UNIVERSIDAD NACIONAL DE COLOMBIA

PREGRADO EN ESTADISTICA

Departamento de Estadística
Facultad de Ciencias

Series de Tiempo Univariadas

Profesor:
Norman Diego Giraldo Gómez
Trabajo 2 de Series de Tiempo
Modelos ARMA-SARMA

Integrantes:

Ricardo William Salazar Espinal	C.C. 1017219472
Alejandro Velásquez Rendón	C.C. 1000913570

Medellín, Colombia

Medellín, Febrero de 2025

Índice

Índice de figuras

Índice de cuadros

1 Introducción

Dada una serie de tiempo, denotada por Y_t, se pretende ajustar a estos datos un posible modelo ARMA Estacional (S-ARMA o, de forma simple, SARMA) a esta serie de tiempo.

2 Punto a) Identificación y estimación

Encuentre el periodo de la serie. Luego aplique la función `auto.arima` con la opción `stationary=TRUE`, `seasonal=TRUE` para identificar un posible SARMA. Estime el modelo con `arima`. Examine los parámetros significativos con la función `coef.test`. Reporte el modelo definitivo. Reporte la tabla de valores estimados y valores p.

3 Solución a)

En primer lugar, se analizará si la serie es estacionaria en covarianza, para concluir sobre esto se usará la prueba de hipótesis KPSS, implementada en la librería `aTSA` con el comando `kpss.test`, los resultados se ven a continuación.

3.1 KPSS

```
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag  stat p.value
   6 0.327     0.1
-----
 Type 2: with drift no trend
 lag  stat p.value
   6 0.901    0.01
-----
 Type 1: with drift and trend
 lag  stat p.value
   6 0.247    0.01
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```

Si tomamos la serie como tipo 1, no se rechaza la hipótesis nula con un valor P de 0.1 (mayor a un nivel de significancia de $\alpha = 0.05$), o sea que la serie Y_t es estacionaria en covarianza.

Por tanto, no hay problema para ejecutar el objetivo del trabajo.

3.2 Periodo de la serie

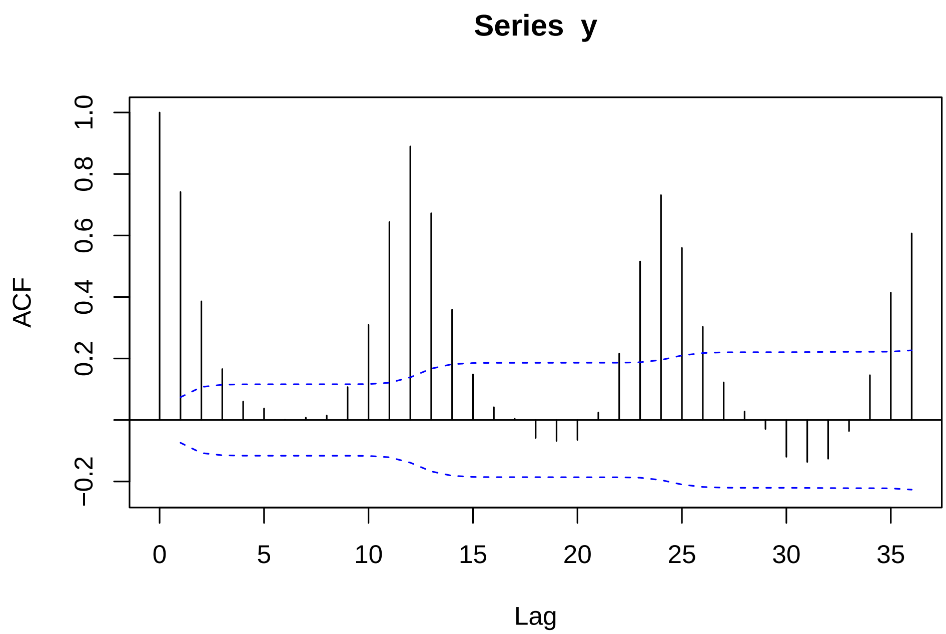


Figura 1: Función FAC de la serie

```
[1] 12
```

Ahora que se sabe que la serie es estacionaria, se procede a buscar el periodo s del posible modelo SARMA, para esto la idea general es buscar los picos relevantes de la función FAC (que se puede ver en la figura 1) de la serie. En este caso se obtiene un valor de $s = 12$, luego, como mínimo se obtiene que el modelo es de la forma:

$$Y_t \sim ARMA(p, q)(p_s, q_s)[12]$$

