ANÁLISIS DE SERIES TEMPORALES

Universidad del Valle

David Arango Londoño
Docente e Investigador
darango.ccafs@gmail.com

Análisis de Series de Tiempo

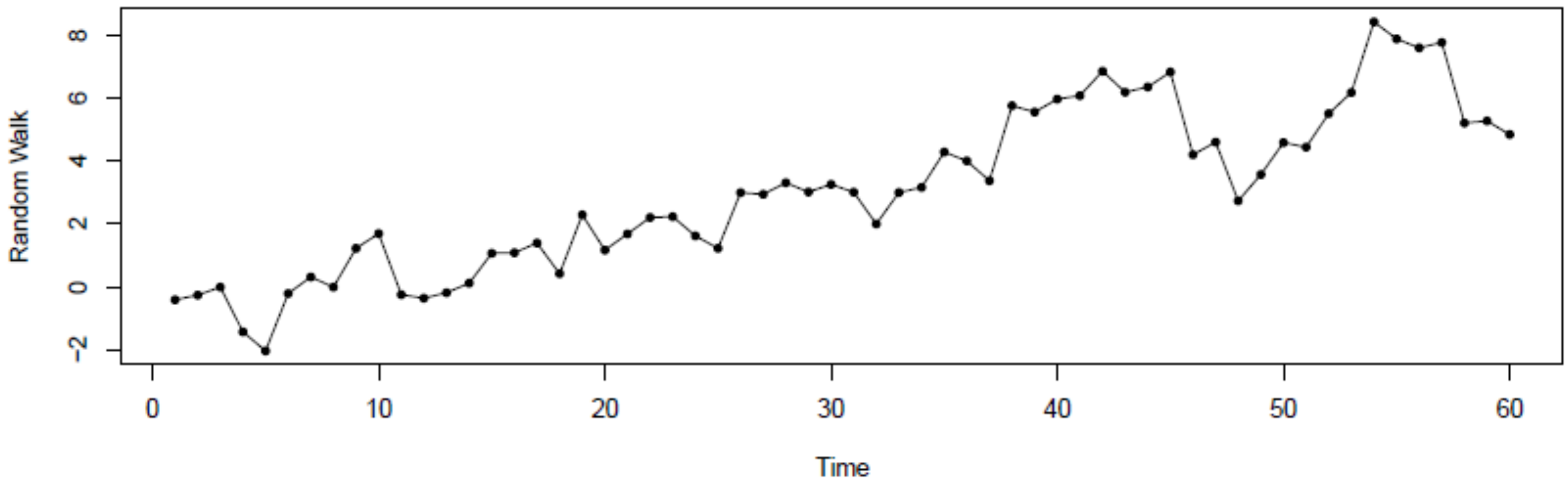
Series de Tiempo

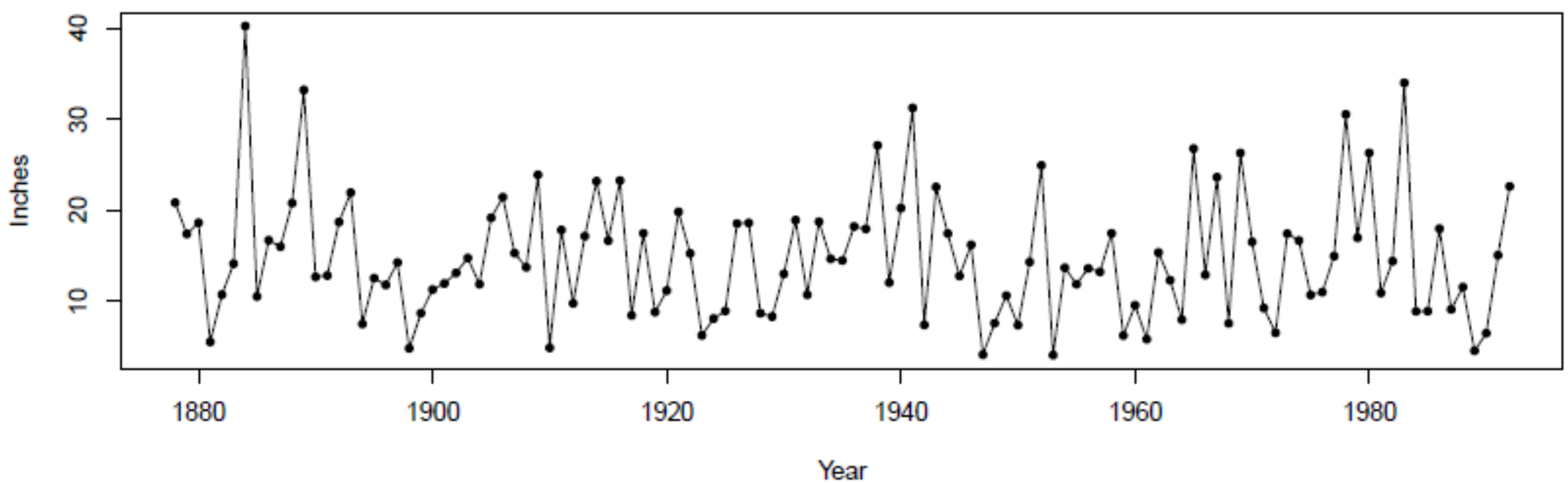
Se llama serie de tiempo al conjunto de datos que surgen del registro de observaciones de una variable cuantitativa en función del tiempo.

Las series pueden ser de dos tipos:

- *Estacionarias:* el valor medio permanece constante aunque se presenta variaciones en torno a ese valor.
- *Evolutivas (No estacionarias):* el valor medio de la serie cambia a través del tiempo.

Series de Tiempo





Series de Tiempo

Las series de tiempo están conformadas por cuatro componentes:

- *Tendencia:* es el comportamiento de una serie a largo plazo donde el valor medio cambia a través del tiempo.

- *Estacionalidad:* representan las fluctuaciones que se repiten periódicamente en intervalos de tiempo de igual amplitud.

- *Ciclo:* representa el patrón de comportamiento que se repite en periodos de diferente duración. Suele ser más irregular que las variaciones estacionales.

- *Ruido:* puede ser el producto de variaciones naturales aleatorias. Se producen en la serie de forma aislada y no permanente de modo que es prácticamente imposible preverlas.

Extracción de Señales

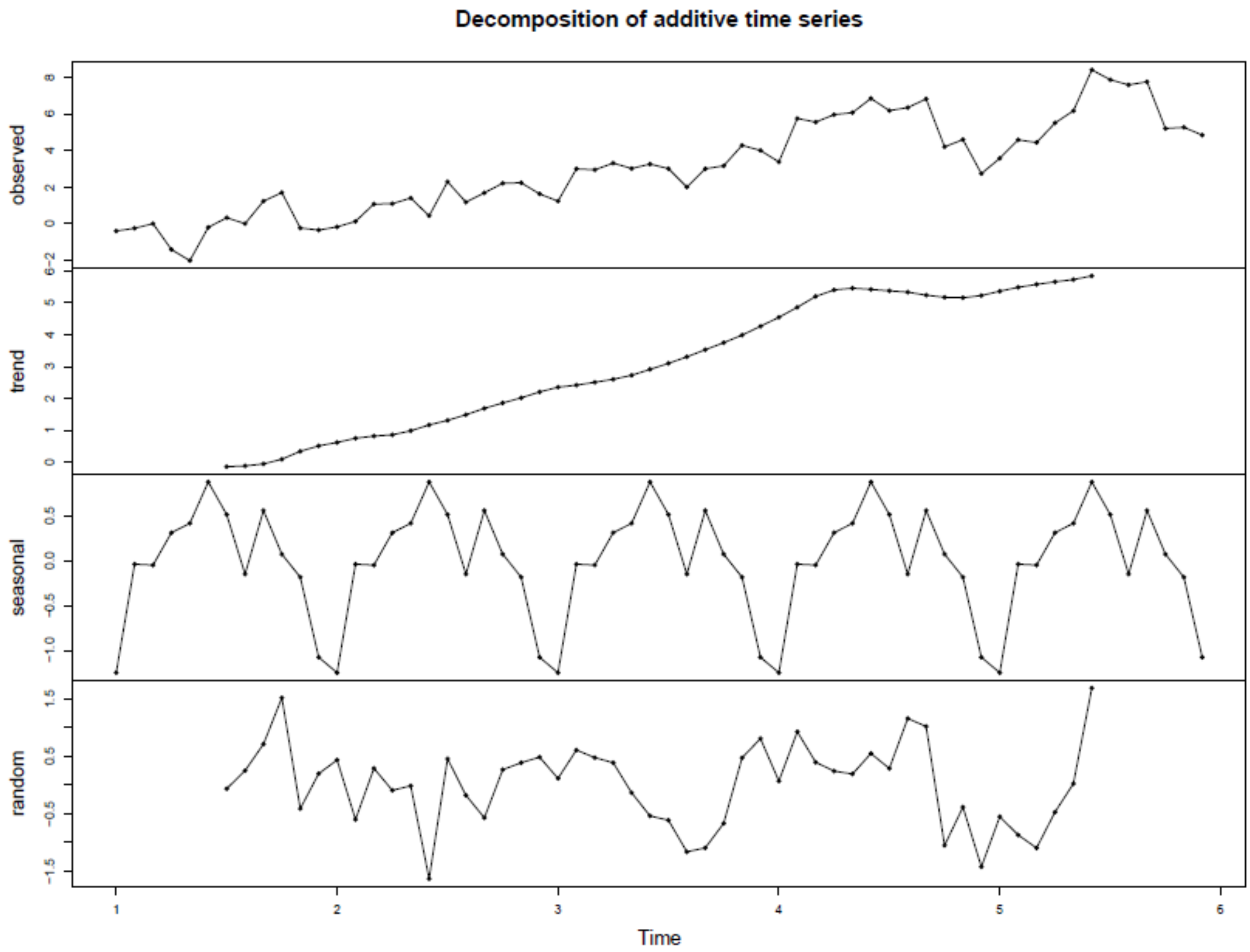
Una forma de abordar el estudio de una serie temporal es mediante la ***descomposición aditiva*** de la serie. El esquema aditivo se expresa de la siguiente forma:

