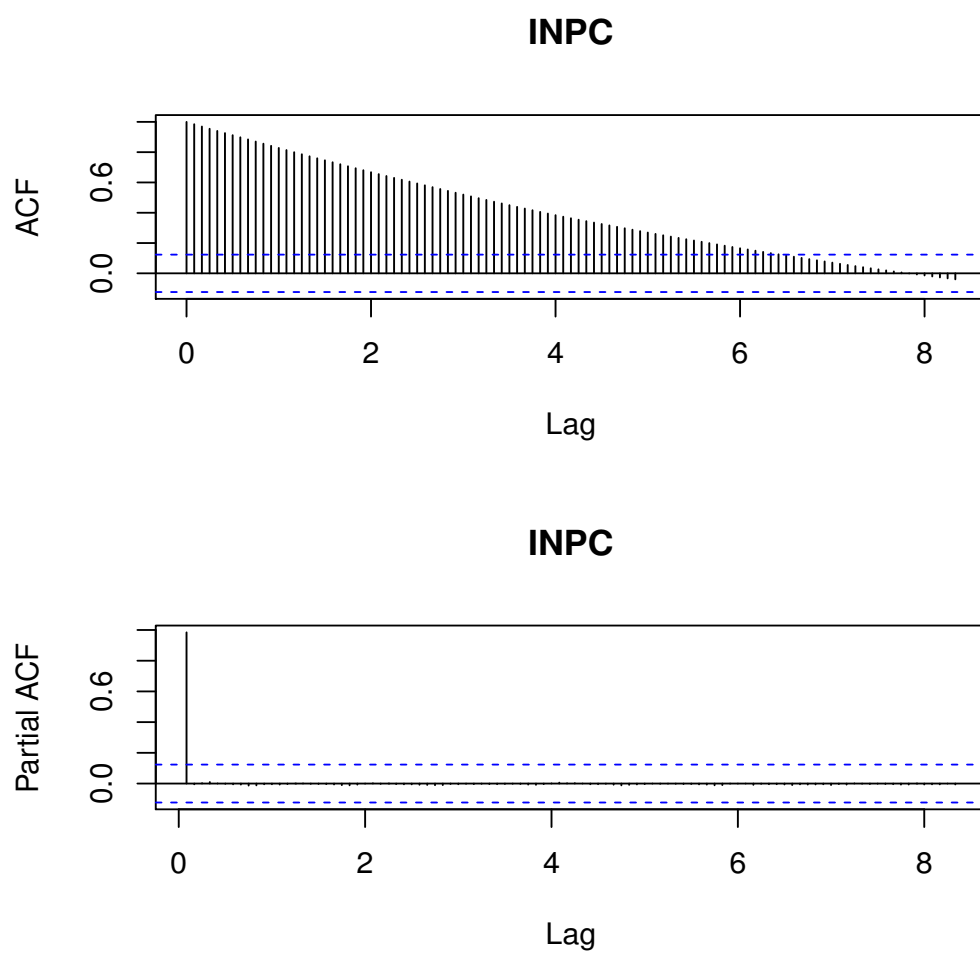


Figura 5: [3.3.1]

Para estabilizar la varianza de la serie aplicamos la función  $\log(x)$  que nos arroja este nuevo resultado.

```
> var(serie)
      INPC
INPC 651.6718
> var(x)
      INPC
INPC 0.1581474
```

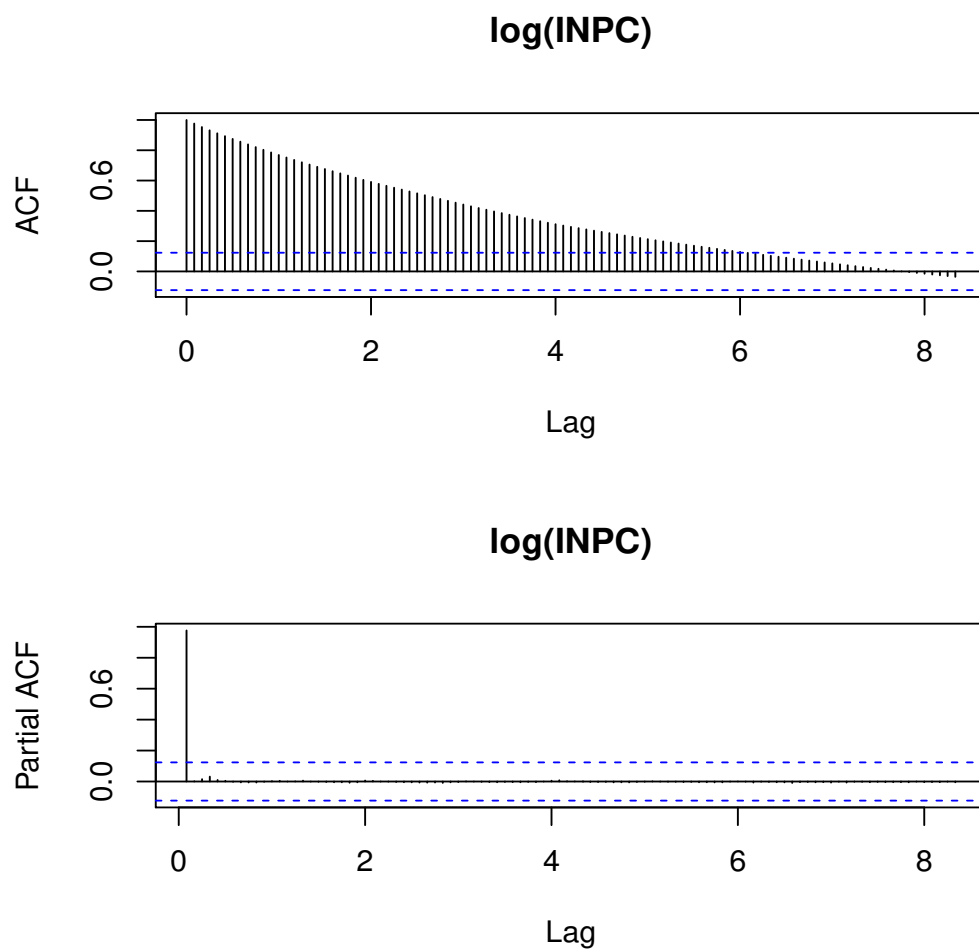


Figura 6: [3.3.2]

El comportamiento de las gráficas anteriores sugiere un comportamiento conocido, un ARIMA estacional, gracias a ello deseamos realizar una diferencia con el objetivo de eliminar por completo las componentes de tendencia y estacionalidad y determinar nuestro modelo.