```
[1] 720
```

Resta hallar los valores de los hiperparámetros del modelo, para esto será de utilidad la figura 1, que muestra la gráfica basada en el criterio de información bayesiano, para elegir los órdenes p y q, sin embargo, no hay valores claros para los rezagos de los errores (q), no hay un claro valor favorito, una mejor forma de abordar este problema es mediante la función `auto.arima` del R, dando como parámetro a $s = 12$, y asumiendo, como ya se sabe, que la serie Y_t es estacionaria en covarianza, y que se quiere hayar un modelo con componente estacional, la función en cuestión nos arroja un resultado de `ARIMA(2, 0, 1)(2, 0, 2)[12]`, es decir que nuestro modelo aparentemente tiene la forma:

$$Y_t \sim ARMA(2, 1)(2, 2)[12]$$

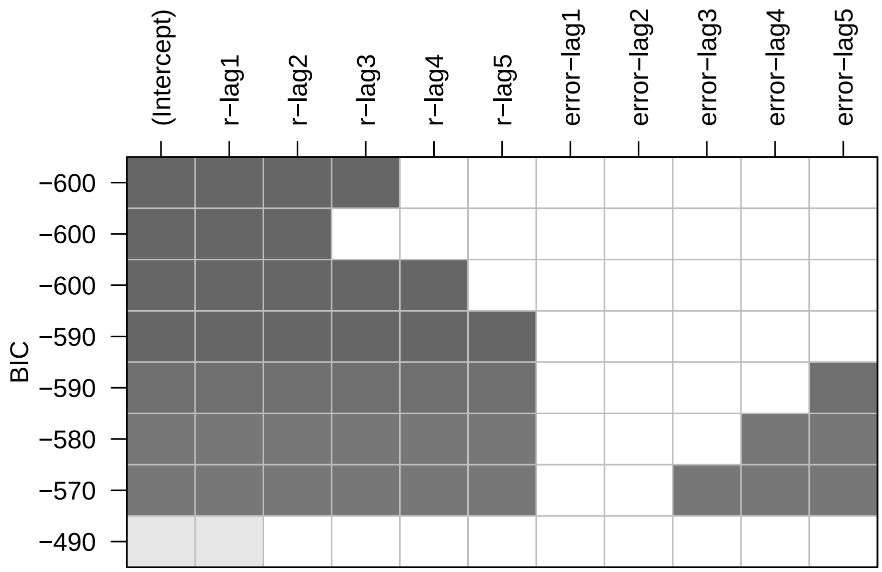


Figura 2: Índice BIC

Sin embargo, al ajustar este modelo, se encontraban problemas en la significancia de los coeficientes estimados del modelo, es decir, algunos de los coeficientes de orden mayor del modelo no daban significativos o no se ajustaban bien a la serie muestral, el orden de los hiperparámetros se fue reduciendo hasta llegar a un modelo adecuado que se explora a continuación.

3.3 Modelo Estimado

Se obtiene que la serie de tiempo se puede ajustar como sigue:

$$Y_t \sim ARMA(2,0)(0,1)[12]$$

	Estimado	**Error Std.**	**Estadístico z**	**Valor P**
ar1	1.032858	0.035221	29.3247	2e-16
ar2	-0.363855	0.035337	-10.2967	2e-16
sma1	0.952788	0.014047	67.8287	$<$ 2e-16
intercept	1.587077	0.829148	1.9141	0.05561

Tabla 1: Coeficientes estimados del modelo SARMA

En la tabla 1 se observan los coeficientes estimados de nuestro modelo ajustado, en la última columna de la misma se ve que todos son significativos, por tanto se puede considerar este modelo como uno válido que puede llegar a ajustar bien a nuestra serie Y_t.

4 Punto b) Validación de supuestos

Generar la gráfica de la fac y el resultado de la prueba Ljung-Box, con un número de rezagos utilizados, por ejemplo `nlag=3 x periodo`.

5 Solución b)

En la figura 3 muestra la función de autocorrelación de los residuales del modelo SARMA propuesto, muestra problemas, pues en muchos lags, hay valores que sobresalen las bandas azules, lo cual no es deseable, y nunca llega a oscilar alrededor del 0, o al menos no lo hace ni hasta el lag 36.

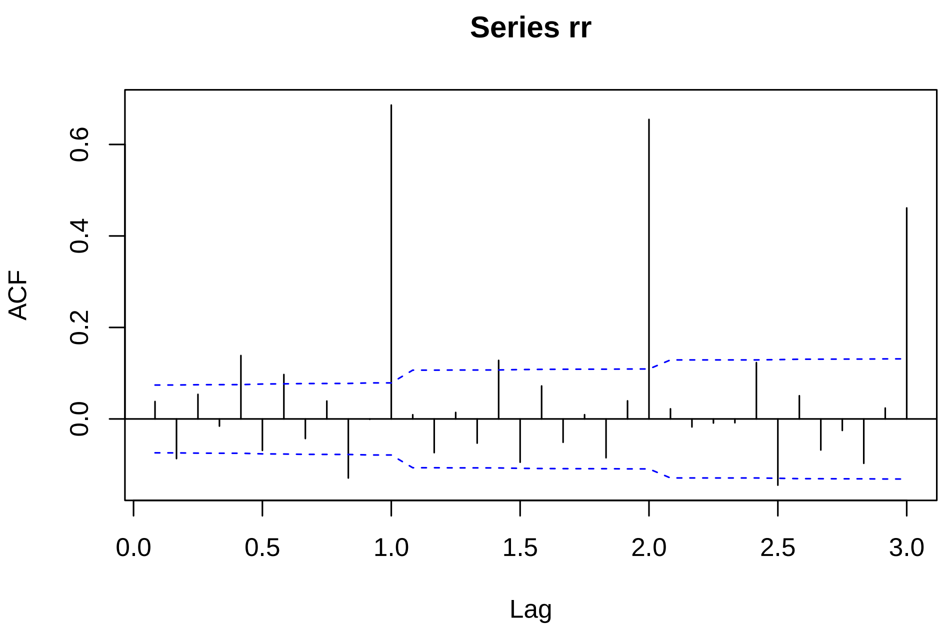


Figura 3: FAC de los residuos

Aunque, como es sabido, basados en una gráfica no podemos sacar conclusiones estadísticamente significativas, por tanto se hará una prueba formal para esto, la prueba en cuestión es la prueba de Ljung-Box para ruido blanco, el juego de hipótesis es el siguiente:

$$\begin{cases} H_0 : Z_n \sim RB(0, \sigma^2) \\ H_1 : no(H_0) \end{cases}$$

Y la conclusión se obtiene gracias a la función `Box.test` de R, cuya salida se imprime a continuación:

