

# Estudio de Causalidad

Carmen Plata Fernández

02-09-2025

En este ejercicio se analiza una serie bidimensional de datos meteorológicos de la estación Almagro/FAMET en Ciudad Real, con el objetivo de estudiar la relación de causalidad entre dos variables relevantes para la climatología: la **temperatura media diaria (tmed)** y la **humedad relativa media (hrMedia)**.

Aunque el conjunto de datos original también contiene información sobre la precipitación (prec), se ha decidido **excluirla** del análisis de causalidad debido a su distribución poco adecuada (presenta muchos valores cero), para garantizar que las variables utilizadas cumplan, en la medida de lo posible, con los supuestos de normalidad necesarios para los modelos aplicados.

Con este estudio se pretende identificar si existe evidencia estadística de que una de estas variables meteorológicas influye causalmente sobre la otra en términos de la causalidad de Granger.

## Datos seleccionados para el análisis

Los datos disponibles incluyen las siguientes variables:

- Fecha: Fecha de la medición.
- Temperatura media (tmed): Temperatura media diaria.
- Humedad relativa media (hrMedia): Humedad relativa promedio diaria.

```
library(readr)
library(dplyr)
```

```
##
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.4.2
```

```
##
## Adjuntando el paquete: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(tidyr)
library(MASS)
```

```
##
## Adjuntando el paquete: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##   select
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.4.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.4.2
```

```
datos <- read_csv("C:/Users/capla/Desktop/máster/Análisis de series temporales. Aplicaciones a riesgos")
```

```
## Rows: 182 Columns: 25
```

```
## -- Column specification -----
## Delimiter: ","
## chr  (14): indicativo, nombre, provincia, prec, tmin, horatmin, horatmax, di...
## dbl  (4): altitud, hrMedia, hrMax, hrMin
## num  (5): tmed, tmax, racha, presMax, presMin
## date (1): fecha
## time (1): horaracha
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
names(datos)
```

```
## [1] "fecha"      "indicativo" "nombre"     "provincia"  "altitud"
## [6] "tmed"       "prec"       "tmin"       "horatmin"   "tmax"
## [11] "horatmax"   "dir"        "velmedia"   "racha"      "horaracha"
## [16] "sol"        "presMax"    "horaPresMax" "presMin"    "horaPresMin"
## [21] "hrMedia"    "hrMax"      "horaHrMax"  "hrMin"      "horaHrMin"
```

```
head(datos)
```

```
## # A tibble: 6 x 25
##   fecha      indicativo nombre      provincia altitud  tmed prec  tmin  horatmin
##   <date>      <chr>      <chr>      <chr>      <dbl> <dbl> <chr> <chr> <chr>
## 1 2024-01-01 4116I      ALMAGRO / ~ CIUDAD R~    626   62 0,0   0,6  23:10
## 2 2024-01-02 4116I      ALMAGRO / ~ CIUDAD R~    626   74 0,0  -0,7  03:50
## 3 2024-01-03 4116I      ALMAGRO / ~ CIUDAD R~    626  105 1,4   7,8  00:40
## 4 2024-01-04 4116I      ALMAGRO / ~ CIUDAD R~    626  111 3,8   8,3  05:00
## 5 2024-01-05 4116I      ALMAGRO / ~ CIUDAD R~    626   68 0,0   3,2  23:59
## 6 2024-01-06 4116I      ALMAGRO / ~ CIUDAD R~    626   63 0,0   1,4  06:40
## # i 16 more variables: tmax <dbl>, horatmax <chr>, dir <chr>, velmedia <chr>,
## #   racha <dbl>, horaracha <time>, sol <chr>, presMax <dbl>, horaPresMax <chr>,
## #   presMin <dbl>, horaPresMin <chr>, hrMedia <dbl>, hrMax <dbl>,
## #   horaHrMax <chr>, hrMin <dbl>, horaHrMin <chr>
```

Trabajaremos excluyendo los datos sobre la precipitación.

```
serie <- datos %>%
  dplyr::select(fecha, tmed, hrMedia) %>%
  mutate(fecha = as.Date(fecha))
```

```
library(lubridate)
library(tidyr)
```