$$Y_t = T_t + S_t + I_t$$

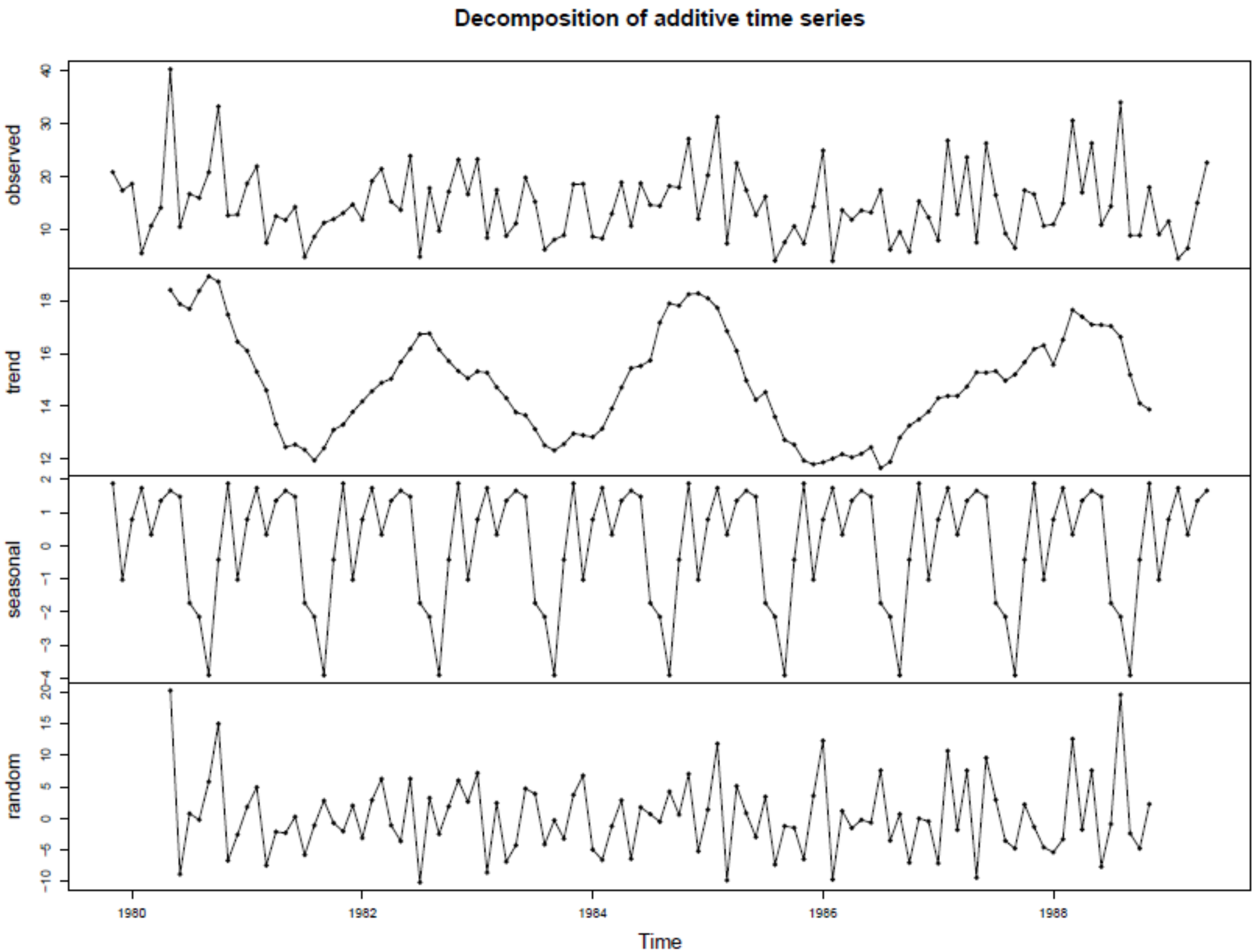
Valor observado = Tendencia + Estacionalidad + Irregular

El componente irregular marca la diferencia entre los alisados exponenciales y las técnicas de descomposición de la serie temporal. Por medio de la descomposición de señales se puede modelar de forma especifica la componente irregular, mientras que en los alisados se asume un carácter determinista de la serie.

Series de Tiempo



Series de Tiempo



Alisado exponencial

El método del *alisado exponencial* es una aproximación determinista al tratamiento de series temporales. Los alisados se emplean para predecir nuevos valores de la serie.

Estos modelos permiten ajustar niveles y comportamientos tendenciales y estacionales que evolucionan en el tiempo, de manera que las observaciones mas recientes tienen mas peso en el ajuste que las mas alejadas.

Hay tres tipos de alisado exponencial:

- Alisado exponencial simple
- Alisado de Holt
- Alisado de Holt-Winters

Alisado exponencial Simple

Se emplea para series *sin tendencia ni estacionalidad*.

El modelo del alisado exponencial simple depende de un parámetro *alpha*, que modula la importancia que tienen las observaciones pasadas sobre el presente. Su valor oscila entre 0 y 1.

Si alpha toma un valor *próximo a 0* las predicciones a lo largo de la serie son muy *similares entre si* y se modifican poco con la nueva información. El caso extremo se produce cuando *alpha es cero*, lo que implica que la predicción es una *constante a lo largo del tiempo*.

Si alpha, por el contrario, toma un valor *próximo a 1* la predicción se va adaptando al último valor observado, por lo que se puede decir que los valores alejados en el tiempo no tienen mucha influencia sobre la predicción.

Alisado de Holt

Se emplea para series *con tendencia y sin estacionalidad*.

El modelo depende de dos parámetros, *alpha* y *beta*. El parámetro beta modula la importancia que tienen las observaciones pasadas sobre la pendiente estimada en tiempo ***t***. Al igual que para alpha, los valores de beta oscilan entre 0 y 1.

Si beta toma un valor próximo a 0 entonces la pendiente es constante o casi constante, es decir, cambia poco a lo largo de la serie temporal.

Si beta toma un valor próximo a 1, la predicción de la pendiente se va adaptando al último valor observado y las observaciones de las pendientes mas alejadas en el tiempo no tienen mucha influencia sobre la predicción.

Alisado de Holt-Winters

Se emplea para series *con tendencia y estacionalidad*.

Depende de tres parámetros: *alpha*, *beta* y *gamma*.

El parámetro gamma modula la importancia que tienen las observaciones pasados sobre la predicción en el tiempo ***t***. Gamma oscila entre 0 y 1.

Si gamma es 0 la predicción en tiempo *t* va a tomar un valor constante que va a depender de todas las observaciones pasadas dentro de ese mismo periodo.

Si gamma es 1 la predicción en tiempo *t* va a depender solamente de la observación hecha en tiempo *t-s*, siendo s la frecuencia en la cual se repite la estacionalidad (por ejemplo *s*=12 para observaciones mensuales).