```
    Box-Ljung test

data:  rr
X-squared = 382.75, df = 12, p-value < 2.2e-16
```

Nuevamente, a un nivel de significancia de $\alpha = 0.05$, se rechaza H_0 (VP=2.2e-16), es decir que los residuos de nuestro modelo no son Ruido Blanco. Este es un supuesto fundamental de los modelos SARMA, sin embargo, se ajustará el modelo propuesto por fines académicos, en parte porque no se hayaron modelos que cumplan este supuesto, y, como se verá más adelante, ambas funciones FAC (la teórica y la muestral) se mostrarán similares.

6 Punto c) Verificación del ajuste del modelo

Reporte la fac calculada con los coeficientes estimados versus la fac muestral. Comente. Reporte la gráfica del periodograma y la densidad espectral estimada. Comente. Para la prueba KS aleatorizada reporte el histograma con la línea vertical que indica el valor observado del estadístico. Reporte el valor p de la prueba. Comente sobre el resultado final.

7 Solucióon c)

Para la resolución de este apartado, se utilizará una función de usuario en R, la cual está anexada en el apéndice de códigos. Esta función calcula los coeficientes del modelo SARMA a partir de los coeficientes previamente estimados de las siguientes componentes:

- **AR**: AutoRegresiva,
- **MA**: Media Móvil,
- **SAR**: AutoRegresiva Estacional,
- **SMA**: Media Móvil Estacional.

Además, la función recibe el **período estacional** de la serie y genera la representación final del modelo **SARMA**, combinando los polinomios autoregresivos y de media móvil con sus respectivos componentes estacionales.

7.1 Funciones de Autocorrelación

En la gráfica 4, se ve la comparación entre ambas funciones de autocorrelación, la FAC teórica (rojo) sigue una estructura similar a la observada (líneas grises), lo que indica que el modelo está capturando algunos de los patrones de autocorrelación en la serie de tiempo.