## 4. Eliminación de componentes de tendencia y estacionalidad

Ya que nuestra serie pudo ser descompuesta en una componente estacional, una componente de tendencia y un proceso de ruido blanco, ahora se puede hacer uso de la diferenciación para deshacernos de dichas componentes.

Ahora quitaremos la componente estacional y de tendencia con el comando `diff`, en donde `lag` es el exponente o mejor dicho el número de retardo.

```
x = log(serie)
d1 = diff(x)
```

Veamos la diferencia entre nuestra nueva serie sin tendencia y sus respectivas ACF Y PACF:

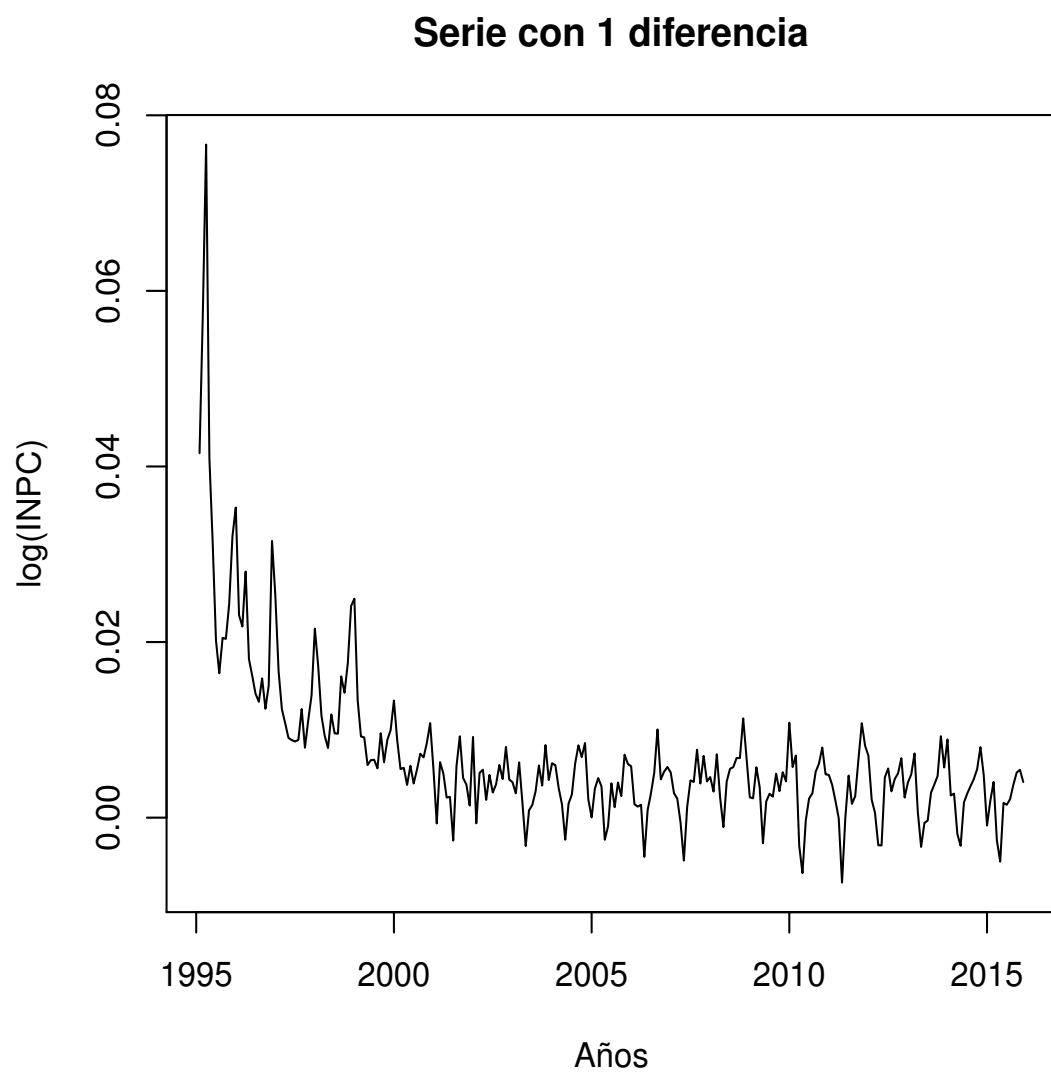


Figura 7: [3.1.1]

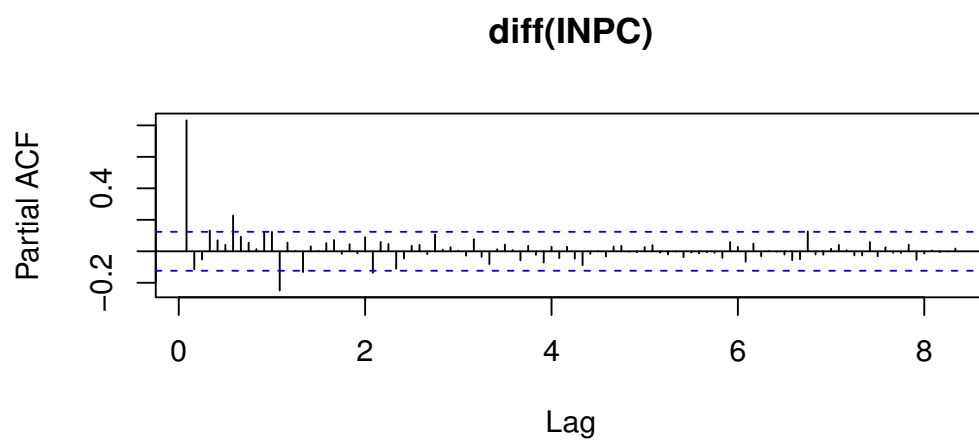
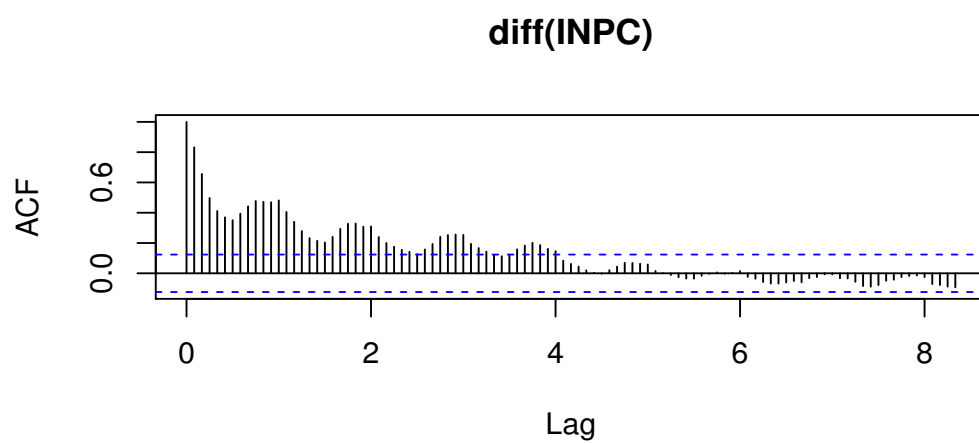


