

# Temas selectos de Econometría y Finanzas. Modulo de series de tiempo

José Antonio García Ramírez

Tarea1, 31/8/2018

1. Describir en términos **intuitivos** la relación entre el teorema de Kolmogorov y el supuesto de procesos ergódicos.

Recordemos que ambos temas, el teorema de Kolmogorov y el supuesto de ergodicidad, hablan de procesos estocásticos así que ambos deben de hablar de colecciones de variables aleatorias definidas en los reales.

Supongamos que vemos a una serie de tiempo solo como un conjunto de mediciones donde interviene el tiempo, si suponemos ergodicidad implícitamente estamos aceptando la idea de que la serie de datos provienen de un proceso estocástico, sin embargo con este supuesto podemos conocer, asintóticamente o bien con un tamaño de muestra alto, el primer momento de tal proceso (el que origina a la serie) lo cual de entrada se agradece pues con una sola realización de la serie, lo que estamos llamando *simple path*, con el supuesto de ergodicidad no lo solo nos informa sobre la esperanza de nuestro *simple path* observado sino del objeto matemático que lo origina, el *ensemble*. Es decir que con el supuesto ergódico al conocer la esperanza de un *simple path* se tiene la del *ensemble* que lo origina. Esta colección finita o infinita de simples path's comparten la esperanza y aceptan un modelo paramétrico (pues de eso nos fiamos en los cursos de series de tiempo), por lo cual son de dimensión finita, y de paso caracterizamos el primer momento con una simple realización. Entonces se tienen las condiciones suficientes para el teorema de Kolmogorov porque se puede especificar la distribución de una familia de variables aleatorias, nuestro *ensemble*, y entonces tenemos la garantía de que existe un proceso estocástico que origina a la serie de tiempo que estamos observando y de la cual suponemos que es ergódica.

Sea de más si la realización observada de nuestra serie de tiempo es estacionaria en sentido débil entonces su varianza y covarianza son finitas y podemos conocer sus segundos momentos y autocorrelaciones que, sea de paso, también se comparten con los demás miembros del *ensemble* pues estos se definen en términos del primer momento que es común.

Resumiendo, el proceso de ergodicidad nos garantiza la existencia de un proceso estocástico y nos informa sobre sus dos primeros momentos y su autocovarianza tan solo al observar una realización que sea prudente.

2. Sea  $Z_t \sim N(0, 1)$  empareje cada una de las siguientes series de tiempo con su correspondiente correlograma de la figura 2.

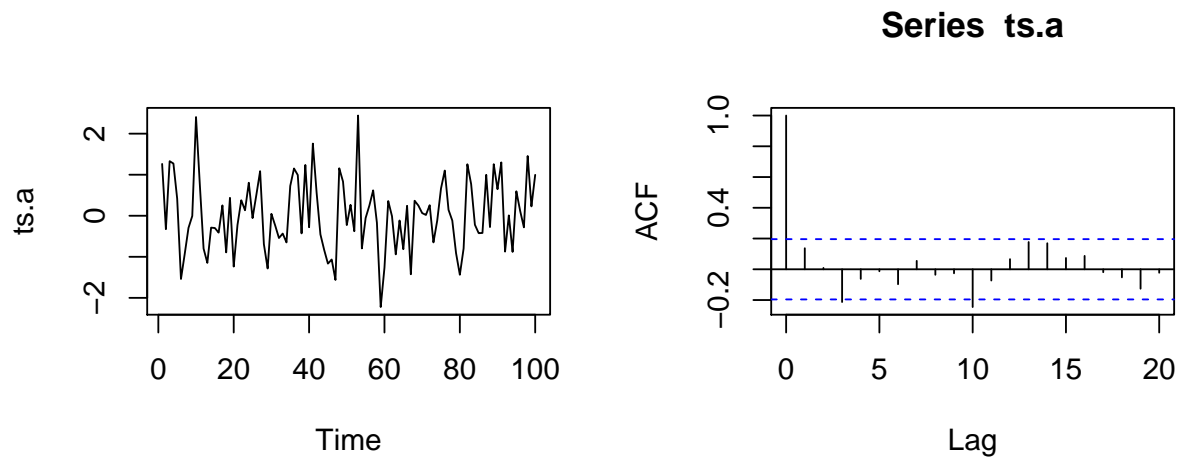
A lo largo del capítulo dos el texto da hints sobre cómo se comportan diferentes series y su respectivo correlograma. Por lo cual es posible responder el ejercicio sin simular, pero para entrar en contexto también simulé cada una de las series y calcule su correlograma.