Adicionalmente se observa que los picos principales en la FAC teórica coinciden con los de la observada, lo que sugiere que el modelo recoge patrones clave en la serie.

La periodicidad estacional parece estar representada en la FAC del modelo, lo cual es fundamental en modelos SARMA que es justamente con el que se está trabajando.

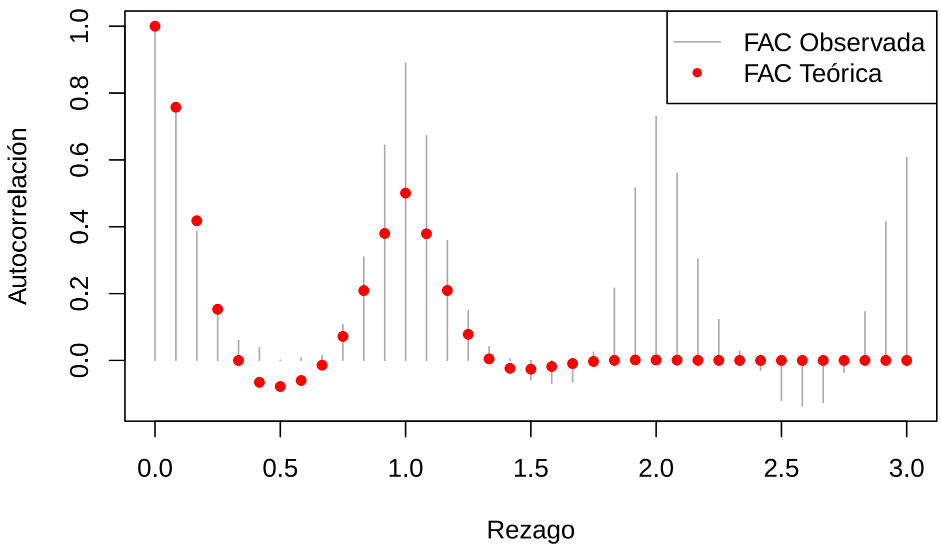


Figura 4: Comparación de las FAC

A pesar de estas bondades el modelo teórico presenta diferencias en la magnitud de algunos coeficientes de autocorrelación, lo cual indica que el modelo no reproduce exactamente la estructura de dependencia temporal.

En conclusión la comparación de la FAC permite evaluar la fidelidad del modelo en la captura de la estructura de correlación de la serie. Aunque el modelo no es una representación exacta, logra recoger los patrones más importantes, lo que justifica su uso como la mejor opción disponible.

7.2 Densidad Espectral Estimada

El espectro teórico (en negro) y el observado (en rojo) siguen un patrón similar en términos de frecuencia, indicando que el modelo SARMA captura adecuadamente las principales características espectrales de la serie temporal.

Las oscilaciones periódicas visibles en ambos espectros sugieren que el modelo logra capturar las componentes cíclicas presentes en la serie temporal.

7.3 Periodogramas Acumulados y prueba KS aleatorizada

Ahora se hará el análisis de los periodogramas acumulados para la serie observada y la serie teórica modelada mediante *SARMA*.

Lo primero que se observa es que ambas curvas (azul y roja) siguen un patrón muy similar, lo que sugiere que el modelo SARMA representa razonablemente bien el comportamiento

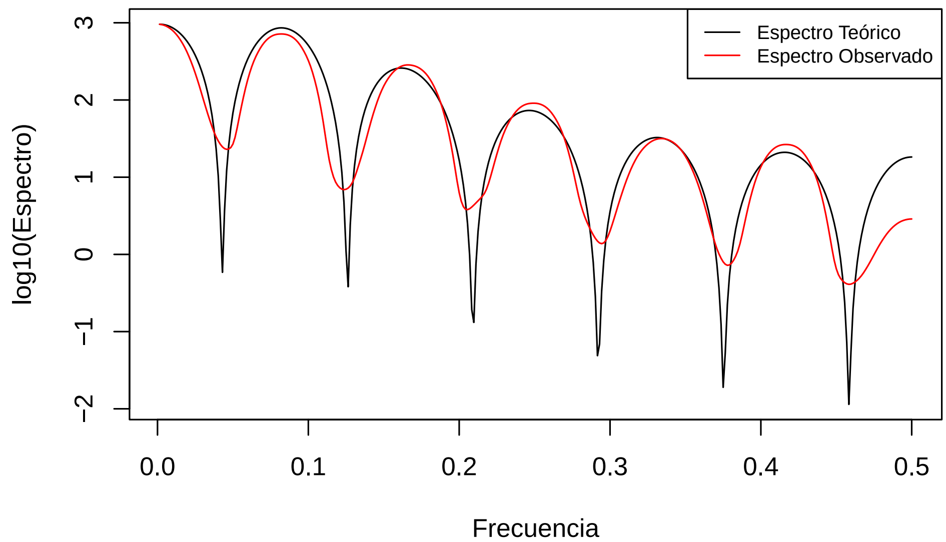
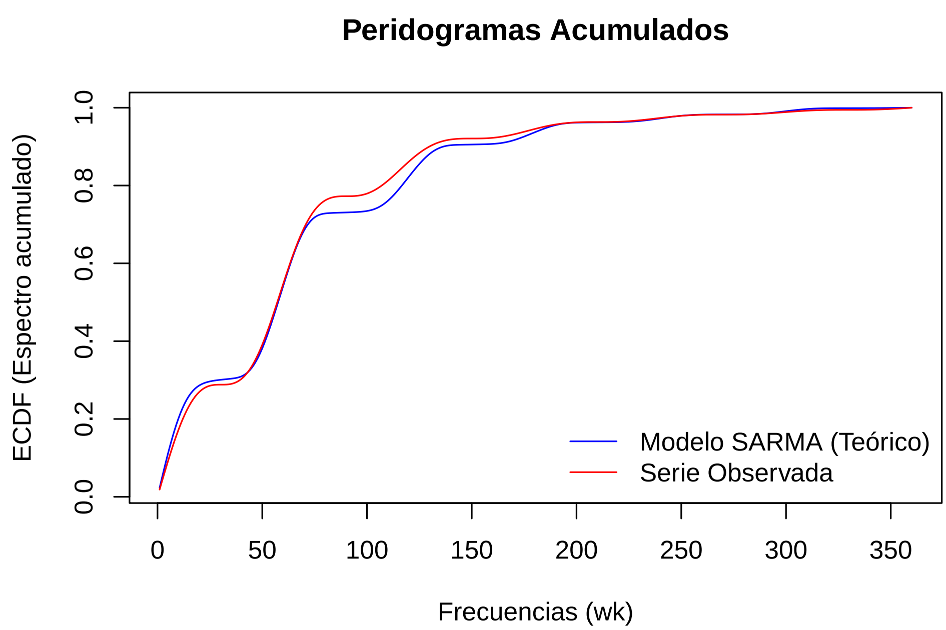


Figura 5: Espectro Observado vs. Teórico

espectral acumulado de la serie observada.



A pesar de esto, hay pequeñas discrepancias entre las curvas, especialmente en algunas secciones (por ejemplo, entre los índices 50 y 100). Esto podría indicar posibles Variaciones en la dinámica de las frecuencias que el modelo SARMA no captura perfectamente.

