

# Series Temporales: Trabajo Guiado

*Juanjo Sierra*

*19 de abril de 2019*

## Paquetes a cargar

Importamos los paquetes que necesitamos para resolver los problemas planteados para la práctica.

```
library(tseries)
```

## Problema

Dadas las cifras de pasajeros que vuelan con una aerolínea por año (en miles, anotadas mensualmente) entre los años 1949 y 1959, se pide estimar la cantidad de pasajeros que esa misma aerolínea tendrá cada mes durante el año 1960.

Se leen los datos de los pasajeros que nos proporciona la aerolínea.

```
serie = scan("Datos/pasajeros_1949_1959.dat")
```

Defino el número de datos que quiero usar para test y para predecir.

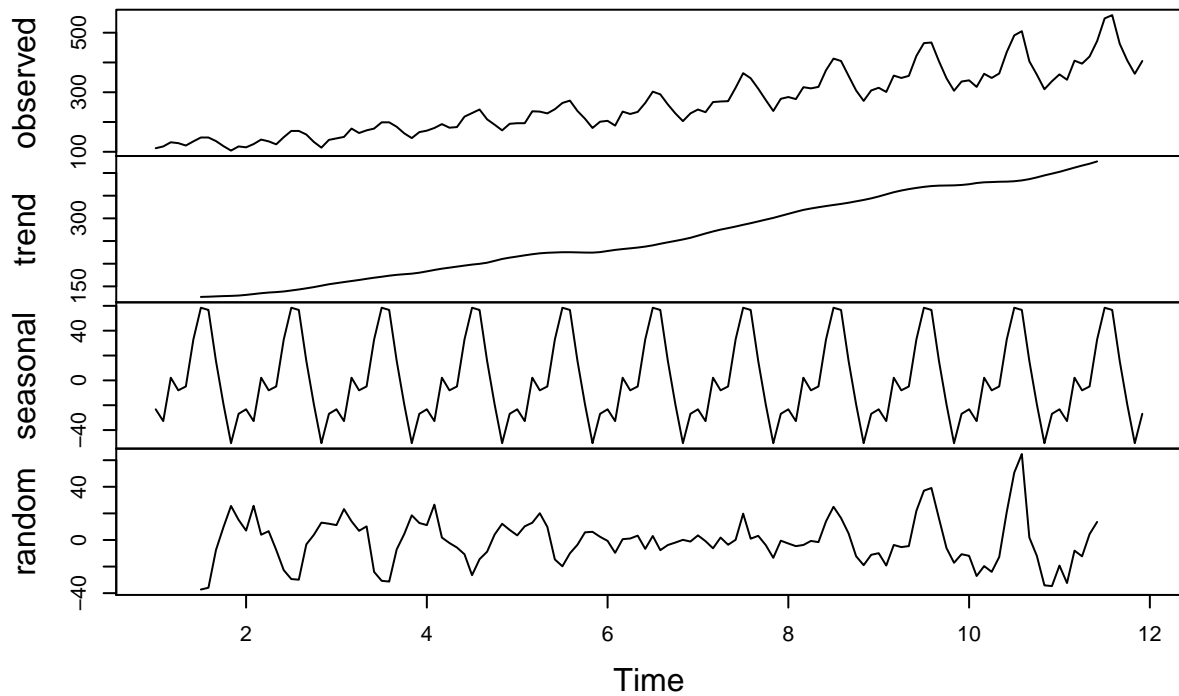
```
nTest = 12
```

```
nPred = 12
```

Creo el objeto “Serie temporal” con la librería `tseries` y su función `ts`. La frecuencia será 12 porque suponemos que la estacionalidad es anual. Utilizando la función `plot` junto a `descompose` se pueden ver la tendencia, la estacionalidad y lo que queda de la serie una vez eliminadas la tendencia y la estacionalidad.

```
serie.ts = ts(serie, frequency=12)  
plot(decompose(serie.ts))
```