Figura 8: [3.3.2]



El comportamiento cosenoidal de la función de autocorrelación indica realizar otra diferencia, siendo el proceso análogo a cuando aplicamos la primera:

```
d2 = diff(d1)
```

### Serie con 2 diferencias

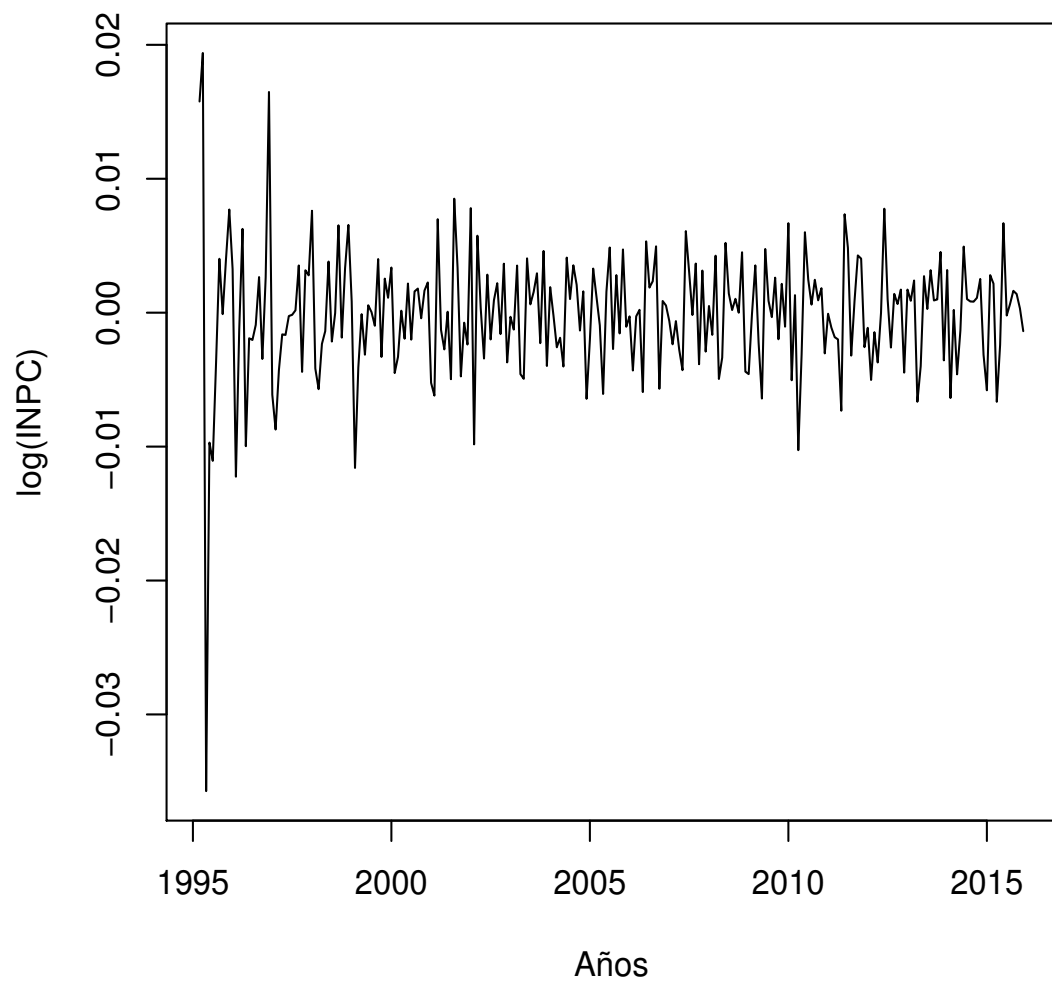


Figura 9: [3.1.1]

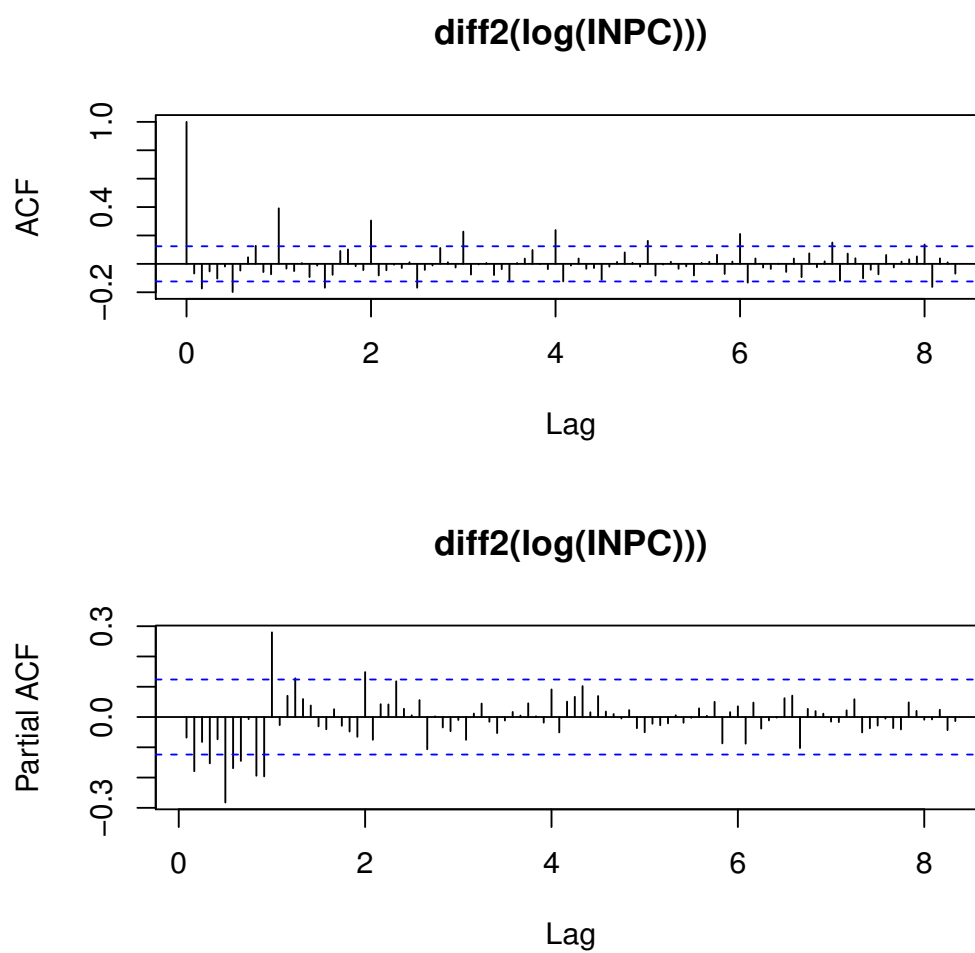


Figura 10: [3.3.2]