De todas maneras, los dos periodogramas acumulados son muy parecidos, por lo que se puede esperar conseguir un buena ajuste.

A continuación se comprobará la igualdad entre las densidades espectrales de ambas series (la observada vs la teórica), mediante la prueba de homegeneidad aleatorizada usando el estadístico de Kolgomorov-Smirnov (K-S). El juego de hipótesis es:

$$\begin{cases} H_o : I_{Y_n}(w_k) = I_Y(w_k), \ k = 1, ..., \lfloor \frac{n}{2} \rfloor \\ H_1 : no(H_0) \end{cases}$$

Asumiendo claro, que Y representa al modelo teórico.

En este caso lo que se busca es obtener un valor p alto para no rechazar la hipótesis nula, de esta manera se podría concluir que el modelo SARMA propuesto es una buena aproximación de la serie observada.

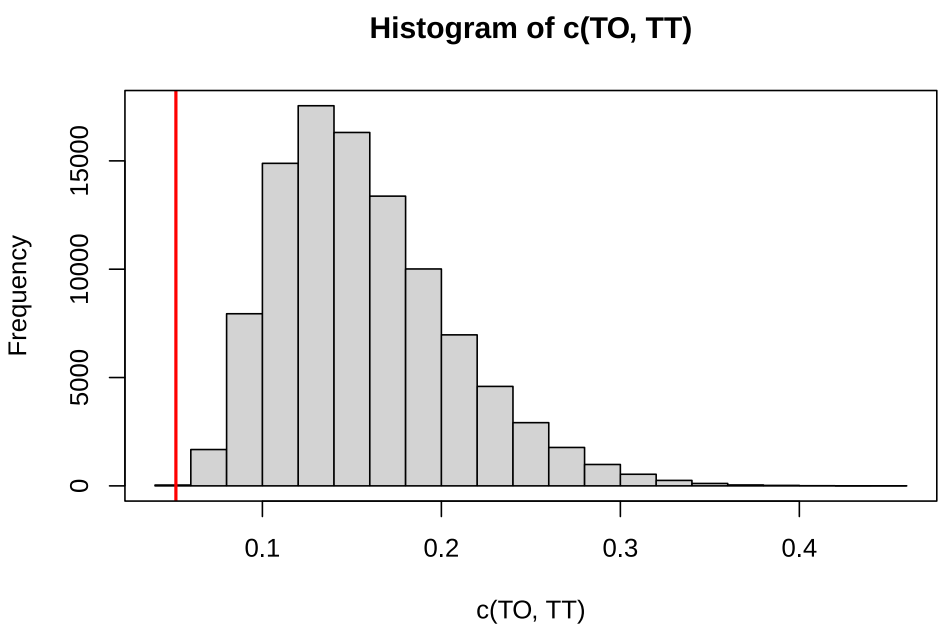


Figura 6: Histograma asociado a la prueba K-S

La prueba de Kolmogorov-Smirnov aplicada a los espectrogramas acumulados del modelo SARMA y de la serie observada respalda la hipótesis nula de que no existen diferencias significativas entre sus distribuciones acumuladas. El valor p alto de 0.99997 indica que la distancia máxima entre las distribuciones acumuladas observada (línea roja en el histograma) y las simuladas es consistente con las distribuciones generadas por el modelo.

En términos prácticos, esto significa que el modelo SARMA es una muy buena aproximación de las características espectrales de la serie observada. Las discrepancias locales identificadas previamente en los gráficos de espectrogramas acumulados son mínimas y no tienen un impacto estadísticamente significativo.

7.4 Conclusión

Este análisis confirma que el modelo SARMA propuesto puede considerarse adecuado para describir el comportamiento espectral de la serie temporal observada.

8 Apéndice-Código Usado

8.1 Llamado de librerías