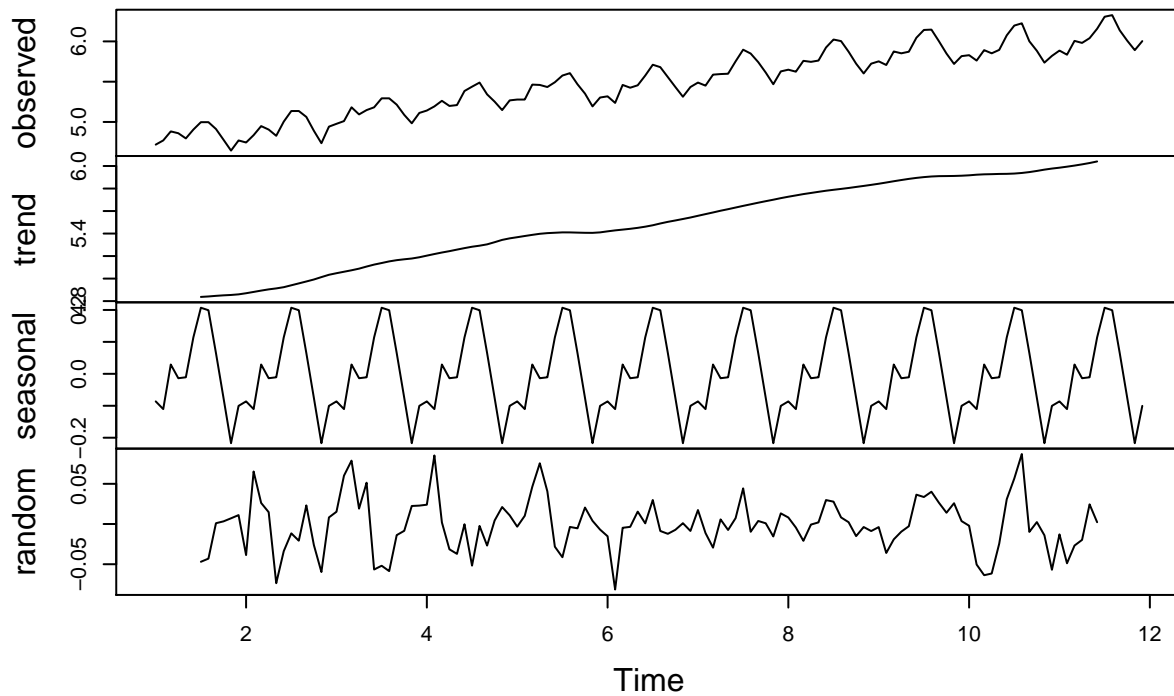
## Decomposition of additive time series



Vamos a construir la serie logarítmica (tanto el objeto ts como la serie original) para tratar de reducir la varianza en la serie resultante de restar la tendencia y estacionalidad. Buscamos que la serie sea estacionaria, y eso implica que no varíe ni en media ni en varianza.

```
serie.ts.log = log(serie.ts)
serie.log = log(serie)
plot(decompose(serie.ts.log))
```

## Decomposition of additive time series



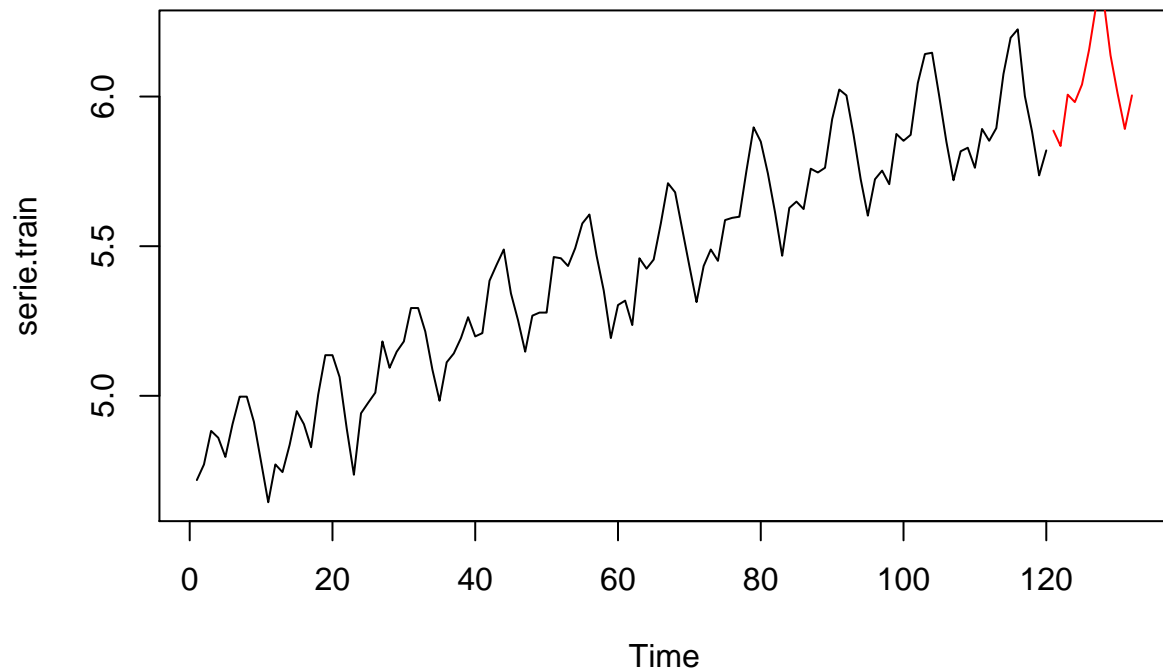
Vamos a seleccionar los datos de train y test del problema, entendiendo test como validación previa a la predicción del año 1960.

```
serie.train = serie.log[1:(length(serie.log) - nTest)]
tiempo.train = 1:length(serie.train)

serie.test = serie.log[(length(serie.log) - nTest+1):length(serie.log)]
tiempo.test = (length(serie.train)+1):length(serie.log)
```