- a)  $X_t = Z_t$ . Esta serie se corresponde con ruido blando, se corresponde con la gráfica r3 de la figura pues la serie en general está acotada en  $[-3, 3]$  alrededor del cero y en su correlograma para todos los lags no se observa nada fuera de los intervalos de confianza salvo el primer cálculo que por definición vale la unidad.
- b)  $X_t = -0.3X_{t-1} + Z_t$ . Esta serie corresponde a un modelo  $AR(1)$  alternante pues el coeficiente autorregresivo de primer orden es negativo así que le corresponde la gráfica r4 pues su correlograma es alternante (positivo, negativo, etc) además se distingue el primer lag en el correlograma como negativo
- c)  $X_t = \sin(\pi t/3) + Z_t$ . Esta serie corresponde con la figura r5 pues su correlograma exhibe periodicidad (además de que el dibujo de la serie presenta estacionalidad).
- d)  $X_t = Z_t - 0.3Z_{t-1}$ . Esta serie es un modelo  $MA(1)$  con coeficiente negativo por lo que destaca el correlograma en  $k = 1$ . Le corresponde la figura r1

- e)  $X_t = 2 - 3t + Z_t$ . Esta serie tiene una marcada tendencia negativa por lo que le corresponde el correlograma que no decrece conforme el lag  $k$  aumenta es decir el correlograma  $r2$

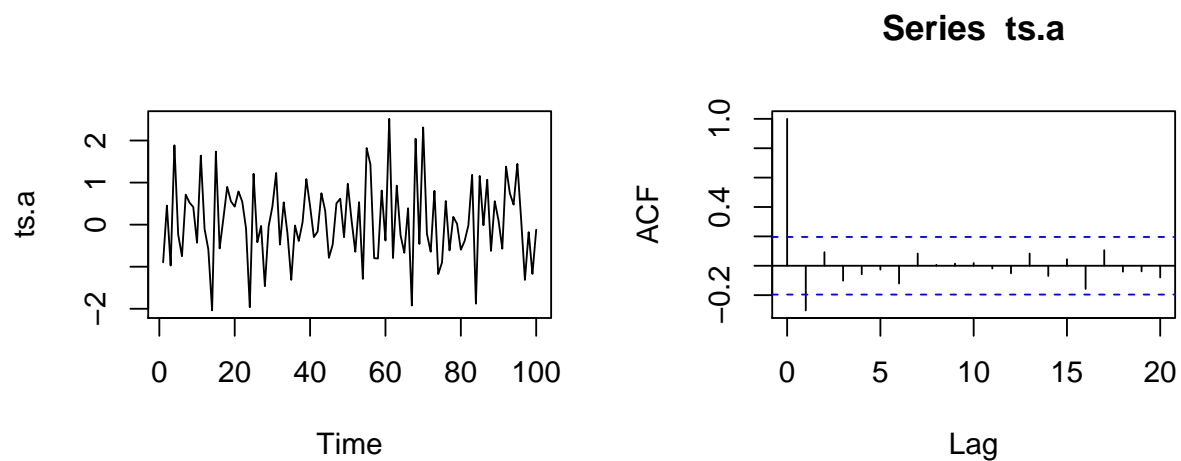
Serie a)

```
par(mfrow = c(1,2))
n <- 100
set.seed(0)
inciso.a <- rnorm(n)
ts.a <- ts(inciso.a, start = 1)
plot(ts.a)
acf(ts.a)
```



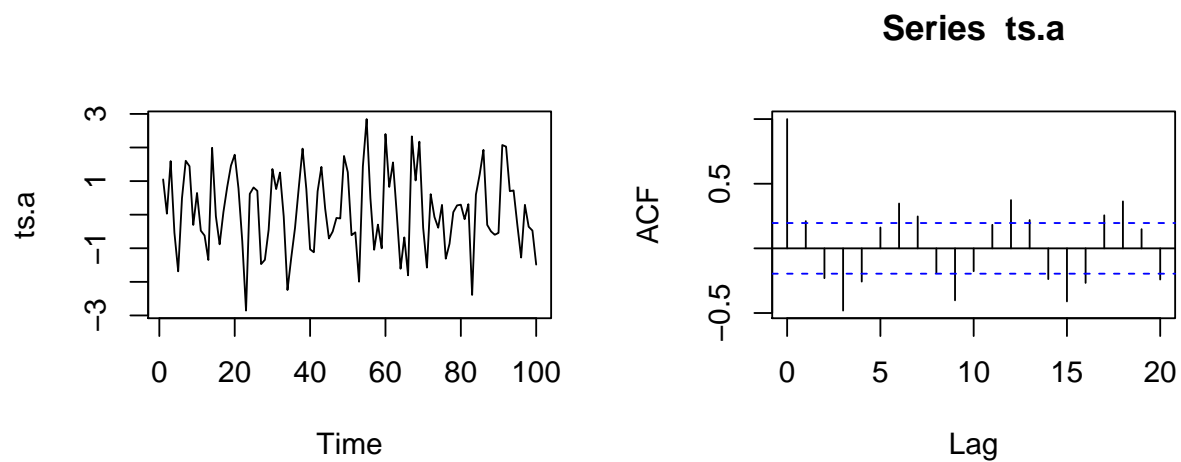
Serie b)

```
par(mfrow = c(1,2))
set.seed(0)
inciso.b <- c(runif(1), rep(-Inf, n))
for(i in 2:length(inciso.b)) inciso.b[i] <- (-0.3)*inciso.b[i-1] + rnorm(1)
ts.a <- ts(inciso.b[2:length(inciso.b)], start = 1)
plot(ts.a)
acf(ts.a)
```