En la gráfica de la función de autocorrelación se pueden apreciar algunos retardos que salen de las bandas de confianza igualmente espaciados a lo largo del eje, lo que nos sugiere ajustar un modelo ARIMA, pero antes de iniciar al ajuste realizamos una diferencia estacional; notamos que dichos retardos salen de las bandas cada 6 unidades, pero las que más resaltan son las que salen cada 12 unidades. Al realizar la diferencia estacional cada 6 observaciones podemos observar que se conserva la estacionalidad como se observa a continuación:

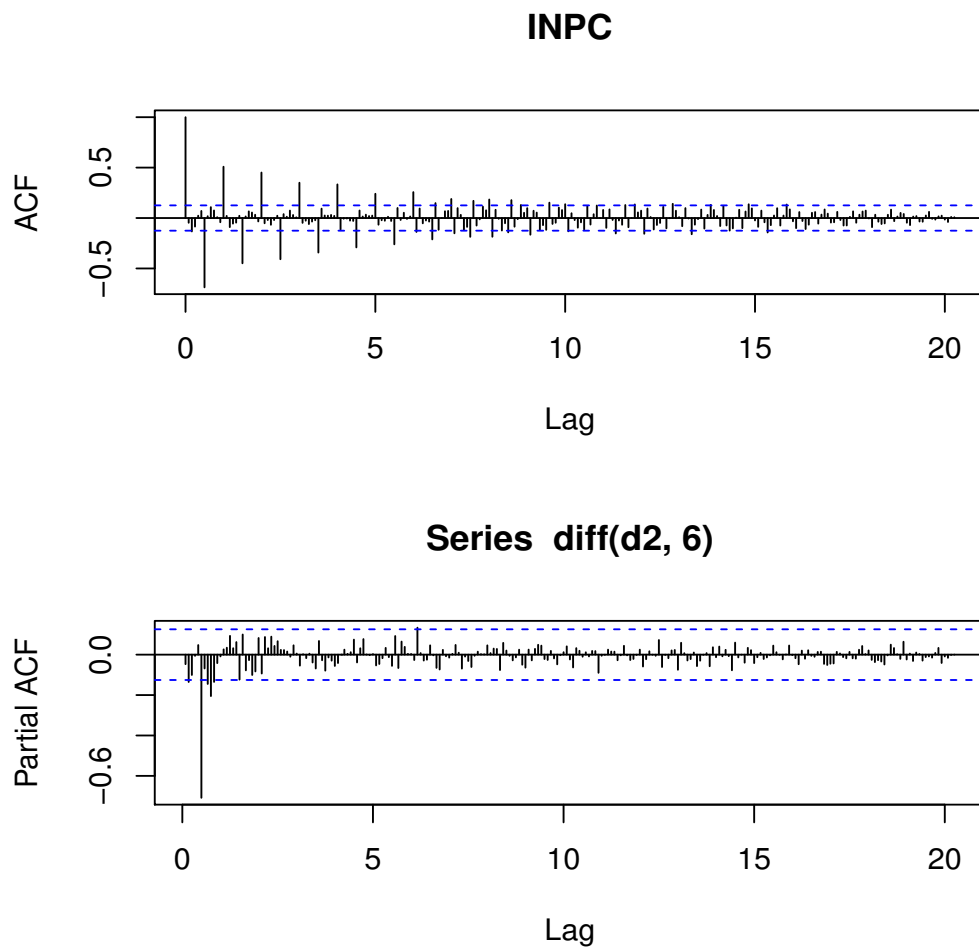


Figura 11:

Pero al realizar una diferencia estacional cada 12 observaciones, desaparece dicho comportamiento además de que logramos obtener una serie estacionaria en media como se muestra en las siguientes gráficas. Por lo que vamos a quedarnos con ésta diferencia para la estimación de los parámetros de la parte regular ( $p,d,q$ ) y la parte estacional ( $P,D,Q$ ).