A continuación los mostramos en una gráfica, resaltando en rojo los datos que se están tomando como test.

```
plot.ts(serie.train, xlim=c(1, tiempo.test[length(tiempo.test)]))
lines(tiempo.test, serie.test, col="red")
```



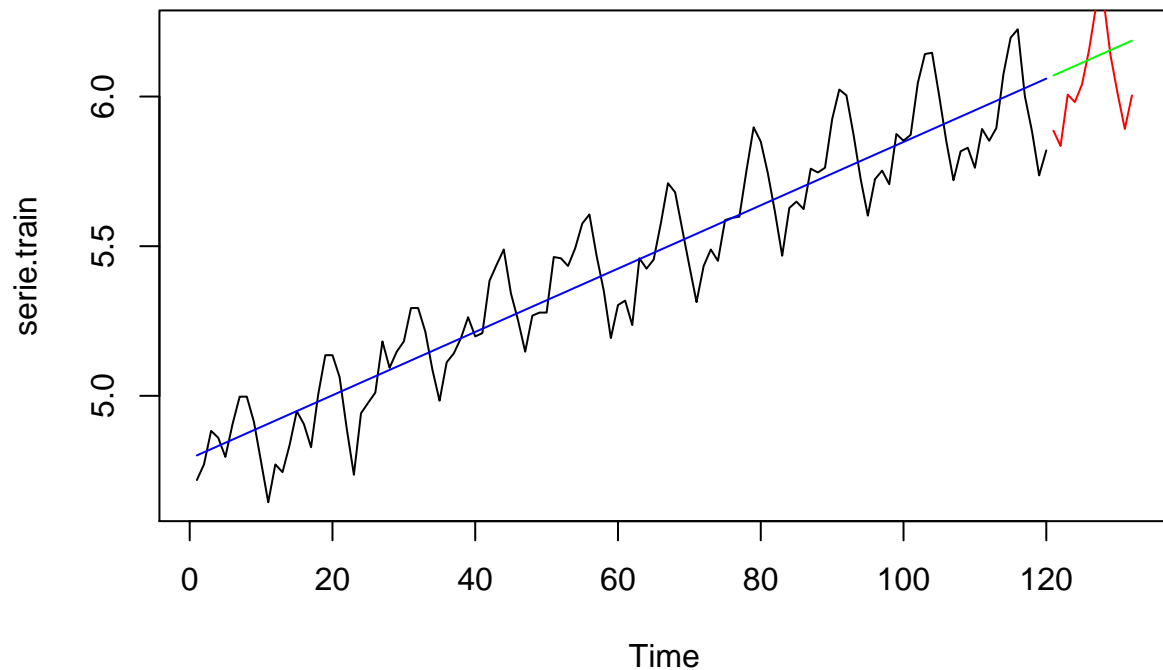
Lo siguiente que se va a hacer es estimar la tendencia. Se va a utilizar un modelo lineal en principio dado que por el aspecto de la serie temporal da la sensación de que puede generalizar bien.

```
parametros.H1 = lm(serie.train ~ tiempo.train)

tendencia.train.H1 = parametros.H1$coefficients[1]+tiempo.train*parametros.H1$coefficients[2]
tendencia.test.H1 = parametros.H1$coefficients[1]+tiempo.test*parametros.H1$coefficients[2]
```