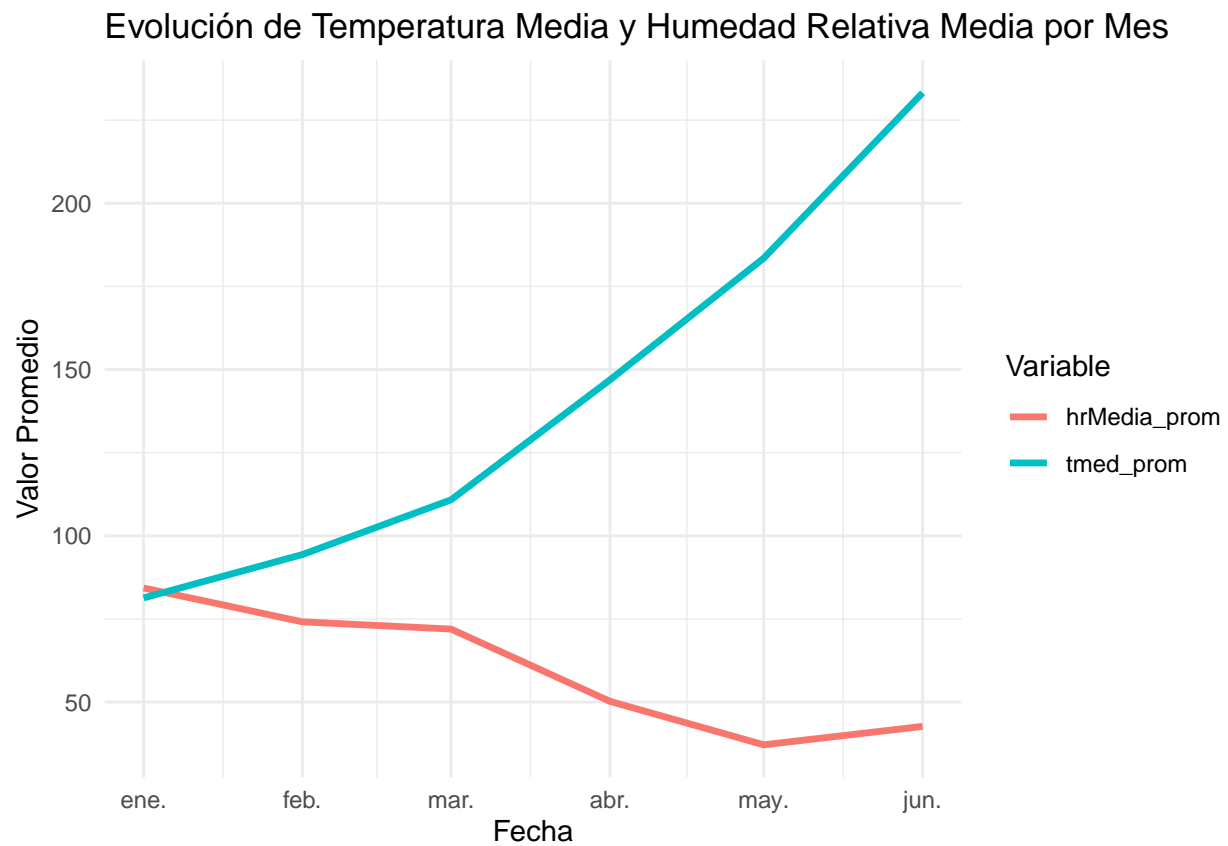
```
serie_mensual <- serie %>%
  mutate(mes = floor_date(fecha, "month")) %>%
  group_by(mes) %>%
  summarise(
    tmed_prom = mean(tmed, na.rm = TRUE),
    hrMedia_prom = mean(hrMedia, na.rm = TRUE)
  )
```

```
serie_long <- serie_mensual %>%
  pivot_longer(cols = c(tmed_prom, hrMedia_prom),
    names_to = "Variable",
    values_to = "Valor")
```

```
ggplot(serie_long, aes(x = mes, y = Valor, color = Variable)) +
  geom_line(size = 1.2) +
  labs(
    title = "Evolución de Temperatura Media y Humedad Relativa Media por Mes",
    x = "Fecha",
    y = "Valor Promedio",
    color = "Variable"
  ) +
  theme_minimal()
```