```
library(kableExtra)
library(rlist)
library(TSA)
library(forecast)
library(lmtest)
library(polynom)
library(astsa)
library(tseries)
```

8.2 Punto a)

```
#Lectura de los datos
L = list.load("L.rdata")
y = L$serie8

aTSA::kpss.test(y) #Prueba kpss

#Gráfica FAC
Rk.est = stats::acf(y,36,ci.type="ma")

#Periodo del modelo
k.mx = which(Rk.est$acf[-c(1,2)]==max(Rk.est$acf[-c(1,2)]))
(p = k.mx+1)

k4 = kernel("modified.daniell", c(4,7,12))
B = spec.pgram(y, k4, taper=0, log = "yes",  ci = 0.8)

(p= 1/B$freq[which.max(B$spec[-seq(1,70)])])


spec <- spec.ar(y,plot=FALSE)
```

```r
abline(v=spec$freq[145])
period <- round(1/spec$freq[145])

#Posible modelo
p = 12
y = ts(y,frequency = p)
arima<-auto.arima(y, stationary=TRUE, seasonal=TRUE)

#Índice BIC
res.arma=armasubsets(y=y,nar=5,nma=5,y.name='r',ar.method='ols')
par(mfrow=c(1,1))
plot(res.arma)

#Modelo
mod = arima(y,order=c(2,0,0), seasonal=list(order=c(0,0,1), period = p))

#Coeficientes estimados
coef<-coeftest(mod)
theta = mod$coef
```

8.3 Punto b)

```r
#FAC residuales
rr = mod$residuals
acf(rr,36,ci.type='ma',drop.lag.0=TRUE)