Y a continuación, mostramos la tendencia estimada en la misma gráfica que la serie temporal, para comprobar cómo se ajusta.

```
plot.ts(serie.train, xlim=c(1, tiempo.test[length(tiempo.test)]))
lines(tiempo.test, serie.test, col="red")
lines(tiempo.train, tendencia.train.H1, col="blue")
lines(tiempo.test, tendencia.test.H1, col="green")
```



Vamos a validar que el modelo es correcto, para apoyarnos en algo más que una hipótesis. Utilizaremos el test de Jarque Bera sobre los residuos que han quedado de generar el modelo lineal y los que quedan en el test.

```
JB.train = jarque.bera.test(parametros.H1$residuals)
JB.test = jarque.bera.test(tendencia.test.H1-serie.test)
```

```
JB.train
```

```
##
##  Jarque Bera Test
##
## data:  parametros.H1$residuals
## X-squared = 1.755, df = 2, p-value = 0.4158
```

```
JB.test
```

```
##
##  Jarque Bera Test
##
## data:  tendencia.test.H1 - serie.test
## X-squared = 0.87957, df = 2, p-value = 0.6442
```

Como el p-value resultante del test de Jarque Bera es superior a 0.05 no se puede afirmar con suficiente confianza que los residuos no sigan una distribución normal, por lo que consideramos normales los errores del modelo lineal para train y para test.

A continuación vamos a comparar las medias del error, para comprobar si el error que se produce en la parte de train es similar al que ocurre en la parte de test. Para ello utilizamos el test de Student.

```
TT = t.test(c(parametros.H1$residuals, tendencia.test.H1-serie.test))
TT
```

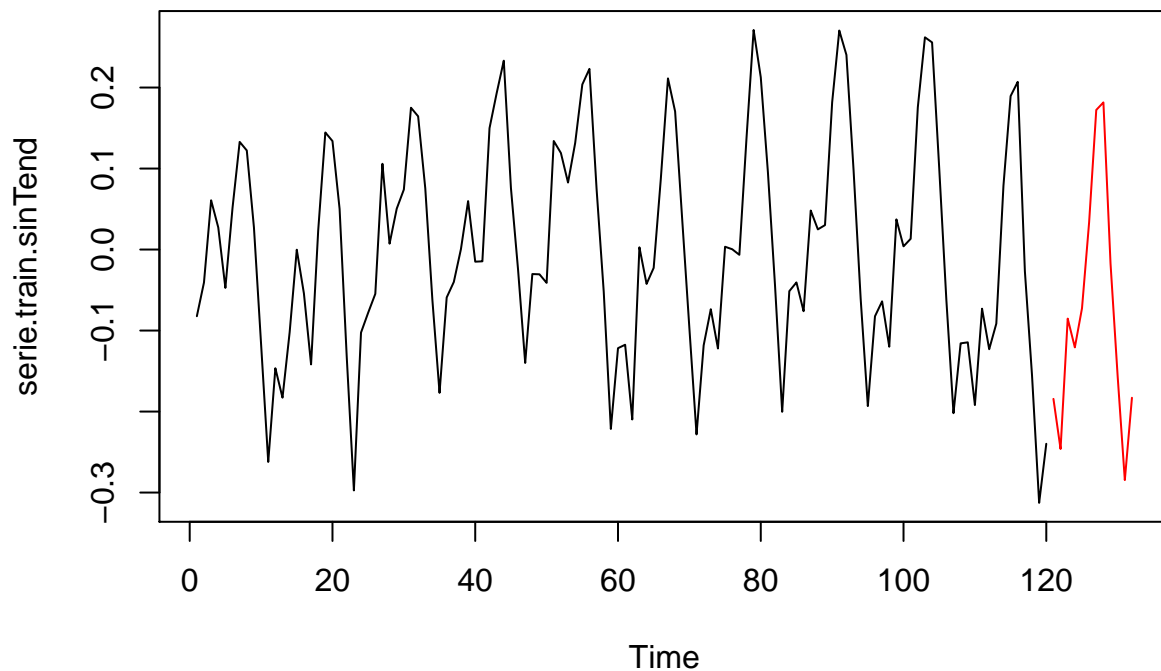
```
##
## One Sample t-test
##
## data: c(parametros.H1$residuals, tendencia.test.H1 - serie.test)
## t = 0.61219, df = 131, p-value = 0.5415
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.01628561 0.03088222
## sample estimates:
## mean of x
## 0.007298304
```