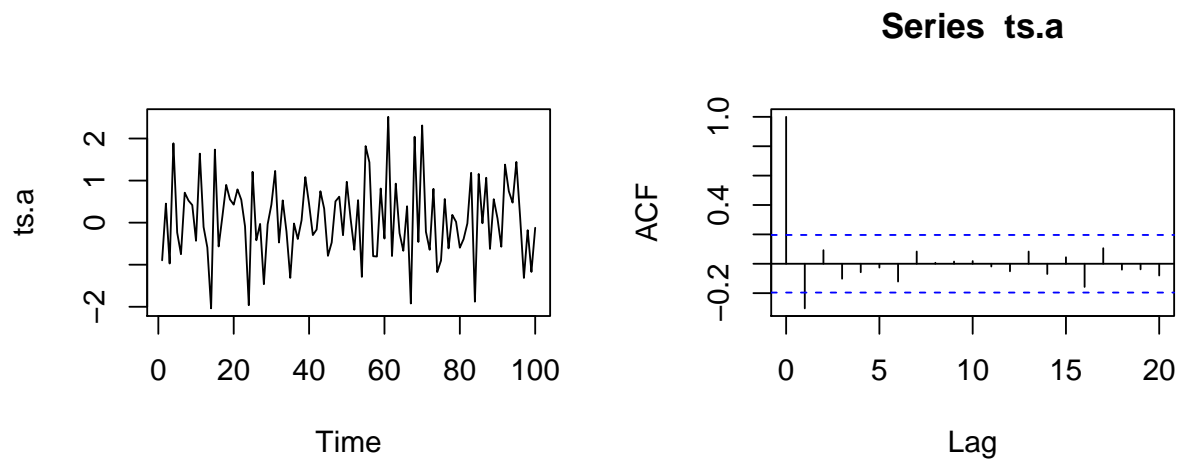
Serie c)

```
par(mfrow = c(1,2))
set.seed(0)
inciso.c <- c(runif(1), sin(pi*1:n/3)) + rnorm(n+1)
ts.a <- ts(inciso.c[2:length(inciso.c)], start = 1)
plot(ts.a)
acf(ts.a)
```



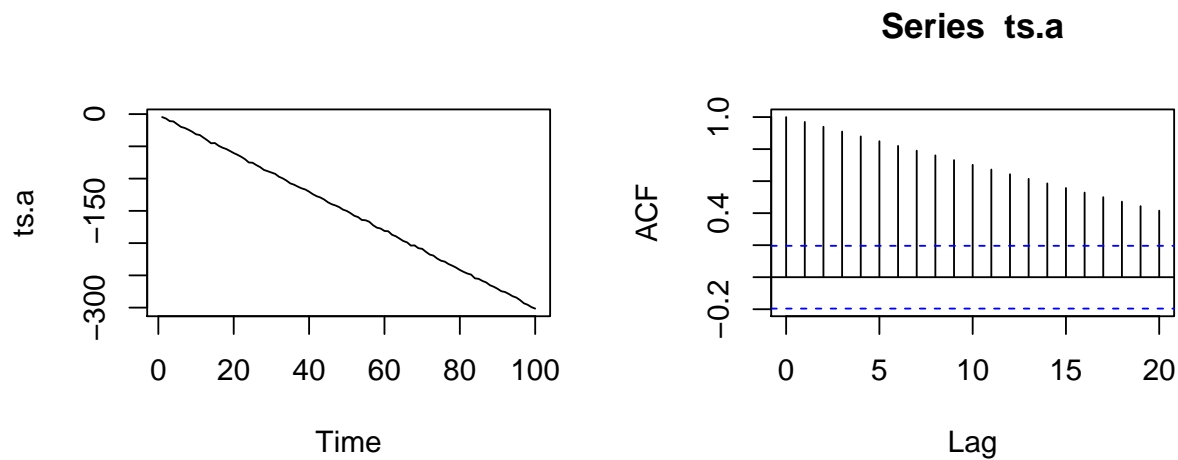
Serie d)

```
par(mfrow = c(1,2))
set.seed(0)
inciso.d <- c(runif(1), rep(0,n))
for(i in 2:length(inciso.d)) inciso.d[i] <- (-0.3)*inciso.d[i-1] + rnorm(1)
ts.a <- ts(inciso.d[2:length(inciso.d)], start = 1)
plot(ts.a)
acf(ts.a)
```