### Con 1 diferencia estacional y 2 normales

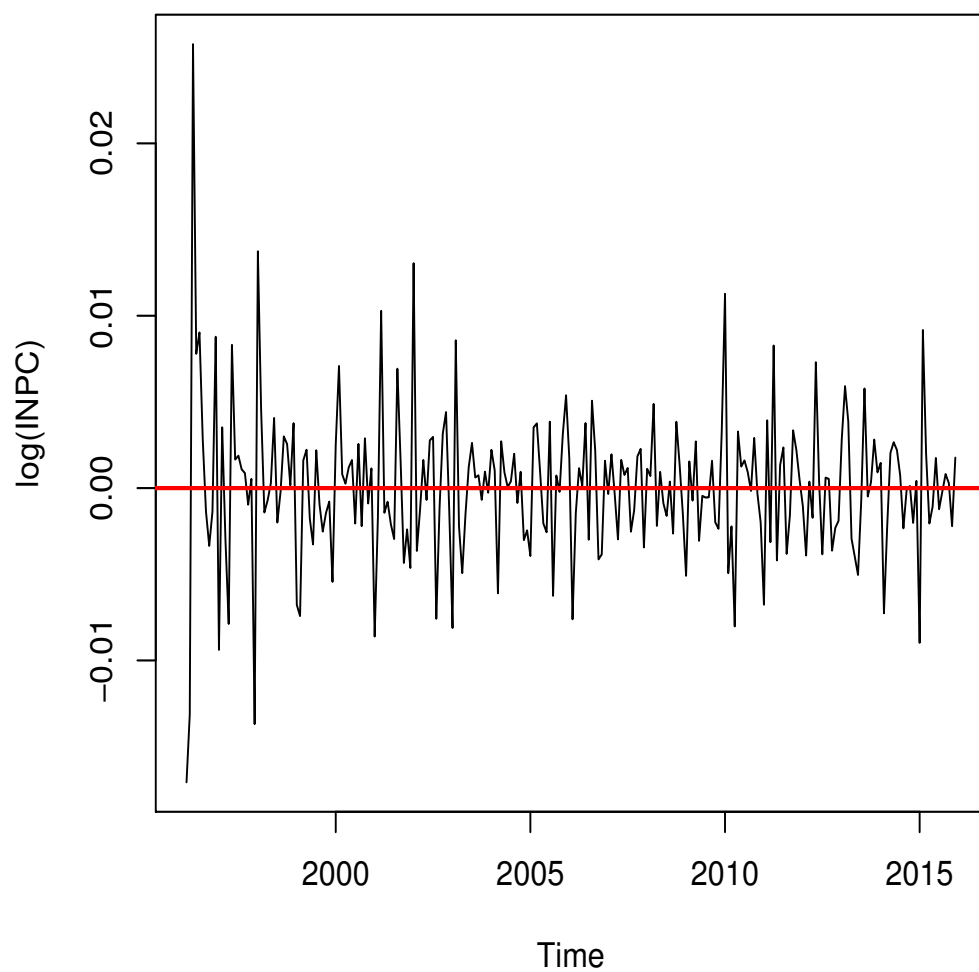


Figura 12:

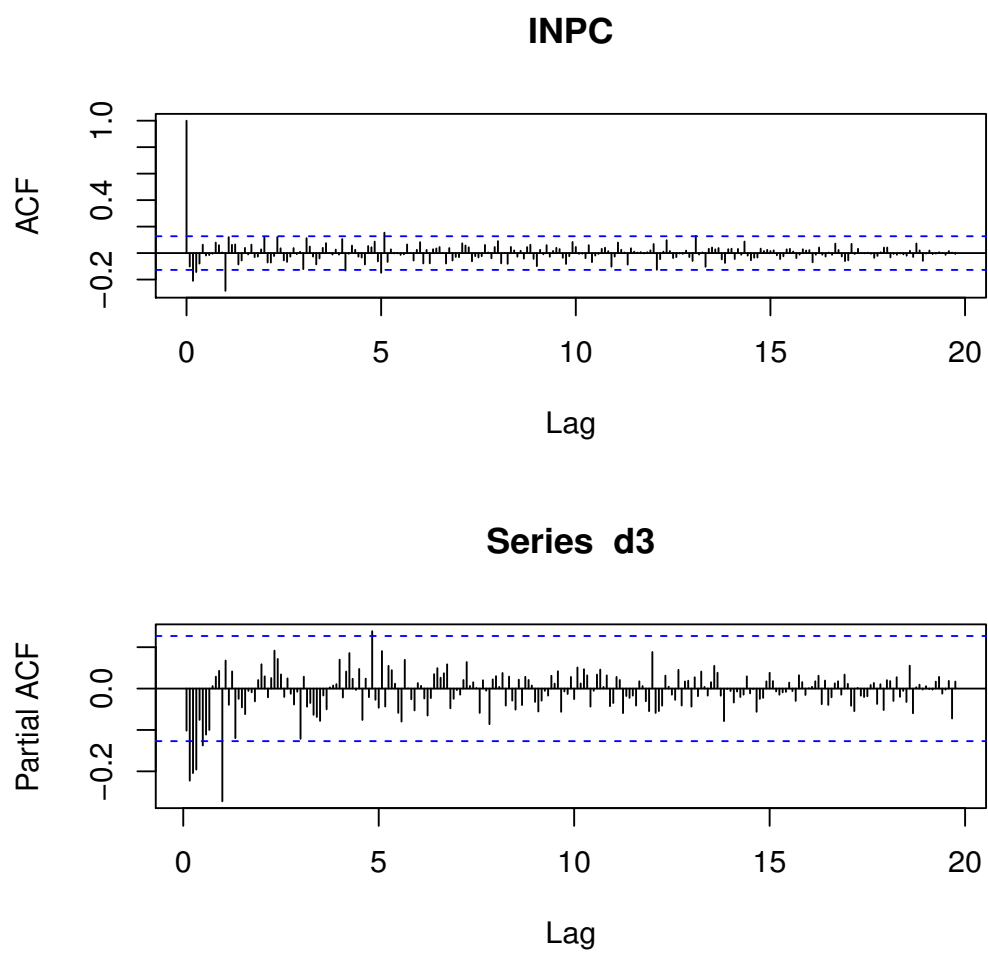


Figura 13:

## 5. Ajuste del modelo y AIC

El criterio de Akaike (AIC) nos ayuda a determinar entre dos o más modelos cuál es mejor, para ello, buscamos aquel que tenga un AIC menor

Durante el análisis de las gráficas anteriores consideramos los siguientes modelos:

<i>Modelo</i>	<i>AIC</i>
ARIMA(12,2,12)(0,1,0)	-1949
ARIMA(6,2,12)(0,1,0)	-1948.59

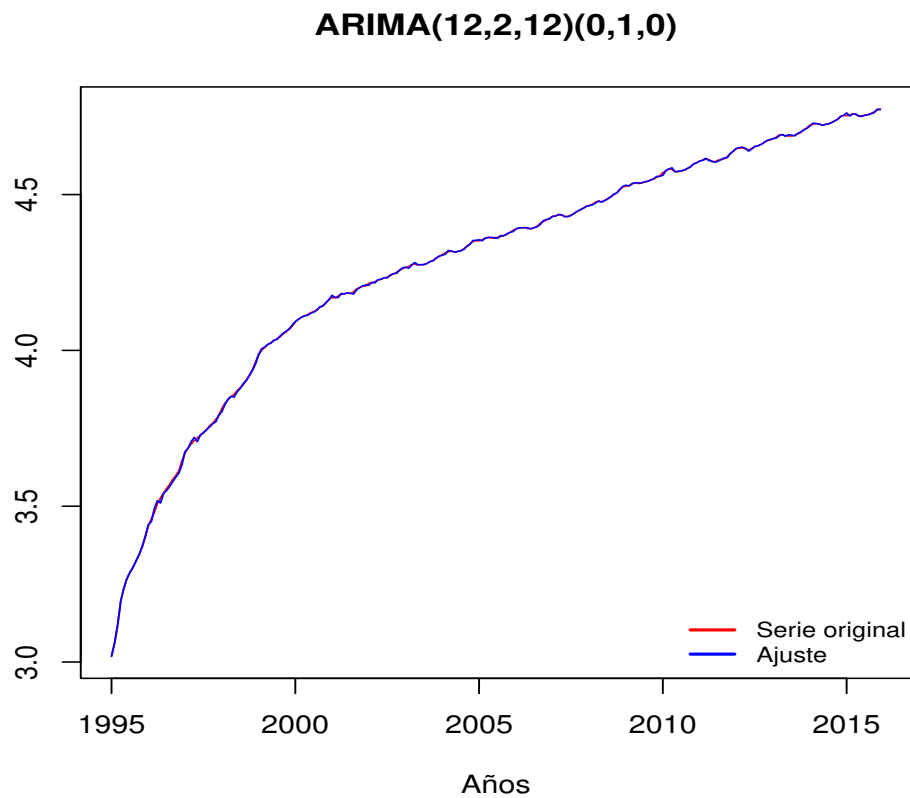


Figura 14:

**ARIMA(6,2,12)(0,1,0)**

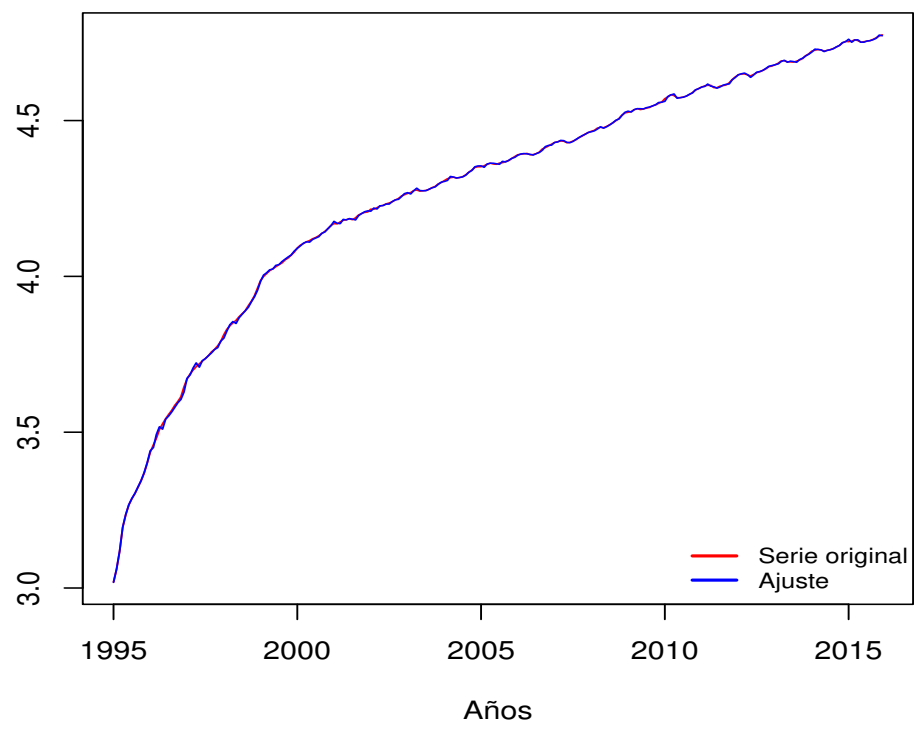


Figura 15:

Como podemos observar los dos modelos se ajustan a muy bien a la serie original, también notamos que los AIC son muy parecidos, es por eso que para el resto del trabajo vamos a trabajar con el modelo  $ARIMA(6,2,12)(0,1,0)$ , debido a que la parte autorregresiva necesita solo de 6 datos anteriores para explicar el dato siguiente, mientras que el otro necesita de 12.

### 5.1. Análisis de los residuales

Una vez que eliminamos nuestras componentes de tendencia y estacional, nuestra serie ahora debe de ser un proceso de ruido blanco, para corroborarlo veremos si cumple con las pruebas de normalidad, allí, nuestra hipótesis nula es que nuestra muestra se comporta como una distribución normal, y para Ljung-Box, nuestra hipótesis corresponde a que son no correlacionadas, trabajando con un  $\alpha = 0,05$

Veamos nuestros residuos de la siguiente manera y sus respectivas funciones de autocorrelación y autocorrelación parcial:

```
residuos = modelo$residuals  
plot(fit2, residuos,type=1)
```

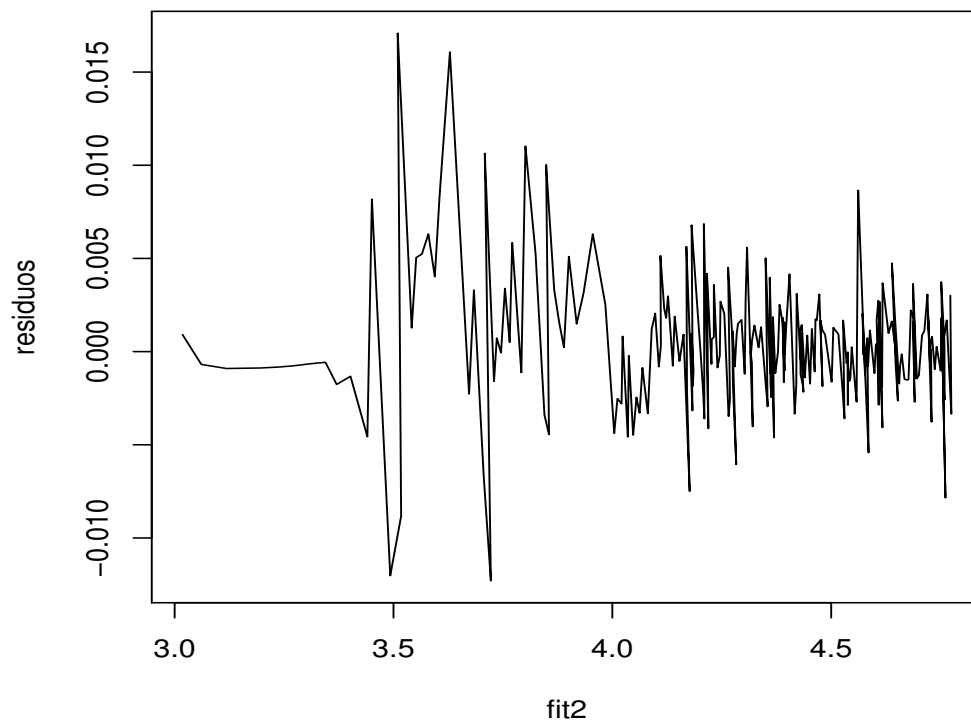
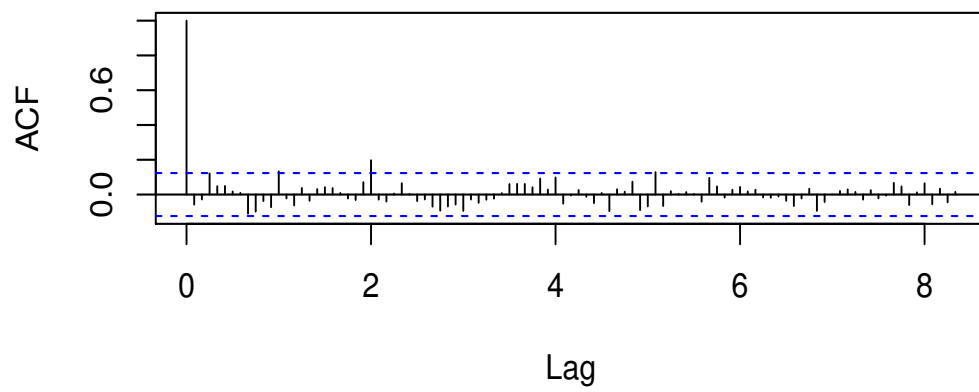