El test de Student indica que la media de los errores es casi 0 (0.0073), y el p-value por encima de 0.05 indica que no hay una desviación significativa con respecto a esta media. Por esto, se puede considerar que no hay una diferencia estructural que invalide el modelo lineal en este caso.

El siguiente paso es eliminar la tendencia de la serie inicial, tanto en entrenamiento como en test. Y comprobamos en una gráfica qué aspecto tiene la serie sin esta tendencia.

```
serie.train.sinTend = serie.train - tendencia.train.H1
serie.test.sinTend = serie.test - tendencia.test.H1

plot.ts(serie.train.sinTend, xlim=c(1, tiempo.test[length(tiempo.test)]))
lines(tiempo.test, serie.test.sinTend, col="red")
```



Una vez hecho esto, procedemos a eliminar la estacionalidad de la serie resultante. Inicialmente asumimos una estacionalidad de 12 valores (anual).

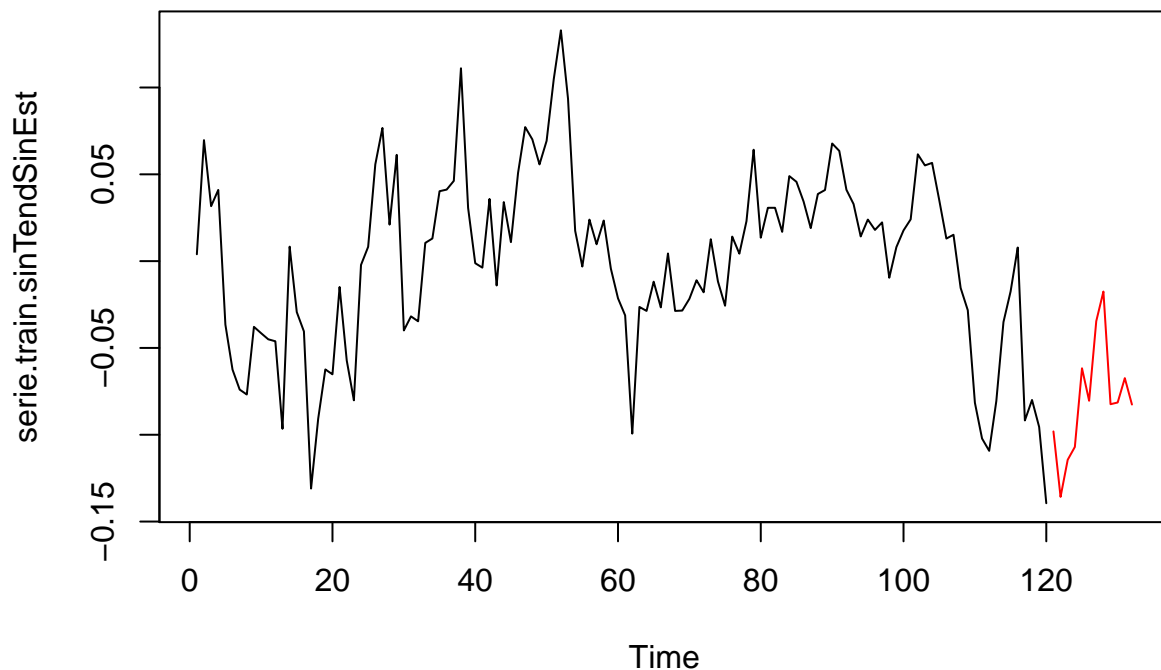
```
k = 12
estacionalidad = decompose(serie.ts.log)$seasonal[1:k]
estacionalidad

## [1] -0.08627390 -0.11042950  0.02915584 -0.01376615 -0.01080726
## [6]  0.11403832  0.20689984  0.19914832  0.06503613 -0.07542627
## [11] -0.21722154 -0.10035385
```