Serie e)

```
par(mfrow = c(1,2))
set.seed(0)
inciso.e <- c(runif(1, min = 10**4, max = 10**6), rnorm(n)) + 2 - 3*(1:(n+1))
ts.a <- ts(inciso.e[2:(length(inciso.e))], start = 1)
plot(ts.a)
acf(ts.a)
```



3. Considere la serie

$$Y_t = 0.4Y_{t-1} + 0.45Y_{t-2} + Z_t + Z_{t-1} + 0.25Z_{t-2}, Z_t \sim WN(0, \sigma^2)$$

a) Expresar  $\{Y_t\}$  en términos del operador  $B$ , backshift, y determine el orden  $(p, d, q)$  de este modelo.

Tenemos que

$$\begin{aligned} Y_t - .4Y_{t-1} - .45Y_{t-2} &= Z_t + z_{t-1} + .25Z_{t-2} \\ (1 - .4B - .45B^2)Y_t &= (1 + B + .25B^2)Z_t \end{aligned}$$

Un modelo ARMA(2,0,2)

b) El modelo se puede simplificar pues

$$(2 + B)(-1\bar{1} + B)Y_t = (2 + B)^2 Z_t$$

Por lo que el modelo es redundante y como comparte soluciones a los polinomios de su parte MA y RA no es invertible ni causal, sin embargo dividiendo ambos lados por  $(1 - B)$

$$(-1\bar{1} + B)Y_t = (2 + B)Z_t$$

Que es un modelo más sencillo ARMA(1,0,1)

c) Determine si el modelo es causal e invertible

Como las raíces de la parte AR del modelo del inciso reducido es  $1.\bar{1}$ , su raíz tiene modulo mayor a la unidad por lo que su parte AR es causal.

Como la raíz de la parte MA, que es -2, que es mayor a la unidad en modulo la parte MA es invertible, entonces  $Y_t$ , del inciso anterior, es causal e invertible.

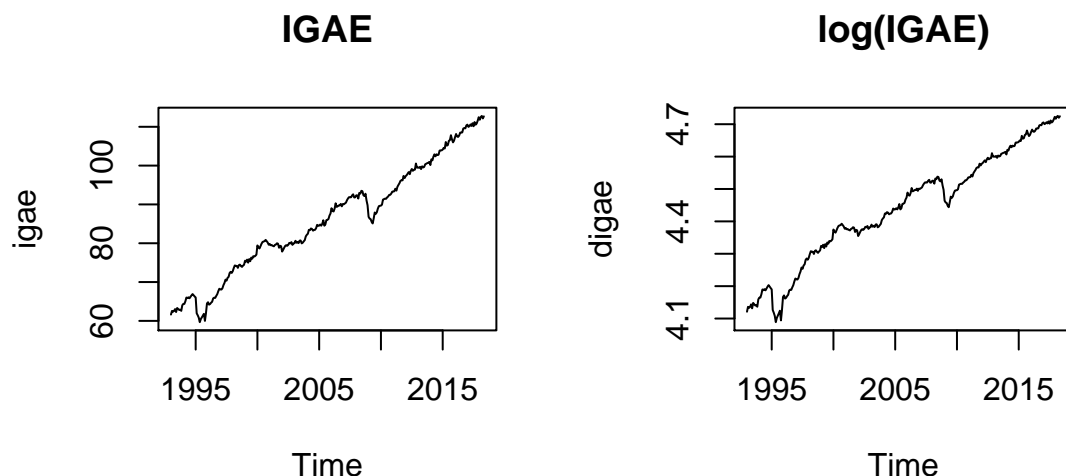
4. Repetir el ejercicio realizado en clase pero para la serie original (IGAE no desestacionalizada).

a) Graficar las series y denotar sus características estocásticas. ¿Es necesaria alguna transformación?

Asumimos que los datos que se nos brindaron son la serie de IGAE no desestacionalizada.