Ejemplo en R

Alisado Exponencial Simple

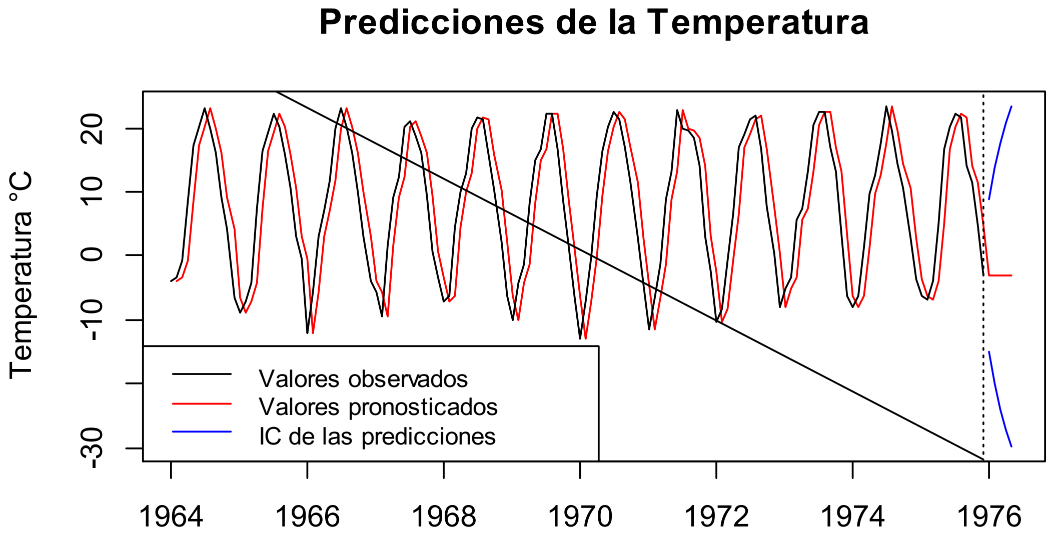
```
data(tempdub)
tempdub<-(tempdub-32)*5/9
plot(tempdub, ylab="Temperatura (°C)", xlab="")

alisim1 <- HoltWinters(tempdub, gamma=FALSE, beta=FALSE)
alisim1
```

El parámetro alpha que minimiza el error cuadrático medio 0.9999339, muy cercano a 1. Esto indica que de acuerdo al modelo, la predicción en el tiempo *t* depende poco de las observaciones pasadas y se parecerá mucho al último valor de la serie.

El coeficiente $a = -3.221693$ representa el valor de la temperatura en el instante $T+1$. Como nuestro modelo no tiene ni tendencia ni estacionalidad, la predicción será aplicable a cualquier instante en el futuro, por lo que las predicciones a futuro son deterministas.

Ejemplo en R

```
pred.aliholt <- predict(alisim1, n.ahead=5, prediction.interval=TRUE)
plot(alisim1, pred.aliholt, ylab="Temperatura °C", xlab="", main="Predicciones de la Temperatura")
labs <- c("Valores observados", "Valores pronosticados",  "IC de las predicciones")
legend("bottomleft", lty=rep(1,3), col=c("black", "red", "blue"), legend=labs, cex=0.8)
```



Ejemplo en R

Alisado de Holt

```
alisim2 <- HoltWinters(tempdub, gamma=FALSE)
alisim2
```

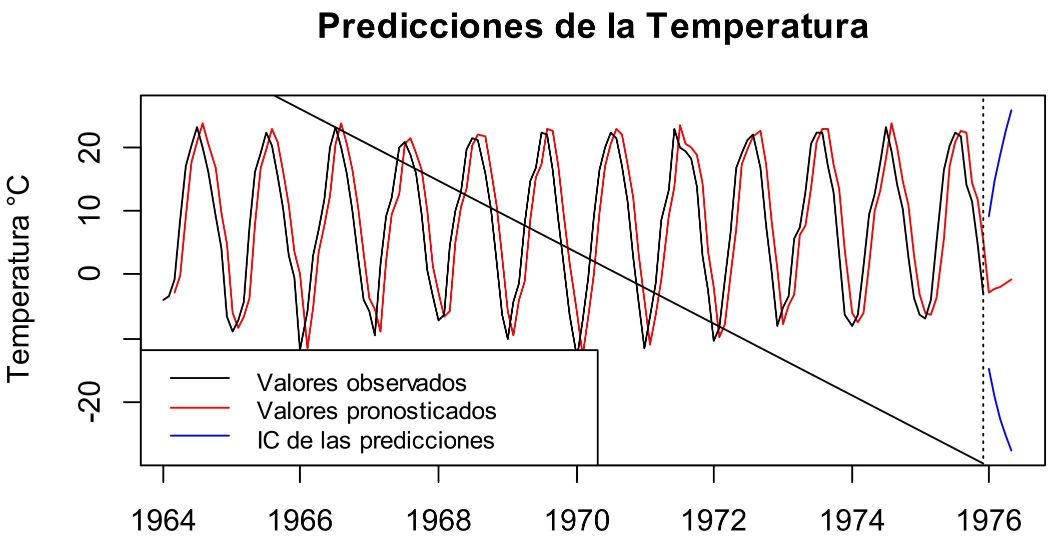
El coeficiente *alpha=1* indica que la predicción en el tiempo *t* (en ausencia de tendencia) depende exclusivamente de lo observado en el tiempo *t-1*.

El coeficiente *beta = 0.001146* indica que la pendiente cambia poco a lo largo de la serie temporal.

Nuestra predicción en el tiempo *T+1* será de -3.2222°C (coeficiente *a*) y la tendencia vendrá determinada por una pendiente de b=0.47034.

Ejemplo en R

```
pred.aliholt <- predict(alisim2, n.ahead=5, prediction.interval=TRUE)
plot(alisim2, pred.aliholt, ylab="Temperatura °C", xlab="", main="Predicciones
de la Temperatura")
labs <- c("Valores observados", "Valores pronosticados", "IC de las
predicciones")
legend("bottomleft", lty=rep(1,3), col=c("black", "red", "blue"), legend=labs,
cex=0.8)
```



Ejemplo en R

Alisado de Holt-Winters

```
alisim3 <- HoltWinters(tempdub)
alisim3
```

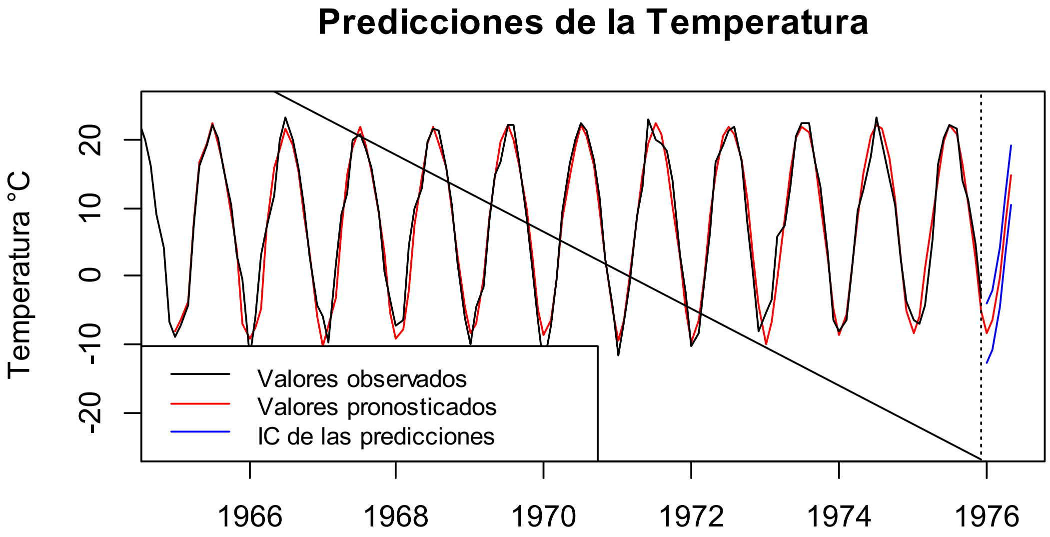
El coeficiente *alpha=0.02021* indica que la predicción en el tiempo *t* depende de las observaciones pasadas.

El coeficiente *beta = 0.1708* indica que la pendiente cambia levemente a lo largo de la serie temporal.

El coeficiente *gamma = 0.2274* indica que las predicciones dependen de las observaciones obtenidas en el periodo estacional.

Ejemplo en R

```
pred.aliholt <- predict(alisim3, n.ahead=5, prediction.interval=TRUE)
plot(alisim3, pred.aliholt, ylab="Temperatura °C", xlab="", main="Predicciones de la Temperatura", ylim=c(-25,25))
labs <- c("Valores observados", "Valores pronosticados", "IC de las predicciones")
legend("bottomleft", lty=rep(1,3), col=c("black", "red", "blue"), legend=labs, cex=0.8)
```



Modelos ARIMA

Existe una clase de modelos paramétricos que permite modelar series temporales estacionarias y no estacionarias: los modelos *ARIMA* (*Autoregressive Integrated Moving Average*). Estos modelos incluyen los modelos *AR* (*Autoregressive*), *MA* (*Moving Averages*) y *ARMA* (*Autoregressive Moving Averages*) para series estacionarias, y los modelos integrados para las series no estacionarias.

Para ajustar modelos *ARIMA* hay que seguir una serie de pasos. Esto es lo que se conoce como metodología *Box-Jenkins*. Los pasos a seguir son:

1. Identificación del modelo ARIMA.
2. Estimación de los parámetros.
3. Validación del modelo. Si el modelo no es valido hay que volver al punto 1. Si el modelo es valido se pasa al siguiente punto.
4. Predicción de nuevos valores.