Aquí se pueden observar los 12 valores que definen la estacionalidad. Para eliminar la estacionalidad de nuestra serie hay que ir eliminando estos 12 valores de forma periódica a lo largo de la serie.

```
# Aprovecho el reciclaje de R para no tener que crear una variable auxiliar
serie.train.sinTendSinEst = serie.train.sinTend - estacionalidad
serie.test.sinTendSinEst = serie.test.sinTend - estacionalidad

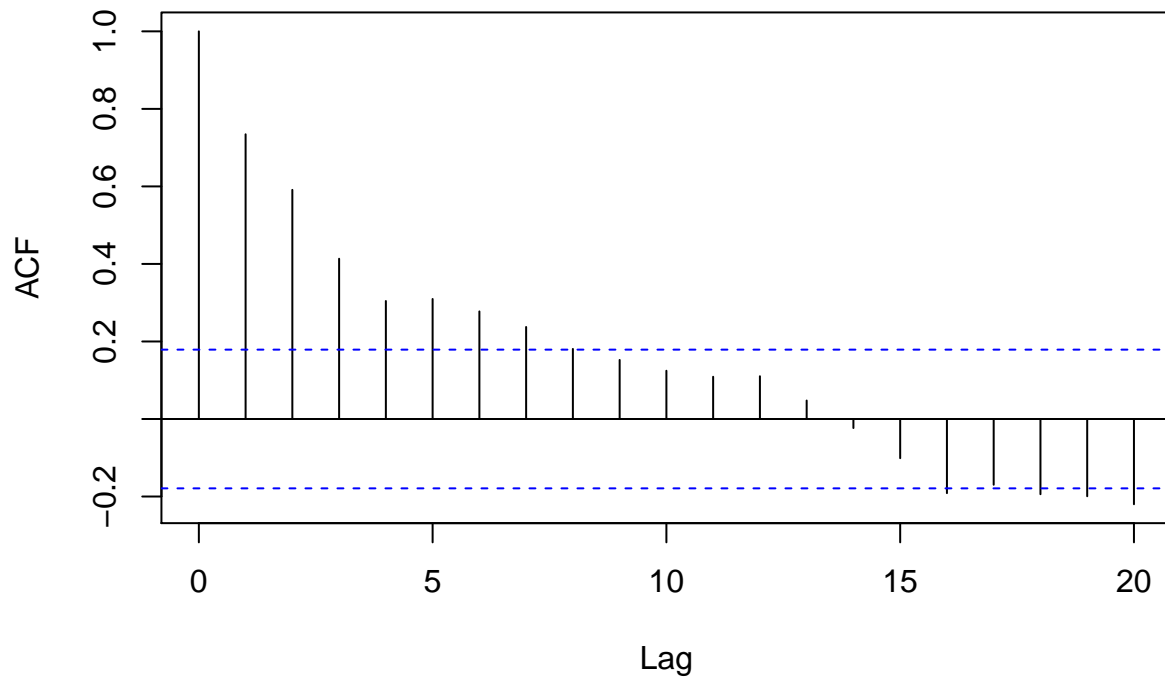
plot.ts(serie.train.sinTendSinEst, xlim=c(1, tiempo.test[length(tiempo.test)]))
lines(tiempo.test, serie.test.sinTendSinEst, col="red")
```



Aquí se puede observar ya limpiamente la serie temporal sin la tendencia y sin la estacionalidad. El siguiente paso es comprobar si la serie es o no estacionaria. Podemos comenzar por un test ACF.

```
acf(serie.train.sinTendSinEst)
```

## Series serie.train.sinTendSinEst



Como no se puede observar claramente que sea estacionaria vamos a hacer el test de Dickey-Fuller aumentado para asegurarnos.

```
adf.series.train = adf.test(serie.train.sinTendSinEst)
adf.series.train
```

```
##
## Augmented Dickey-Fuller Test
##
## data: serie.train.sinTendSinEst
## Dickey-Fuller = -1.8407, Lag order = 4, p-value = 0.6427
## alternative hypothesis: stationary
```