Box.test(x = rr, lag = 12, type="Ljung-Box")
```

8.4 Punto c)

```r
#Función de usuario
gen.sarma.coef = function(ar,ma,ars,mas,p){

require(polynom)

period = p
bs <- polynomial(c(rep(0,period),1))
b1 <- polynomial(c(0,1))

  arpoly <- polynomial(c(1,-ar))
  mapoly <- polynomial(c(1,ma))
  sarpoly <- polynomial(c(1,-ars))
  smapoly <- polynomial(c(1,mas))
```

```r
  fullarpoly <- arpoly*predict(sarpoly,bs)
  fullmapoly <- mapoly*predict(smapoly,bs)

  mo <- list()
  mo$ar <- -coef(fullarpoly)[-1]
  mo$ma <- coef(fullmapoly)[-1]

return(mo)}

#Uso de la función anterior
ar.e = theta[1:2]
ma.e = 0
ars.e =  0
mas.e = theta[3]

mo = gen.sarma.coef(ar=ar.e,ma=ma.e, ars=ars.e, mas=mas.e, p=p)

#Gráfica de las FAC
Rk = ARMAacf(ar = mo$ar, ma = mo$ma, lag.max = 36)
par(mfrow=c(1,1))
Rk.est = stats::acf(y,36,ci.type='ma', plot=FALSE)

plot(Rk.est$lag, Rk.est$acf, type='h', col='darkgray', xlab="Rezago",
     ylab="Autocorrelación")
points(Rk.est$lag, Rk, pch=20, cex=1.2, col='red')

# Agregar leyenda
legend("topright", legend=c("FAC Observada", "FAC Teórica"),
       col=c("darkgray", "red"), pch=c(NA, 20), lty=c(1, NA), lwd=c(1, NA))

#Gráfica de los espectogramas
V = arma.spec(ar=mo$ar, ma=mo$ma, n.freq=length(B$freq), plot=FALSE)

a = max(B$spec)/max(V$spec)

plot(B$freq, log10(a * V$spec), type='l', lwd=1, col='black',
     xlab="Frecuencia", ylab="log10(Espectro)")

lines(B$freq, log10(B$spec), col='red', lty=1, lwd=1)

# Agregar leyenda
legend("topright", legend=c("Espectro Teórico", "Espectro Observado"),
       col=c("black", "red"), lty=c(1, 1), lwd=c(1, 1), cex=0.75)
```

```r
#Periodogramas acumulados
ecdf_1 <- cumsum(B$spec) / sum(B$spec)
ecdf_2 <- cumsum(V$spec) / sum(V$spec)
wk <- seq(1, length(ecdf_1))

# Graficar
plot(wk, ecdf_1, type = 'l', col = 'blue',
     xlab = "Frecuencias (wk)",
     ylab = "ECDF (Espectro acumulado)",
     main = "Peridogramas Acumulados") # Título del gráfico
lines(wk, ecdf_2, col = 'red') # Agregar la segunda línea

# Leyenda
legend("bottomright", legend=c("Modelo SARMA (Teórico)", "Serie Observada"),
       col = c("blue", "red"), lty = 1, bty = "n")

#Prueba de homogeneidad aleatorizada
Z <- c(B$spec,V$spec)
n <- length(B$spec)
m <- length(V$spec)
N <- length(Z)

TS <- function(Iwk1,Iwk2){
ecdf_1 <- cumsum(Iwk1)/sum(Iwk1)
ecdf_2 <- cumsum(Iwk2)/sum(Iwk2)
abs_dif <-  abs(ecdf_1- ecdf_2)
distancia_ks <- max(abs_dif)
return(distancia_ks)
}

TO <- TS(B$spec,V$spec)
TT <- vector()
K = 100000

for(i in 1:K){
  set.seed(i)
  Z.pi <- sample(Z, N, replace = FALSE)
  TT[i] <- TS(Z.pi[1:n], Z.pi[(n+1):(n+m)])
}

#Histograma
hist(c(TO,TT))
abline(v = TO, lwd = 2, col = "red")
box()
```

```r
pvalue = mean(TT>TO)
```

Referencias

Gomez, N.D.G. *Introduccion a Series de Tiempo con aplicaciones en R*. Universidad Nacional de Colombia.