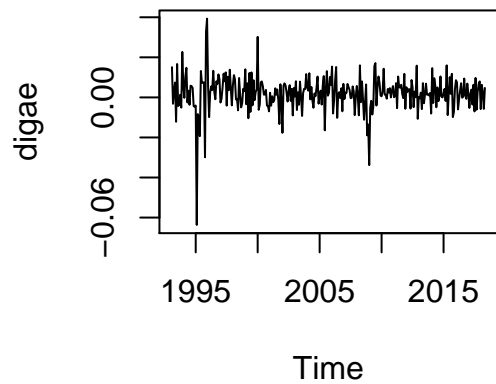
Para comenzar notamos que la varianza no es constante por lo que aplicamos la funcion ln para disminuir la varianza.

```
par(mfrow = c(1,2))
IGAE <- read.csv('IGAE.csv')
igae <- ts(IGAE$IGAE, start = c(1993,1), frequency = 12 ) #convertimos nuestros datos a una bonita serie
plot(igae, main='IGAE')
digae <- log(igae)
plot(digae, main='log(IGAE)')
```



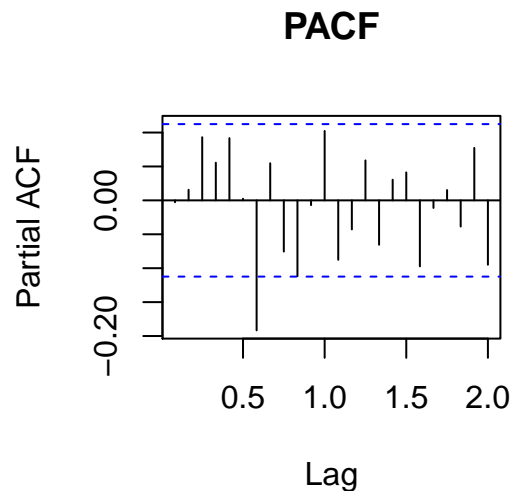
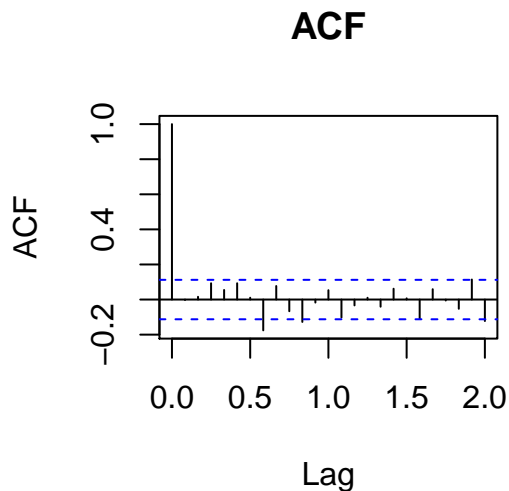
Notamos que la tendencia es evidente, se intento quitarla por medio de mínimos cuadrados, sin embargo se obtuvieron mejores resultados diferenciando la serie.

```
par(mfrow = c(1,2))
digae <- diff(digae)
plot(digae)
```



b) Estimar las ACF y las PAC

```
par(mfrow = c(1,2))
acf(digae, main = "ACF")
pacf(digae, main = "PACF")
```



Del correlograma podemos ver que el séptimo lag sobre sale lo que sugiere un modleo  $AR(7)$  y de la PACF notamos que el lag séptimo también sobre sale por lo que pensariamos en un modelo  $MA(7)$  sin embargo obtuvimos mejores residuales con un modelo  $MA(8)$

c) De acuerdo a lo anterior, construir un modelo ARIMA que consideremos correcto.

```
arma <- arima(digae, c(7, 1, 8))
```

d) Verificar si el error es ruido blanco.

Utilizamos una noble prueba de bondad de ajuste, la de Kolmogorov y verificamos que los residuales son normales la primer grafica parece indicar que no hay correlación entre ellos y los valores ajustados y la segunda que no hay correlación entre ellos.

```

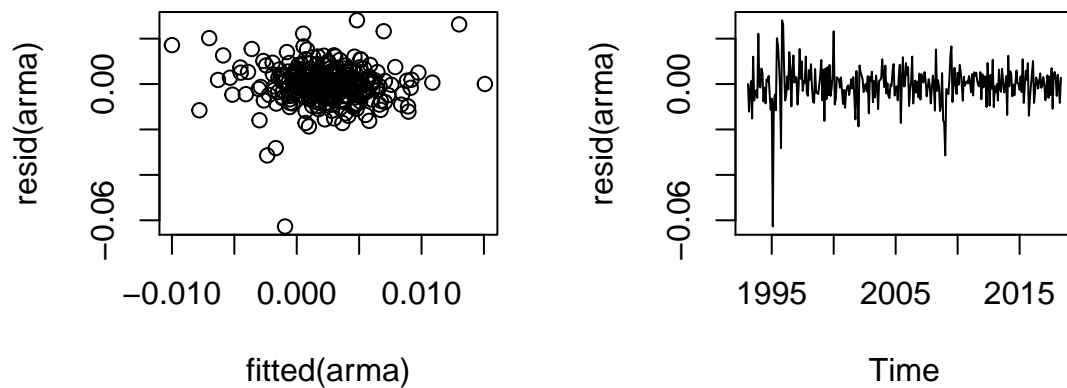
par(mfrow = c(1,2))
ks.test(resid(arma), 'pnorm')

##
## One-sample Kolmogorov-Smirnov test
##
## data: resid(arma)
## D = 0.4888, p-value < 2.2e-16
## alternative hypothesis: two-sided

plot(fitted(arma), resid(arma), main="Valores ajustados vs residuales")
plot(resid(arma))