Figura 16:



### Series residuos



### Series residuos

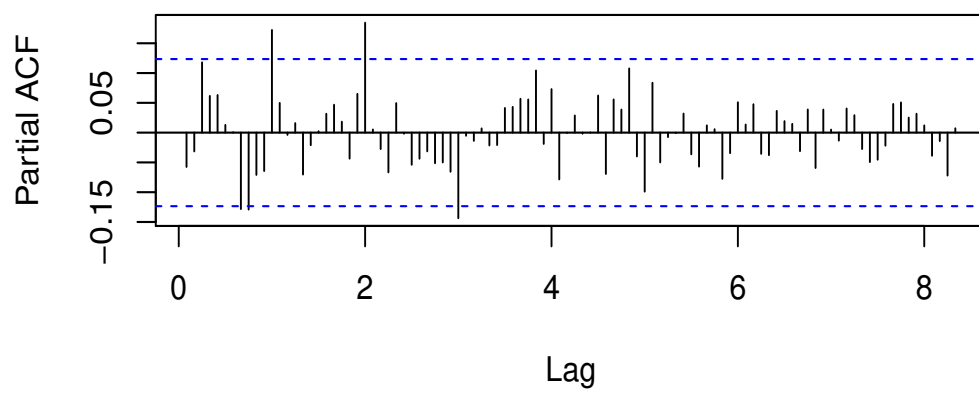


Figura 17:

Pero más importante aún, es asegurarnos que no se rechace la hipótesis, es decir, debemos fijarnos en los p valores de las pruebas y asegurarnos que estén por encima de nuestro  $\alpha$ , para ello aplicaremos pruebas de normalidad y el diagnóstico de nuestro modelo via:

```
tsdiag(modelo)
Box.test(residuos, lag = 1, type = "Ljung-Box")
```

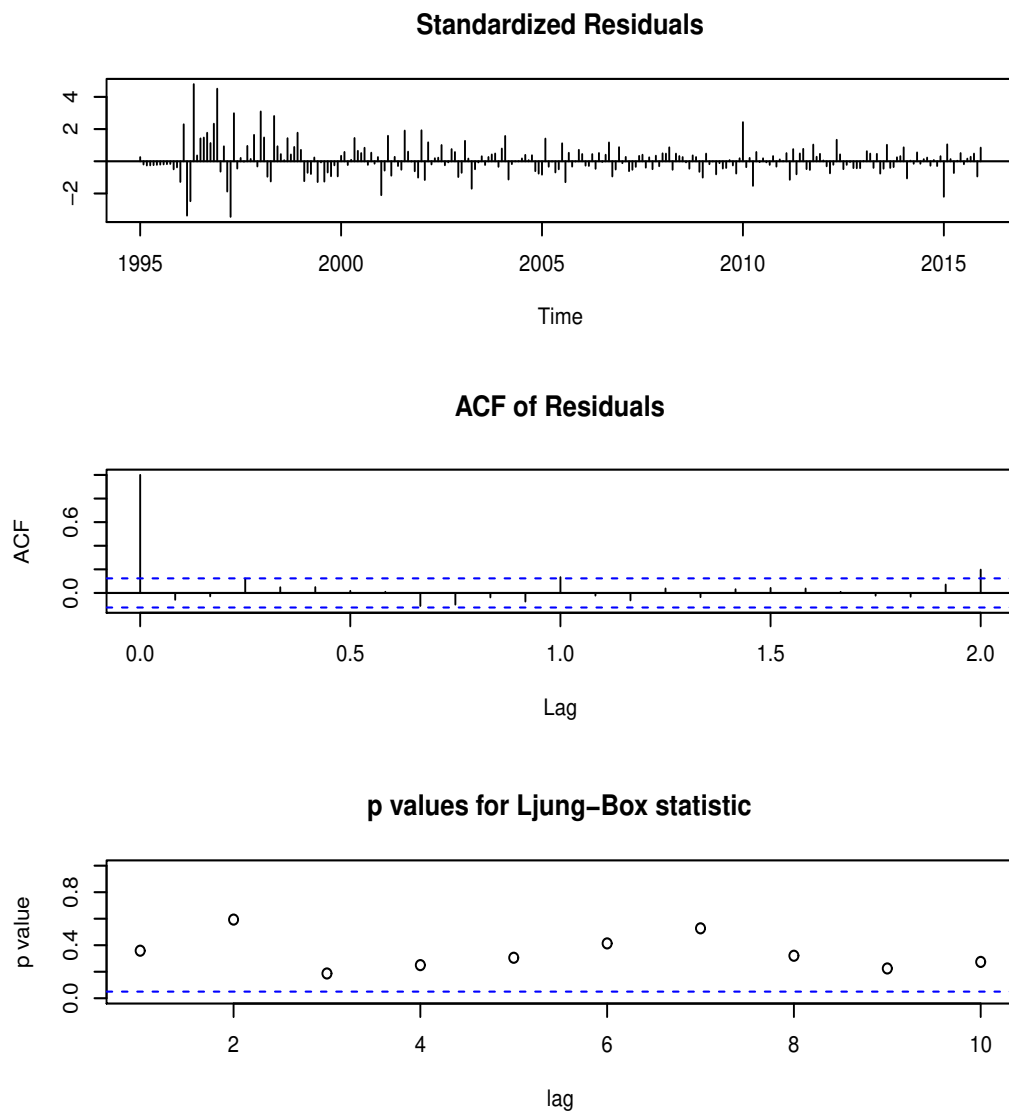


Figura 18:

Fijandonos específicamente en la Ljung-Box, podemos decir entonces que no rechazaremos nuestra hipótesis, es decir, que no existen correlaciones entre sí.