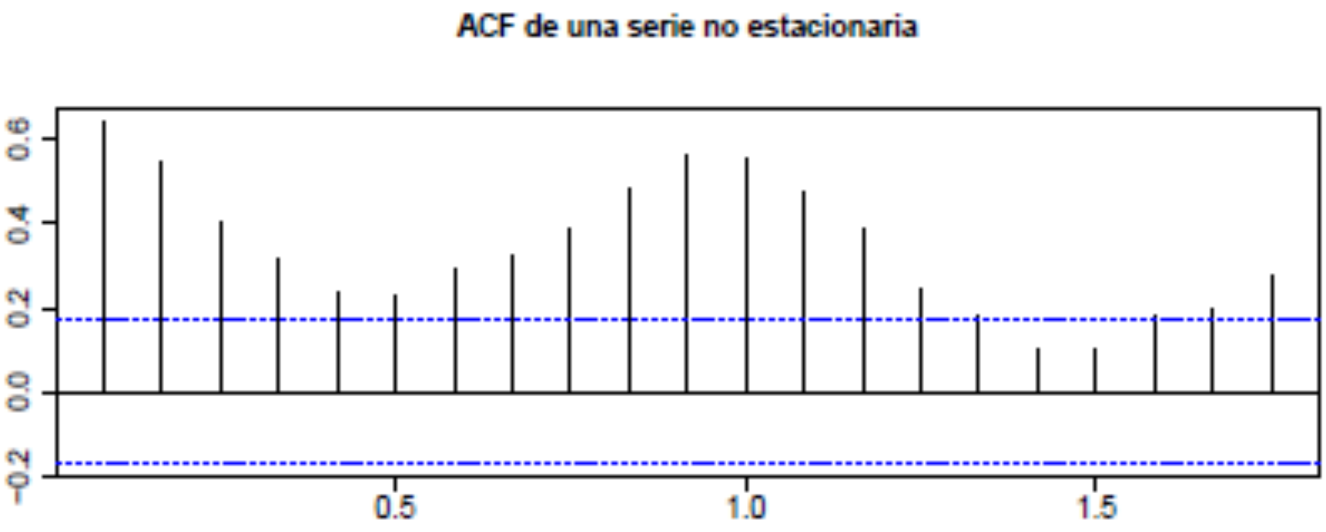
Series de Tiempo

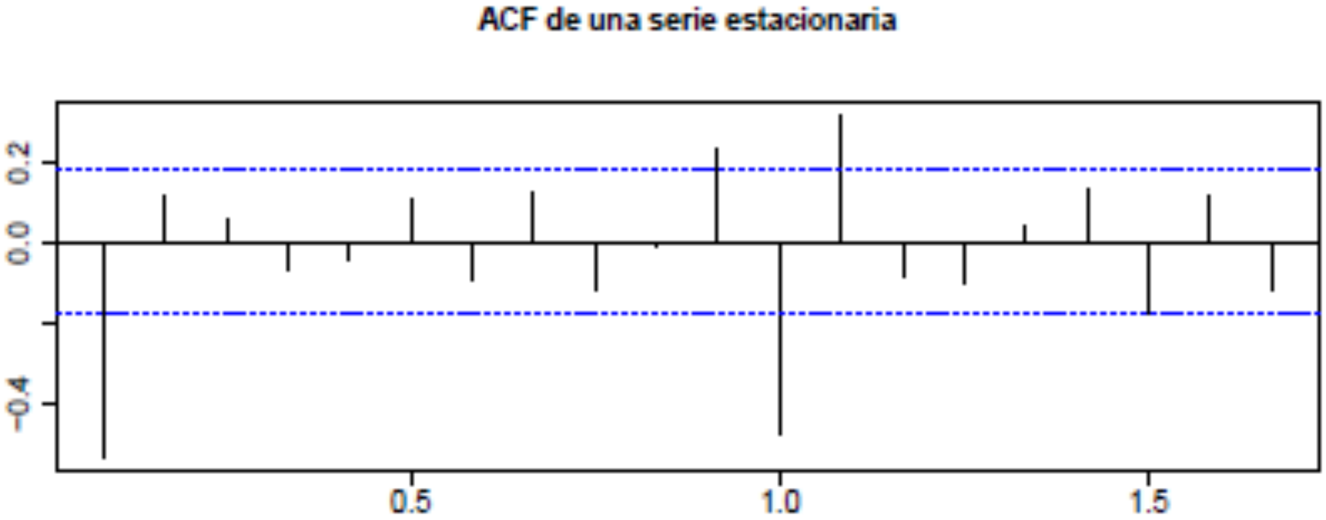
Una de las restricciones impuestas habitualmente a una serie de tiempo es que sea ***estacionaria***. Se dice que una serie es estacionaria si su media y su varianza son constantes en el tiempo.

Para probar estacionariedad, se emplea la función de autocorrelación ***FAC*** la cual mide el grado de dependencia lineal entre observaciones próximas.

En la FAC de una serie estacionaria, la dependencia temporal debe acercarse a cero rápidamente. Sin embargo, un decrecimiento lento de las autocorrelaciones indica que la serie no es estacionaria.

Series de Tiempo





Identificación del Modelo

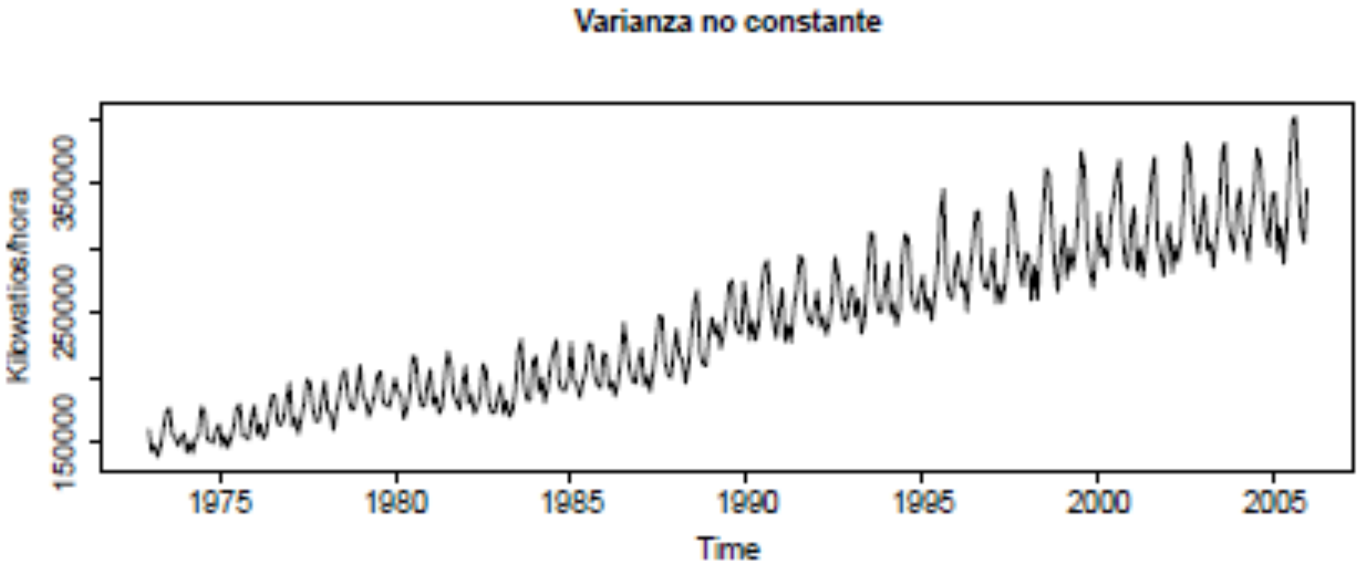
Transformación de la variable respuesta:

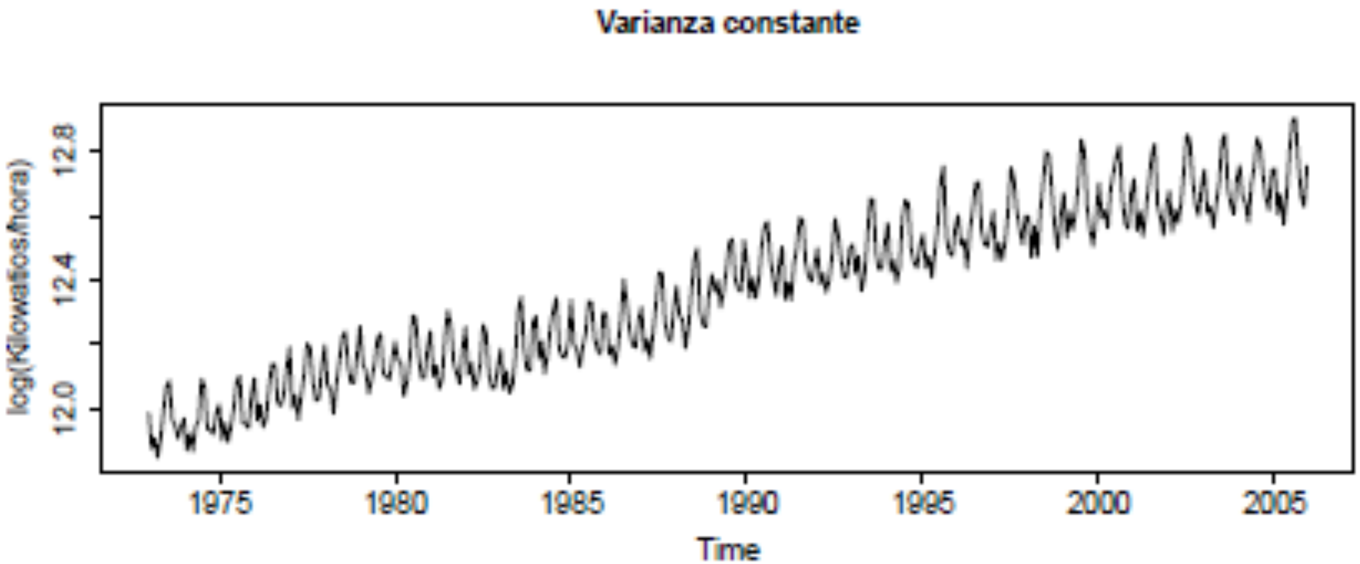
La hipótesis de varianza constante se exige en las condiciones de estacionariedad, pero es frecuente observar que la varianza aumenta con el Nivel (media) de la serie. En este caso la transformación con logaritmo neperiano ayuda a homogeneizar su comportamiento.

Para decidir si hace falta la transformación con el logaritmo neperiano se comparan los gráficos de la serie original y transformada.

Otras transformaciones de la variable respuesta también son posibles, como la familia de transformaciones Box-Cox, que incluye potencias y logaritmos.

Identificación del Modelo





Identificación del Modelo

Convertir la serie en estacionaria:

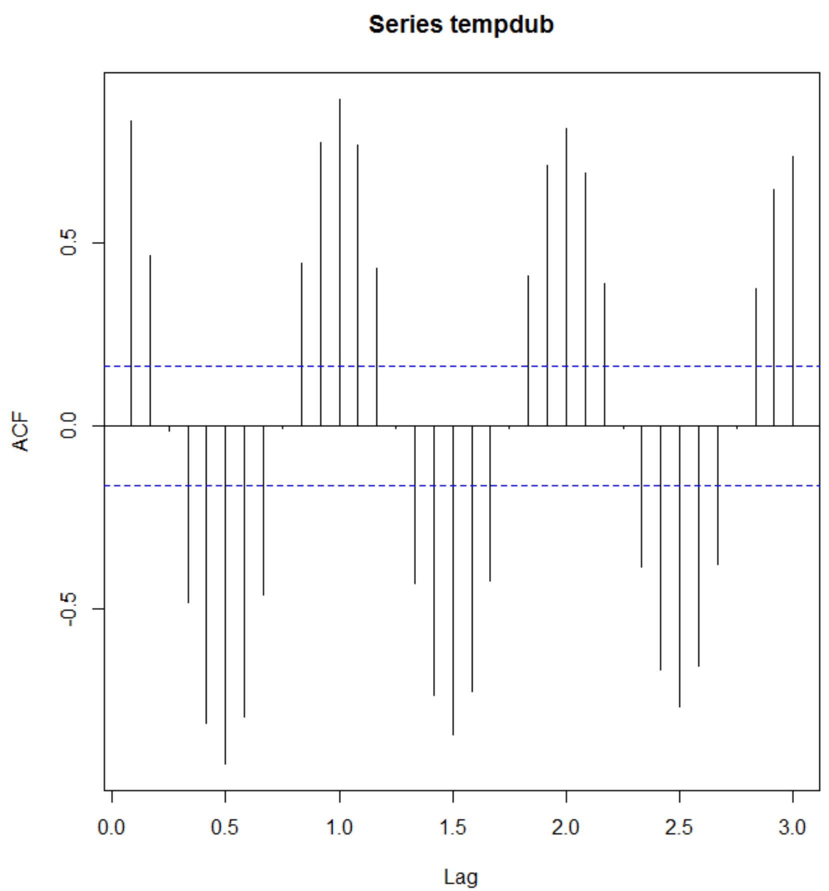
El gráfico de una serie estacionaria no debe presentar comportamientos tendenciales o estacionales. Si presenta alguno de estos comportamientos, se puede tomar ***d*** diferencias regulares $(Y_t - Y_{t-1})$ y/o ***D*** estacionales $(Y_t - Y_{t-s})$ para intentar convertirla en estacionaria.

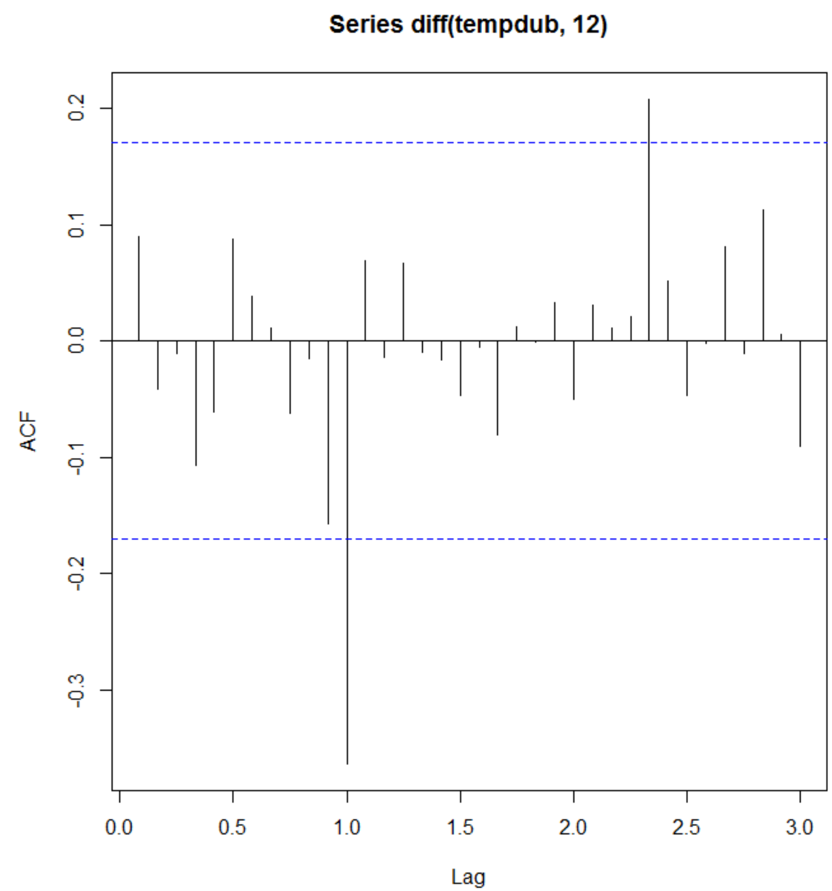
Se debe seleccionar aquella diferenciación que minimice la varianza.

```
var(tempdub)
var(diff(tempdub,1)) #d=1
var(diff(tempdub,2)) #d=2
var(diff(tempdub,3)) #d=3
var(diff(tempdub,12)) #d=0 y D=1
var(diff(diff(tempdub,1),12)) #d=1 y D=1
var(diff(diff(tempdub,2),12)) #d=2 y D=1
var(diff(diff(tempdub,3),12)) #d=3 y D=1
```

Identificación del Modelo

Una vez elegido el grado de diferenciación se debe verificar que la serie sea estacionaria mediante la FAC.





Identificación del Modelo

Orden de los polinomios autorregresivos y de medias móviles de la estructura regular y estacional:

Si tenemos una serie estacionaria o una vez que hemos convertido a estacionaria una serie no estacionaria, lo siguiente es definir el orden de los modelos autorregresivos (***AR***) y de medias móviles (***MA***) de la estructura regular y estacional de la serie.

Para predecir el valor de la variable en una serie estacionaria podemos utilizar la siguiente información:

Los valores pasados de la serie: $Y_1, Y_2, \ldots, Y_{t-1}$
Las innovaciones pasadas: $a_1, a_2, \ldots, a_{t-1}$

Dependiendo de que para predecir Y_t se utilicen los valores pasados de la serie o las innovaciones pasadas, se tendrán los modelos ***AR*** o los modelos ***MA***, respectivamente.

Universidad del Valle

Identificación del Modelo

Modelos autorregresivos AR(p) :

En los modelos AR(p) cada observación Y_t depende linealmente de las *p* observaciones anteriores:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t$$

donde ϕ_j es el coeficiente que relaciona la observación Y_t con la observación Y_{t-j} y a_t es ruido blanco (los residuales del modelo tienen media cero, varianza constante y son independientes).