```

### Valores ajustados vs residuales

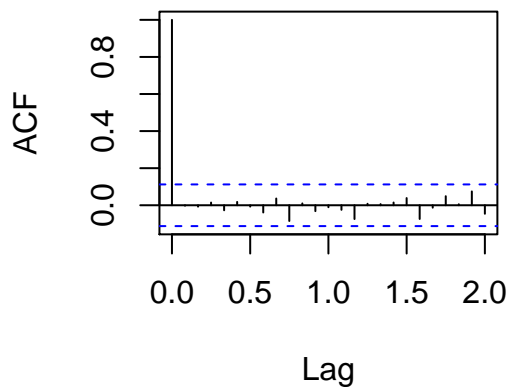


```

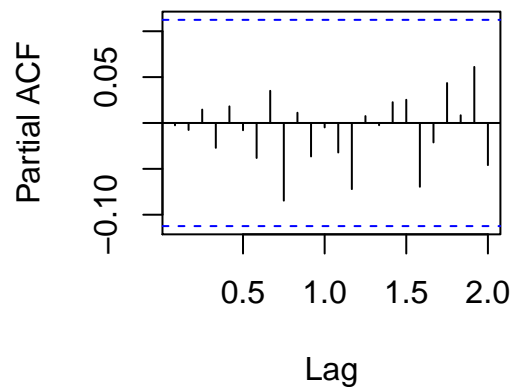
acf(resid(arma), main = "ACF residuales")
pacf(resid(arma), main = "PACF residuales")

```

### ACF residuales



### PACF residuales



Y todas las correlaciones estan dentro de los intervalos de confianza

e) Con la librería forecast, realizar la estimación automática.

```
(arma.auto <- auto.arima(digae, stepwise = FALSE))
```

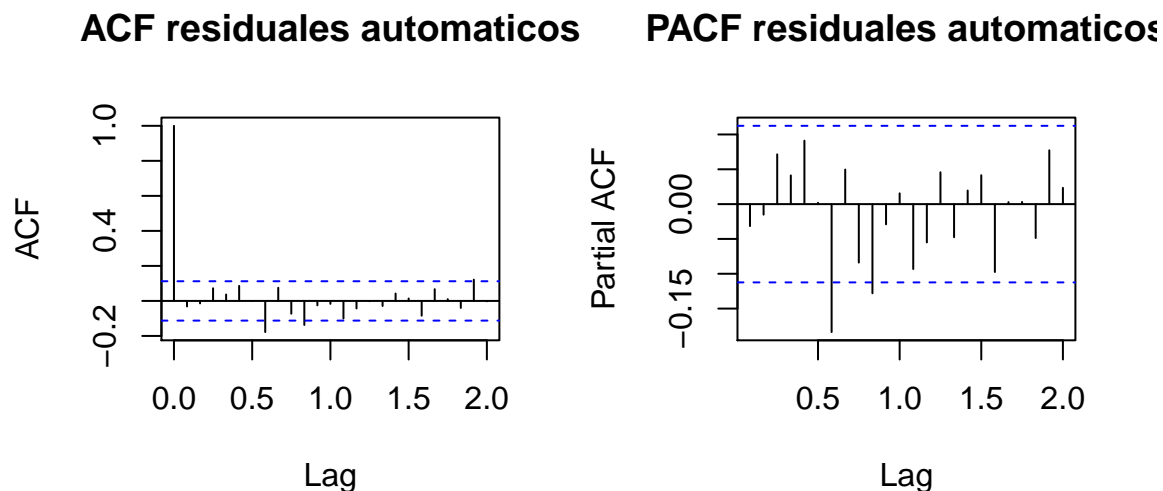
```
## Series: digae
## ARIMA(1,0,1)(2,0,0)[12] with non-zero mean
##
## Coefficients:
##          ar1          ma1          sar1          sar2          mean
##          0.7080      -0.6590      0.0795      -0.1349      2e-03
## s.e.    0.1871      0.1961      0.0594      0.0577      6e-04
##
## sigma^2 estimated as 7.666e-05:  log likelihood=1011.28
## AIC=-2010.55   AICc=-2010.27   BIC=-1988.25
```