Y para las pruebas de normalidad aplicaremos distintas pruebas para asegurar que estamos tratando con un proceso de ruido blanco:

```
jarque.bera.test(residuos)
shapiro.test(residuos)
ad.test(residuos)
lillie.test(residuos)
cvm.test(residuos)
```

```
> shapiro.test(residuos)
```

Shapiro-Wilk normality test

```
data:  residuos
W = 0.92258, p-value = 3.513e-10
```

```
> ad.test(residuos)
```

Anderson-Darling normality test

```
data:  residuos
A = 4.432, p-value = 5.008e-11
```

```
> lillie.test(residuos)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data:  residuos
D = 0.11345, p-value = 2.421e-08
```

```
> cvm.test(residuos)
```

Cramer-von Mises normality test

```
data:  residuos
W = 0.73814, p-value = 3.405e-08
```

Gracias a éste análisis, podemos aceptar la hipótesis con un nivel de confianza del 95 % de que trabajamos con un proceso de ruido blanco, y así, las inferencias que hemos hecho a lo largo de este trabajo pudieron ser aplicadas de manera correcta.

## 6. Predicción

Al utilizar el modelo  $ARIMA(6,2,12)(0,1,0)$  y ejecutando los comandos correspondientes en R para conseguir las predicciones, obtuvimos los siguientes resultados para el año 2016:

Fecha	INPC	Predicción	90 % IC.L	90 % IC.U	95 % IC.L	95 % IC.U
ene-16	118.984	118.879	118.178	119.584	118.045	119.72
feb-16	119.505	119.093	117.646	120.558	117.371	120.841
mar-16	119.681	119.499	117.332	121.705	116.921	122.133
abr-16	119.302	119.233	116.419	122.115	115.888	122.675
may-16	118.77	118.646	115.276	122.115	114.641	122.791
jun-16	118.901	118.851	114.943	122.893	114.209	123.683
jul-16	119.211	118.994	114.555	123.605	113.723	124.509
ago-16	119.547	119.24	114.254	124.444	113.323	125.466
sep-16	120.277	119.677	114.054	125.577	113.008	126.74
oct-16	121.007	120.161	113.762	126.92	112.576	128.258
nov-16	N.D.	121.001	113.688	128.784	112.338	130.331
dic-16	N.D.	121.407	113.131	130.288	11.611	132.062

Gráficamente se aprecian de la siguiente forma:

## Serie de tiempo

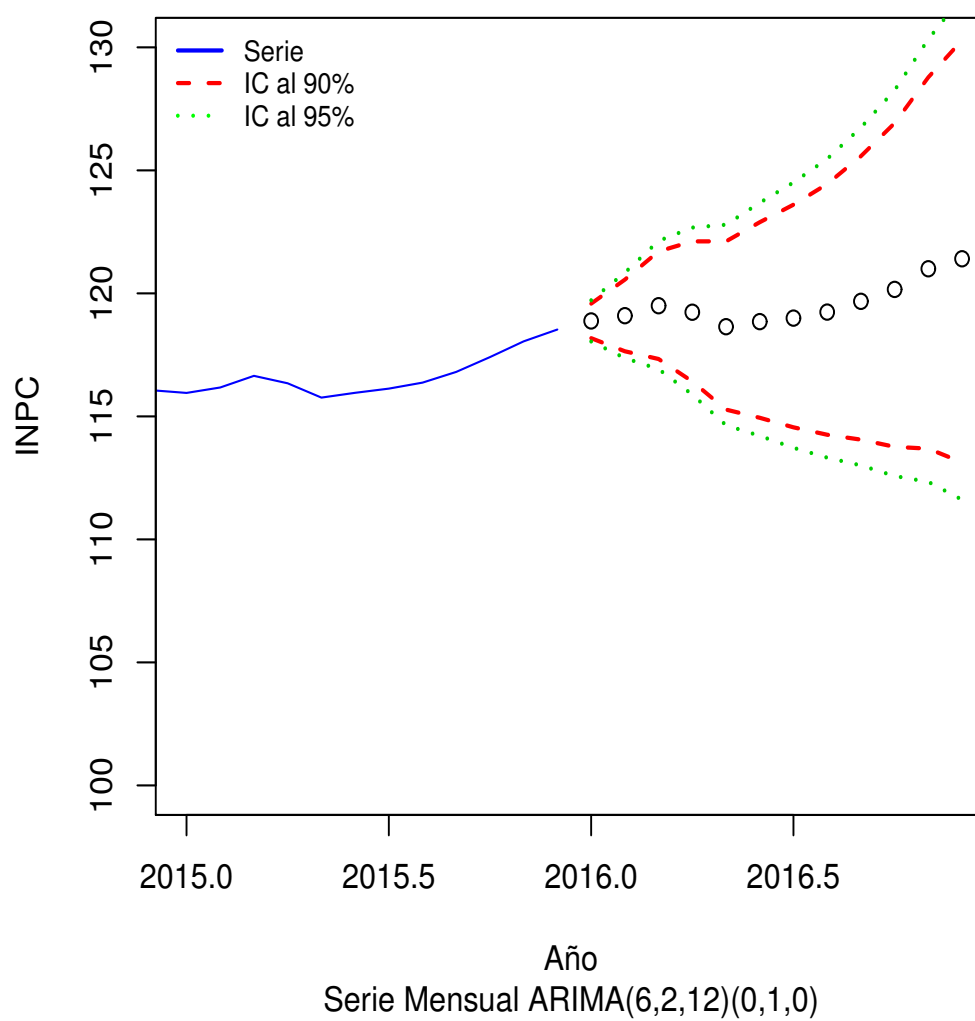


Figura 19:

Para la elaboración de este proyecto se solicitaron 6 predicciones del INPC, las predicciones empiezan a partir de enero del 2016, se hizo de esta forma para comparar el comportamiento de nuestro modelo respecto a los datos reales, ya que contamos con información hasta octubre del 2016. Nosotros presentamos 6 datos adicionales, los correspondientes al periodo junio – diciembre del 2016, estos datos se obtuvieron con la finalidad de comprobar que a un plazo mayor nuestro modelo siga arrojando buenas predicciones.

## **7. Conclusiones**

La serie del INPC resultó ser un poco fácil de manejar ya que, a pesar de contar con componentes como estacionalidad y tendencia, no necesitamos de muchas diferencias ni transformaciones para eliminarlas.

Consideramos que el modelo está teniendo un comportamiento adecuado al menos para obtener información correspondiente a un año ya que es muy cercano al dato real, tomando en cuenta que este último cae dentro de los intervalos de confianza.

Recomendamos el análisis periódico de este modelo, debido a que está replicando un índice económico, el cual se ve afectado por diversos factores nacionales como internacionales.