Identificación del Modelo

Modelos de medias móviles MA(q) :

En los modelos MA(q) cada observación Y_t depende linealmente de las q perturbaciones anteriores:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q \, a_{t-q}$$

donde θ_j es el coeficiente que relaciona la observación Y_t con la perturbación a_{t-j} y a_t es ruido blanco.

Identificación del Modelo

Modelos autorregresivos y de medias móviles ARMA(p,q) :

En los modelos ARMA(p,q) cada observación Y_t depende linealmente de las p anteriores y q últimas innovaciones:

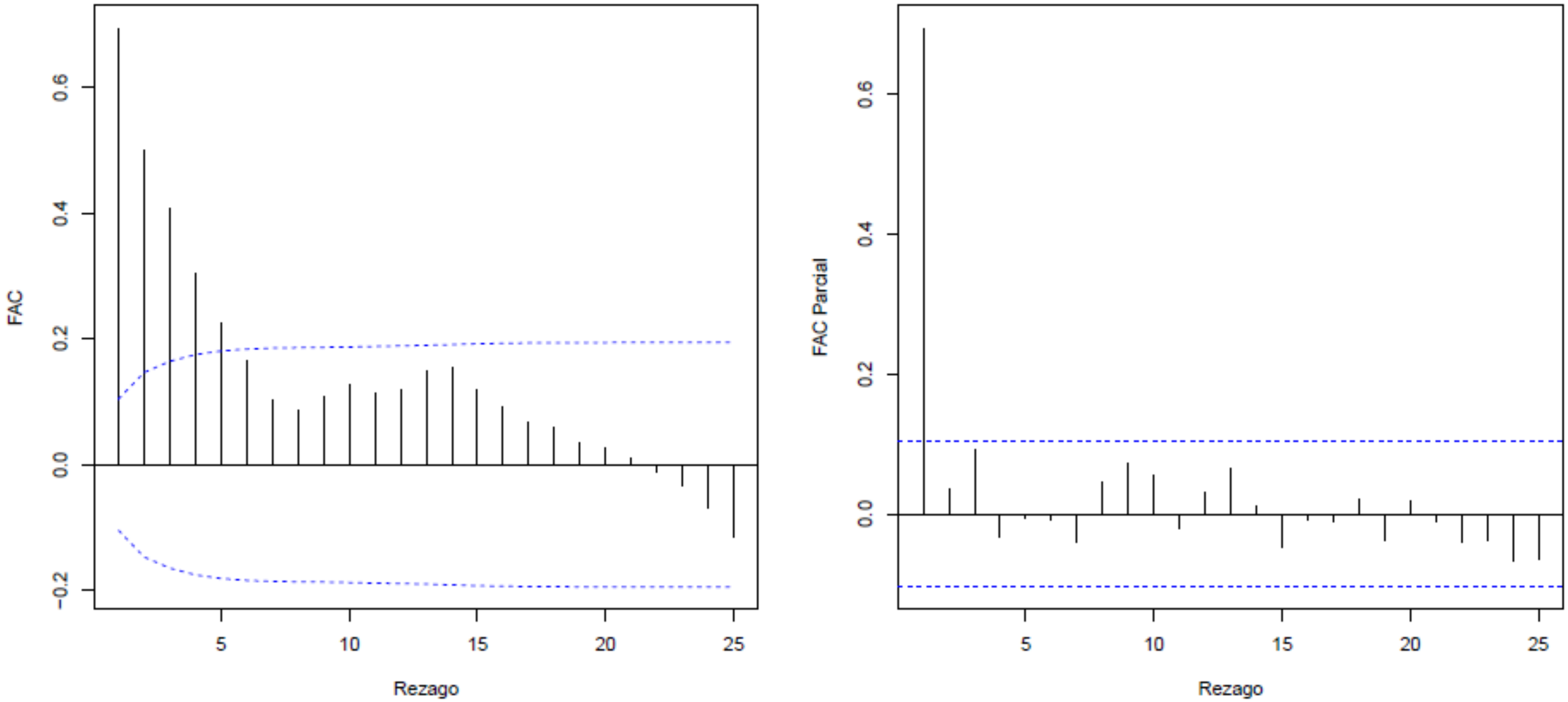
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t$$

En la práctica es muy difícil conocer de cuantas observaciones (p) e innovaciones (q) pasadas depende la observación t.

Para identificar la estructura de autoccorrelación de los datos se puede emplear la Función de Autocorrelación FAC para identificar procesos MA y la Función de Autocorrelación Parcial FACP para identificar procesos AR.

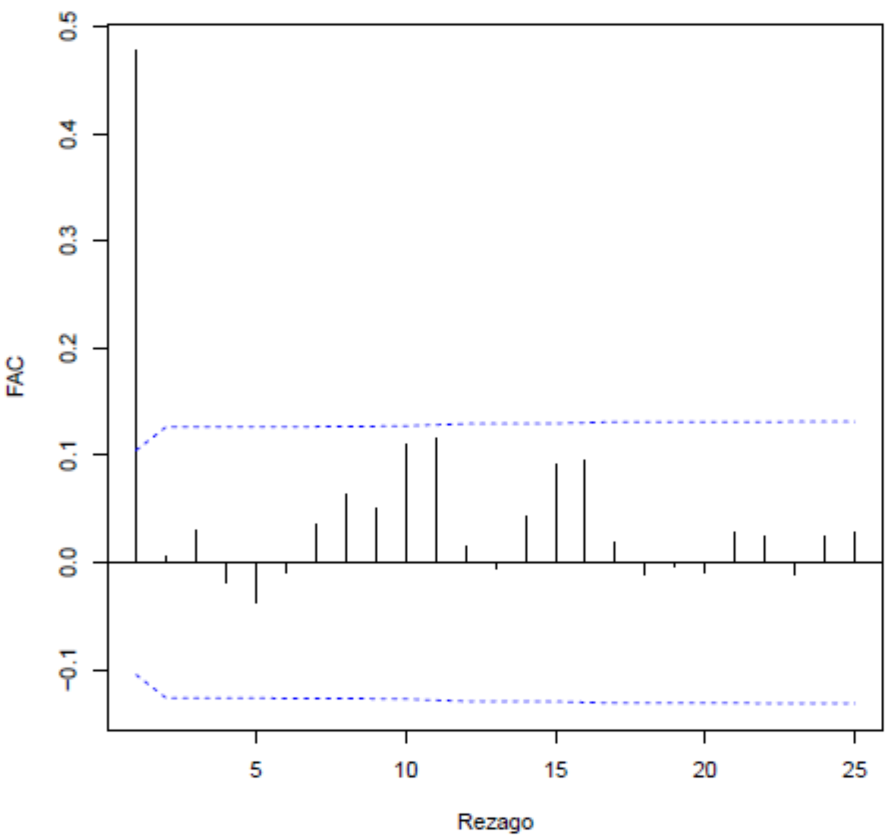
Identificación del Modelo

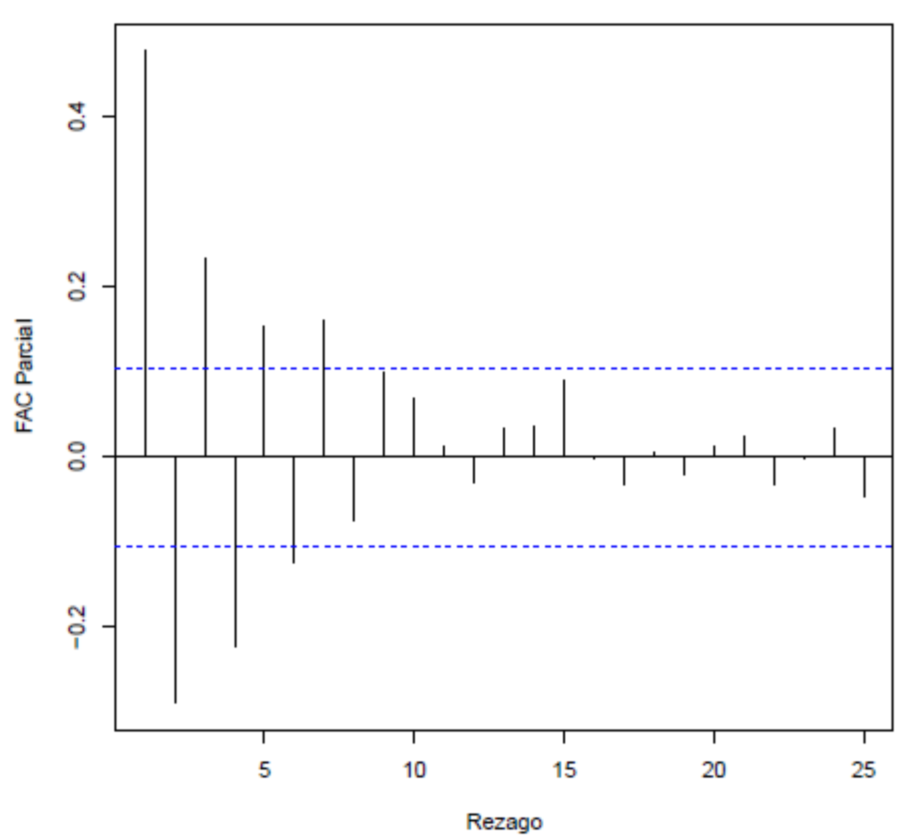
FAC y FACP para un modelo AR(1):
Sobresale solo la primera autocorrelación parcial y decae lentamente la FAC.