f) Comparar el modelo estimado en el punto 3 con el obtenido en el punto anterior. Decidir: ¿Qué modelo es mejor?

```
par(mfrow = c(1,2))
ks.test(resid(arma.auto), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data:  resid(arma.auto)
## D = 0.484, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
acf(resid(arma.auto), main = "ACF residuales automaticos")
pacf(resid(arma.auto), main = "PACF residuales automaticos")
```



Aunque los residuales automaticos son normales estos no estan descorrelacionados como podemos ver en el séptimo lag del ACF y el sexto del PACF

g) Eliminar las últimas 12 observaciones, estimar el modelo seleccionado como el mejor y realizar predicciones de esas 12 observaciones.

```
H <- 12
digae_f <- digae[1:(length(digae)- H )]
```



```
digae_r <- digae[-(1:(length(digae)- H ))]
fore <- forecast(arima(digae_f, c(7,1, 7)), h = H)
fore$mean
```

```
## Time Series:
## Start = 293
## End = 304
## Frequency = 1
## [1] 0.0012917293 -0.0010662271 0.0028335415 0.0033670260 -0.0008652967
## [6] 0.0021961160 0.0044564948 0.0006011415 0.0025580533 0.0031356812
## [11] 0.0015572998 0.0024414559
```

h) Estimar el promedio de error de pronóstico.

El error medio es

```
mean(fore$mean- digae_r)
```

```
## [1] 2.776836e-05
```

```
(mean(exp(fore$mean)- exp(digae_r))) #error en la misma escala
```

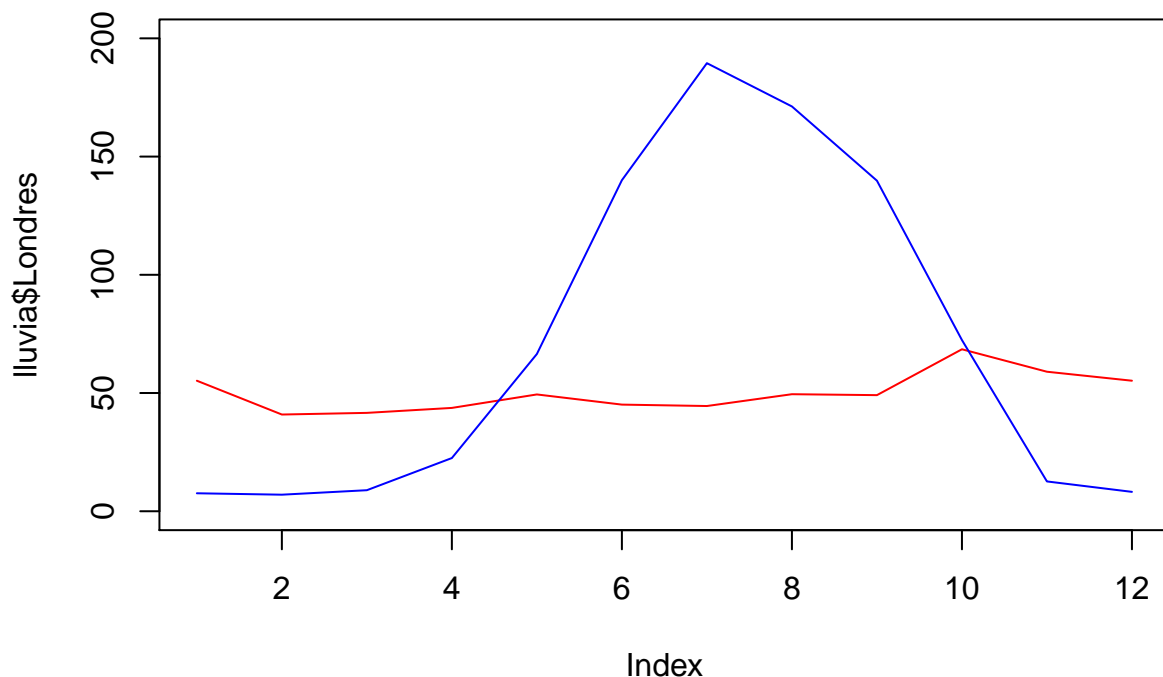
```
## [1] 1.112211e-05
```

Sin embargo recordemos que aunque el error es pequeño los datos originales han sido transformados.