El p-value es 0.64, por lo que al buscar una confianza del 95% no se puede asegurar que la serie sea estacionaria. Lo normal es que la serie no sea estacionaria en media, para hacerla estacionaria vamos a diferenciarla. Cuando lo hayamos hecho volveremos a pasarle el test.

```
serie.train.sinTendSinEstDif = diff(serie.train.sinTendSinEst)
serie.test.sinTendSinEstDif = diff(serie.test.sinTendSinEst)
```

```
adf.series.train.2 = adf.test(serie.train.sinTendSinEstDif)
```

```
## Warning in adf.test(serie.train.sinTendSinEstDif): p-value smaller than
## printed p-value
```

```
adf.series.train.2
```

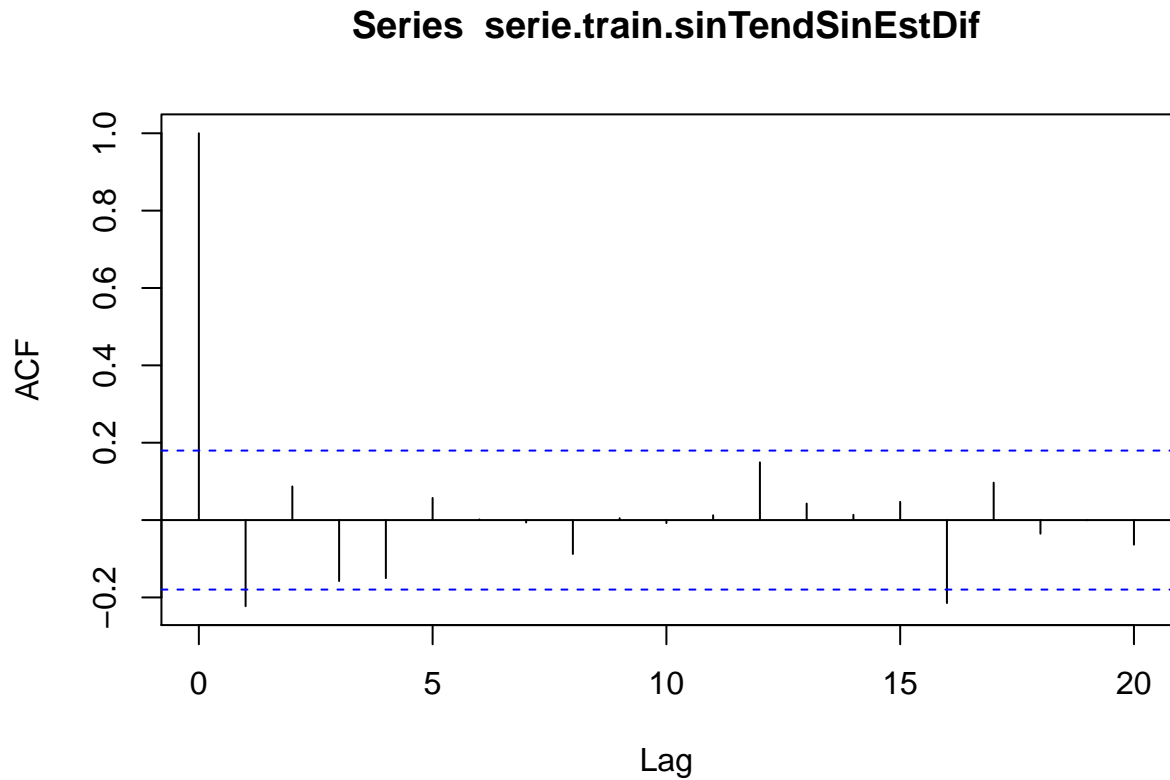
```
##
## Augmented Dickey-Fuller Test
```



```
##  
## data: serie.train.sinTendSinEstDif  
## Dickey-Fuller = -6.2153, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary
```

Ahora que el p-value se encuentra por debajo de 0.05 sí que se puede asumir que la serie es estacionaria. Si comprobamos con un nuevo test ACF y un test PACF las gráficas resultantes de esta serie obtenemos algo que encaja mucho más con una serie estacionaria.

```
acf(serie.train.sinTendSinEstDif)
```



```
pacf(serie.train.sinTendSinEstDif)
```

**Series serie.train.sinTendSinEstDif**