Identificación del Modelo

FAC y FACP para un modelo MA(1):
Sobresale solo la primera autocorrelación y decae lentamente la FACP.





Identificación del Modelo

Modelos estacionarios estacionales:

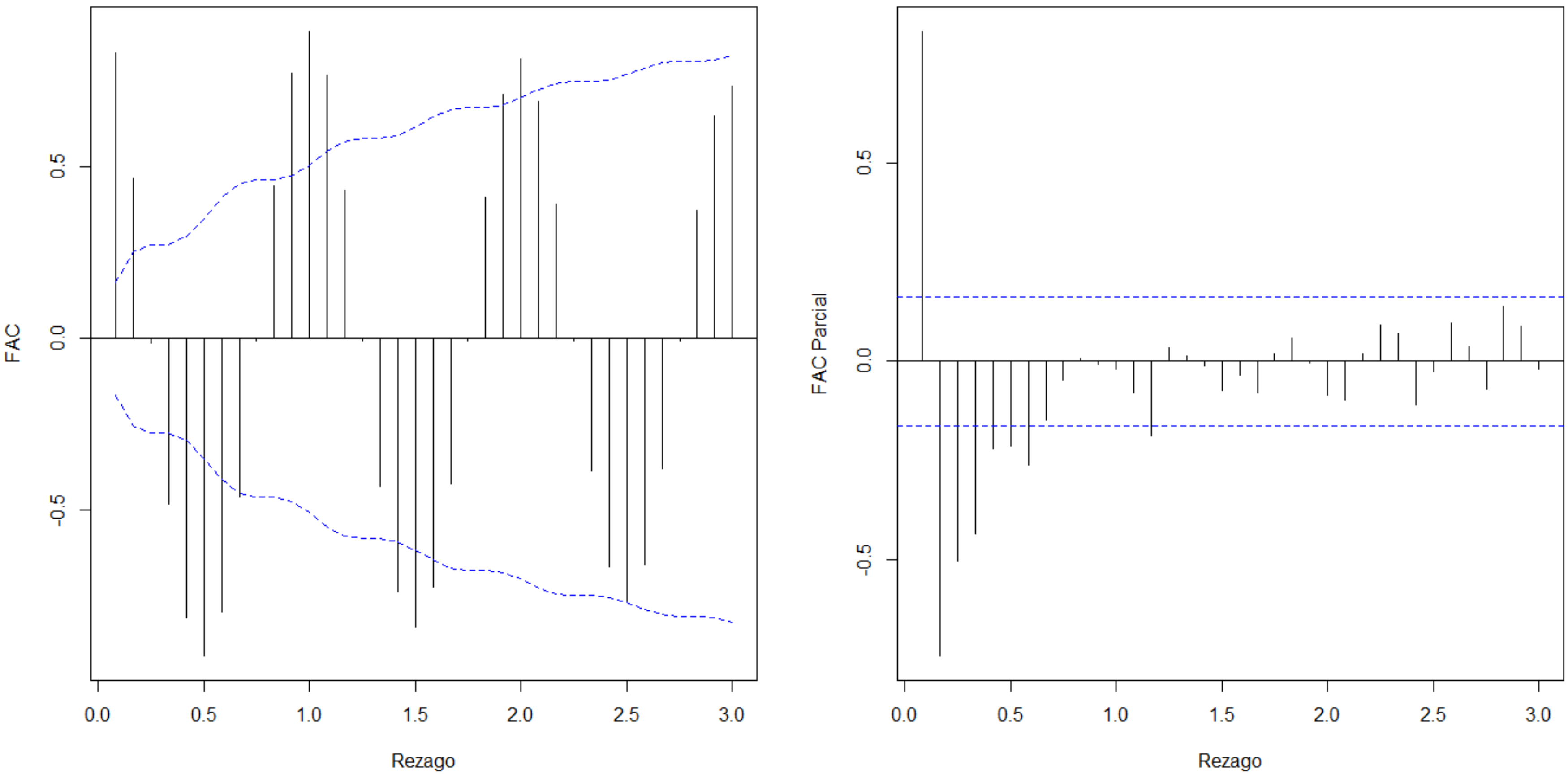
En muchas series las observaciones presentan dependencia con las observaciones previas, pero también con las que ocurrieron hace un día, una semana, un mes o un año. En estas situaciones la ACF debe reflejar dependencia temporal larga.

De esta forma, se puede especificar un modelo ARMA con información de la variable y las innovaciones en rezagos estacionales (*s*, *2s*, *3s*,...), de manera equivalente a los modelos ARMA regulares.

La estructura ARMA estacional, ARMA(P,Q)s , aparece en la ACF y la PACF con autocorrelaciones sobresalientes cada *s* periodos.

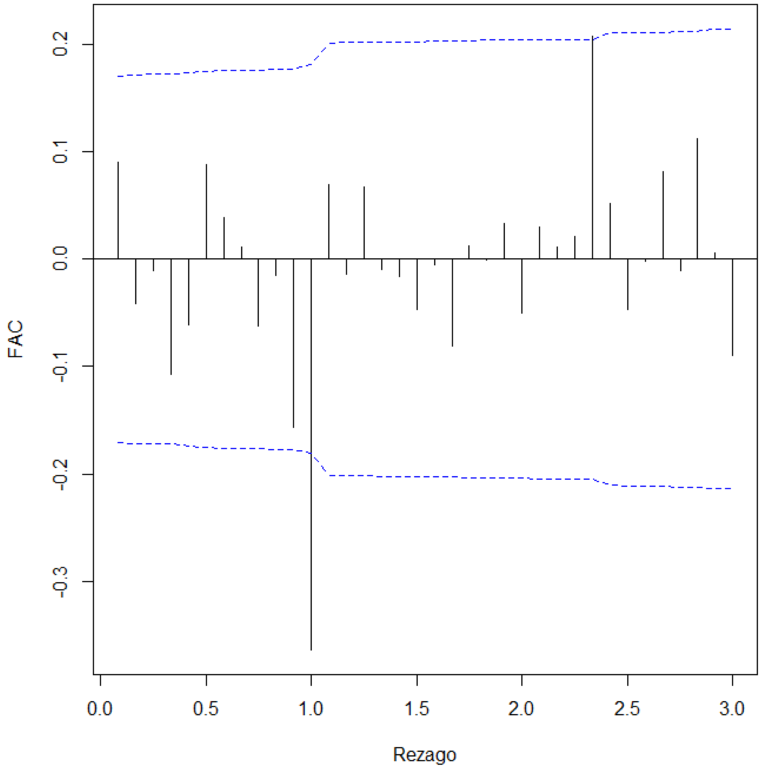
Identificación del Modelo

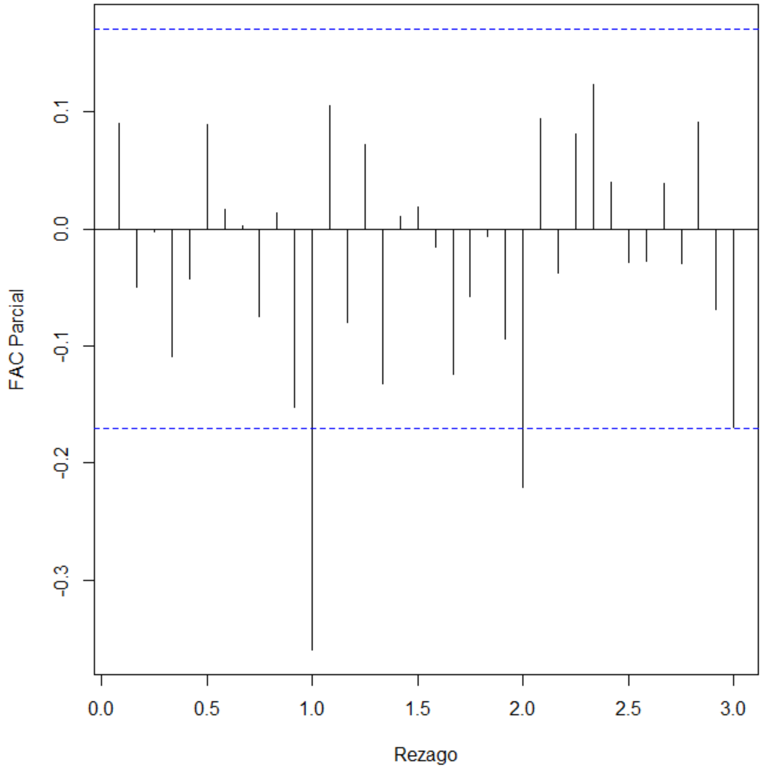
FAC y FACP de un Modelo estacional:



Identificación del Modelo

Regresando al ejemplo de la Temperatura, una vez que se ha convertido a estacionaria la serie realizando una diferencia estacional, las FAC y FACP son las siguientes:





Identificación del Modelo

La parte regular de las funciones no muestran no muestran valores significativos. Se evidencia una componente estacional en la FAC y dos en la FACP.