5. Se adjunta un pequeño archivo de 12 datos que contiene los promedios de precipitación pluvial (mensuales) de la Ciudad de México y Londres. Si fuese el caso y con esa información, ¿ podrías argumentar estadísticamente que llueve más en alguna de las dos ciudades?

Primero graficamos nuestras observaciones, donde notamos que la componente del tiempo es muy importante pues nos muestra que en el intervalo observado llovió desigualmente en México (a través de los meses) mientras que en Londres la lluvia parece ser uniforme todo el año.

```
lluvia <- read.csv('Precipitacion.csv')
plot(lluvia$Londres, type='l', col = 'red', ylim=c(0,200))
points(lluvia$CDMX, type='l', col='blue')
```



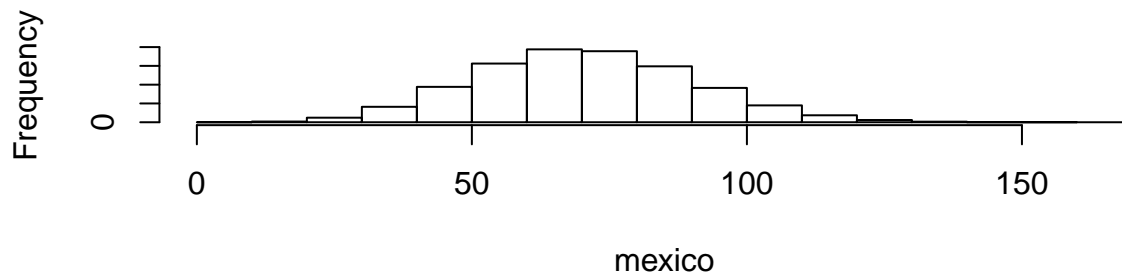
Si bien el número de observaciones es pequeño la información a priori que disponemos es importante pues sabemos que la lluvia es un fenómeno estacional, con ciclo anuales, por lo que buscar más datos parece poco informativo. En particular si prestamos atención a la pregunta si llueve mas en Mexico o en Londres, realice un test paramétrico de igualdad de medias con varianza desconocida (una prueba  $t$ ) pero realice un proceso de bootstrap para incrementar el número de observaciones y utilizar los resultados asintóticos sobre la media.

Como resultado presenté las distribuciones de las medias en el proceso de bootstrap, y finalmente los intervalos de confianza para ambas, los cuales no se intersectan y la media estimada para México es mayor.

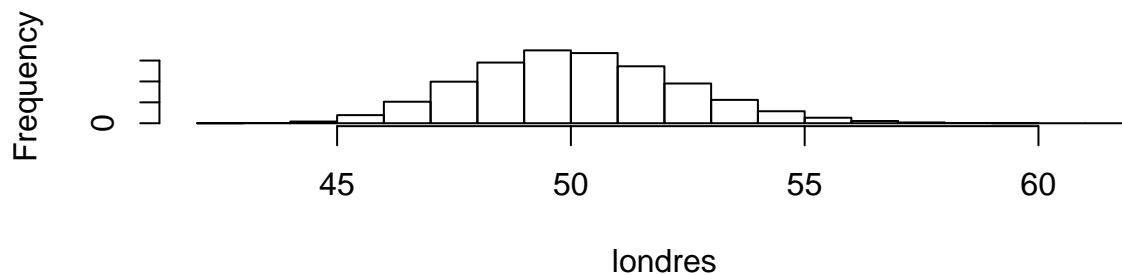
~Nunca imagine que en México, en promedio lloviese más que en Londres.

```
n <- 100000
mexico <- londres <- rep(-Inf, n)
for (i in 1:n)
{
  mexico[i] <- mean(sample(lluvia$CDMX, 12, replace = TRUE))
  londres[i] <- mean(sample(lluvia$londres, 12, replace = TRUE))
}
par(mfrow = c(2,1))
hist(mexico, main='Distribución de la media-bootstrap Mexico')
hist(londres, main='Distribución de la media-bootstrap Londres')
```

## Distribución de la media-bootstrap Mexico



## Distribución de la media-bootstrap Londres



```
t.test(mexico)
```

```
##
## One Sample t-test
##
## data:  mexico
## t = 1135, df = 99999, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  70.40125 70.64481
## sample estimates:
## mean of x
##  70.52303
```

```
t.test(londres)
```

```
##
## One Sample t-test
##
## data:  londres
## t = 7044.7, df = 99999, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  50.12084 50.14873
## sample estimates:
## mean of x
##  50.13479
```