#Estimación del modelo y detección automática en R:

```
library(forecast)
mod1<-auto.arima(tempdub)
mod2<-auto.arima(tempdub, d=0, D=1)
mod3<-arima(tempdub, order=c(0,0,0), seasonal=list(order=c(0,1,1)))
```

Universidad del Valle

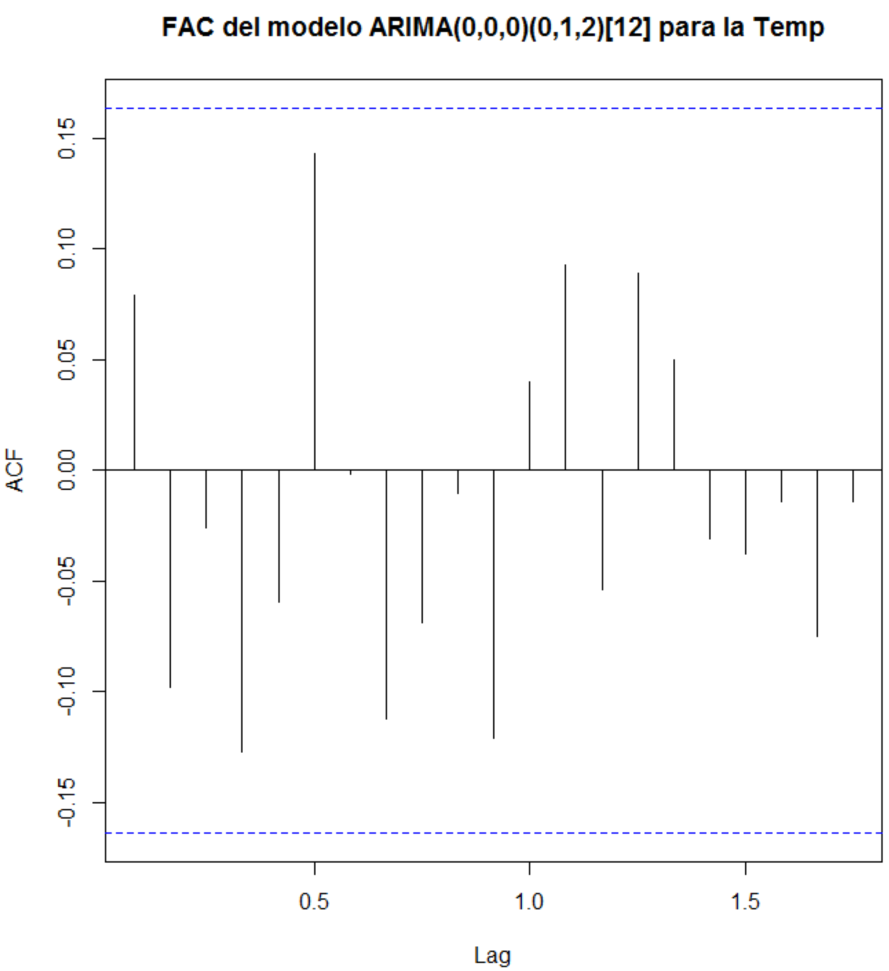
Validación del Modelo

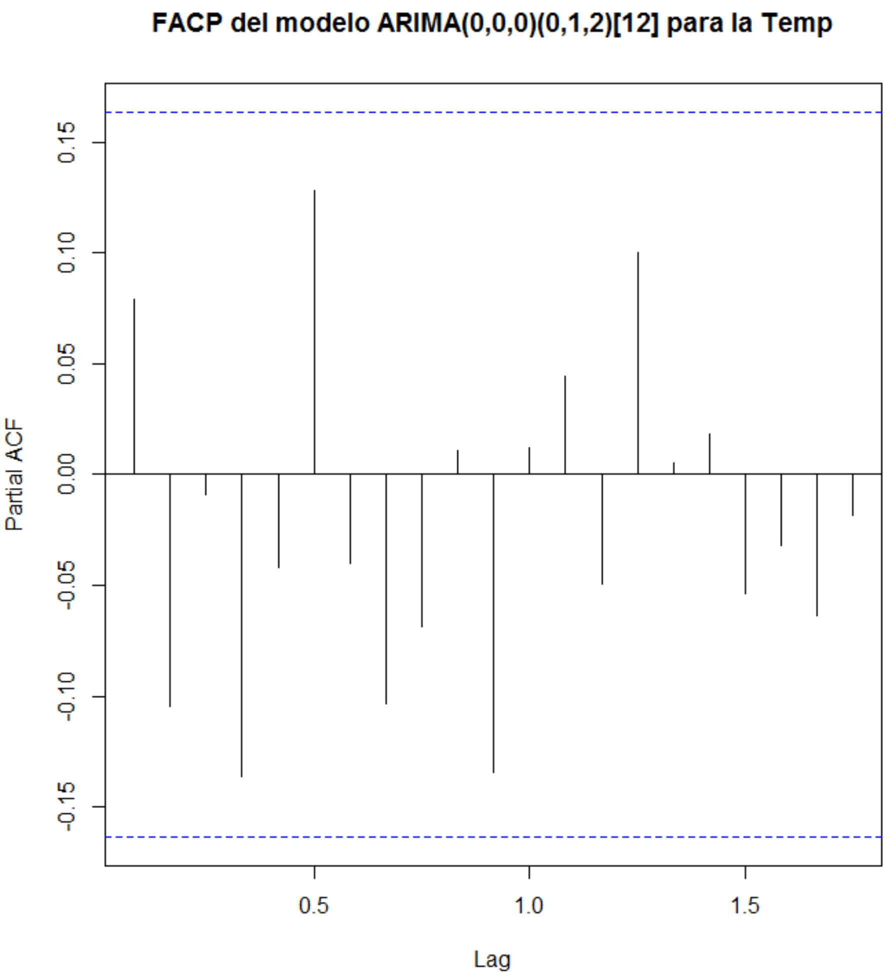
Si una serie esta bien identificada, cuando se ajusta el modelo los residuos no deben tener estructura, es decir, deben parecerse a un ruido blanco.

Un ruido blanco es una serie estacionaria en la que ninguna observación depende de las otras y, por tanto, todos los valores de la ACF y la PACF son nulos. La FAC y la FACP deben ser muy similares y los valores no son significativamente distintos de cero (valores dentro de sus bandas de confianza).

```
acf(rstandard(mod3))
pacf(rstandard(mod3))
```

Validación del Modelo





Validación del Modelo

El contraste de *Ljung-Box-Pierce* evalúa la hipótesis nula de que las primeras autocorrelaciones son cero. La hipótesis alternativa de esta prueba implica que alguna de las correlaciones es distinta de cero y, por tanto, no se puede asumir que los residuos sean ruido blanco.

```
library(stats)
Box.test(rstandard(mod3), lag=12, type="Ljung-Box")

library(TSA)
LB.test(mod3, lag=12)
```

Se espera que el valor-p de la prueba sea superior al nivel de significancia.

Validación del Modelo

La representación gráfica de la serie de residuos estandarizados. El gráfico de los residuos debe mostrar que los residuos varían en torno a cero, sin tendencias, la varianza es constante y no hay valores atípicos. Aproximadamente el 95% de los residuos estandarizados deben estar entre -2 y 2 desviaciones estándar.

```
plot(rstandard(mod3), xlab="", ylab="", main="Residuos estandarizados",
  type="o", ylim=c(-3,3))
abline(h=2, lty=3, col="red")
abline(h=-2, lty=3, col="red")

sum(rstandard(mod3)>2)
sum(rstandard(mod3)< -2)

8/length(rstandard(mod3))
```

Predicción

Para cada predicción tenemos el valor puntual y el intervalo de predicción.

La predicción esta sujeta al menos a dos fuentes de error:

- El carácter aleatorio de las innovaciones.
- Fallos en la especificación del modelo.

```
plot(tempdub, xlim=c(1964, 1977))
pred <- predict(mod3, 12)
plot(pred$pred, xlab="", ylab="", main="Predicciones para 1976", col="red")
lines(pred$pred+pred$se, lty=3, col="blue")
lines(pred$pred-pred$se, lty=3, col="blue")

#Funciona cuando se tiene un modelo AR o MA regular
plot.Arima(mod3, n.ahead=12, xlab="", ylab="", main="Predicciones para 1976", type="l")
```

Universidad del Valle