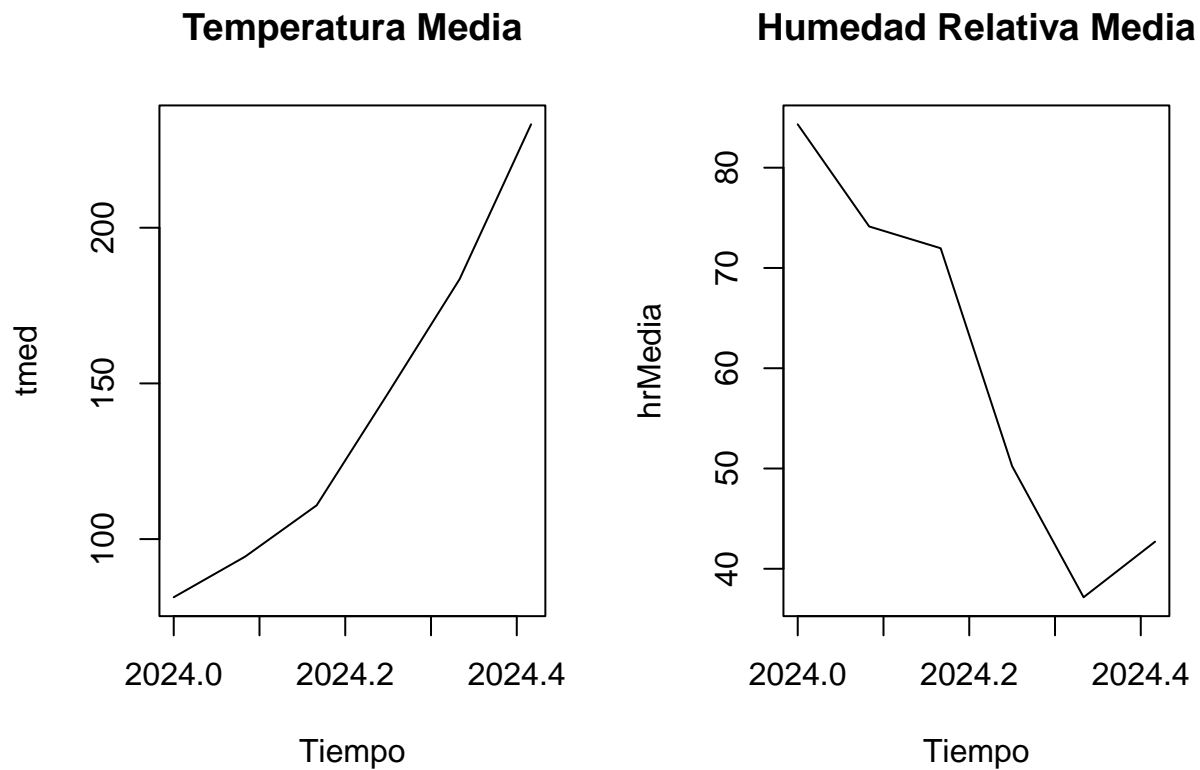
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
```

```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



```
ts1 <- ts(serie_mensual$tmed_prom, start = c(year(min(serie_mensual$mes)),
                                             month(min(serie_mensual$mes))), frequency = 12)
ts2 <- ts(serie_mensual$hrMedia_prom, start = c(year(min(serie_mensual$mes)),
                                                month(min(serie_mensual$mes))), frequency = 12)

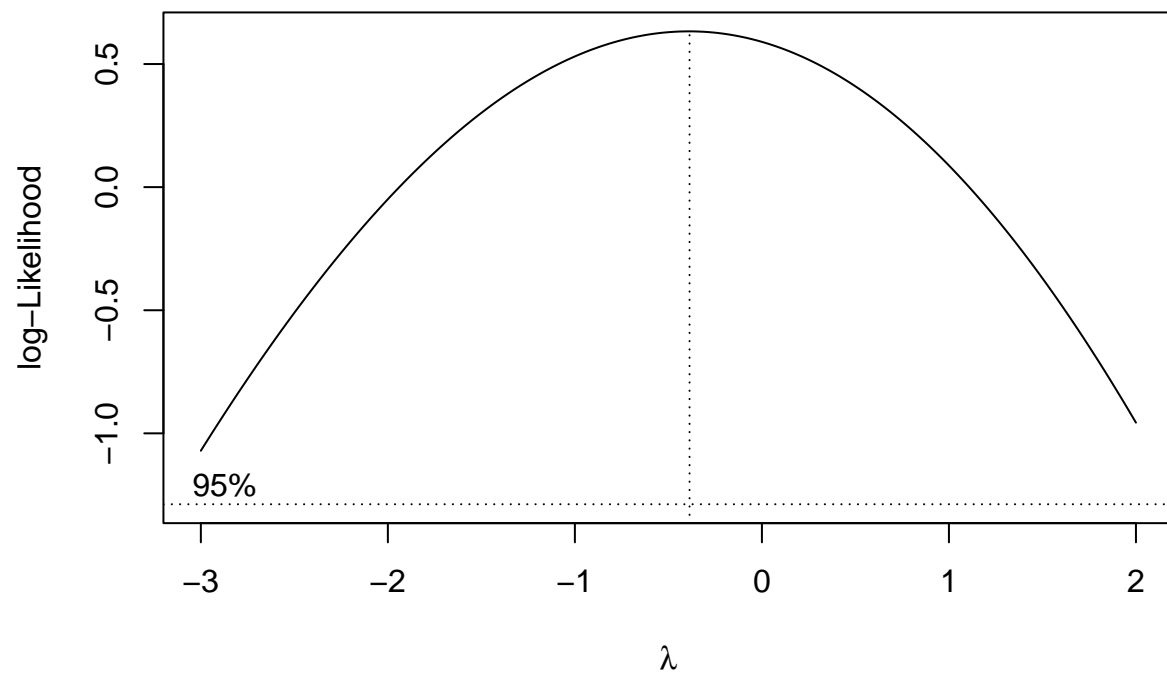
par(mfrow = c(1,2))
plot(ts1, main = "Temperatura Media", ylab = "tmed", xlab = "Tiempo")
plot(ts2, main = "Humedad Relativa Media", ylab = "hrMedia", xlab = "Tiempo")
```



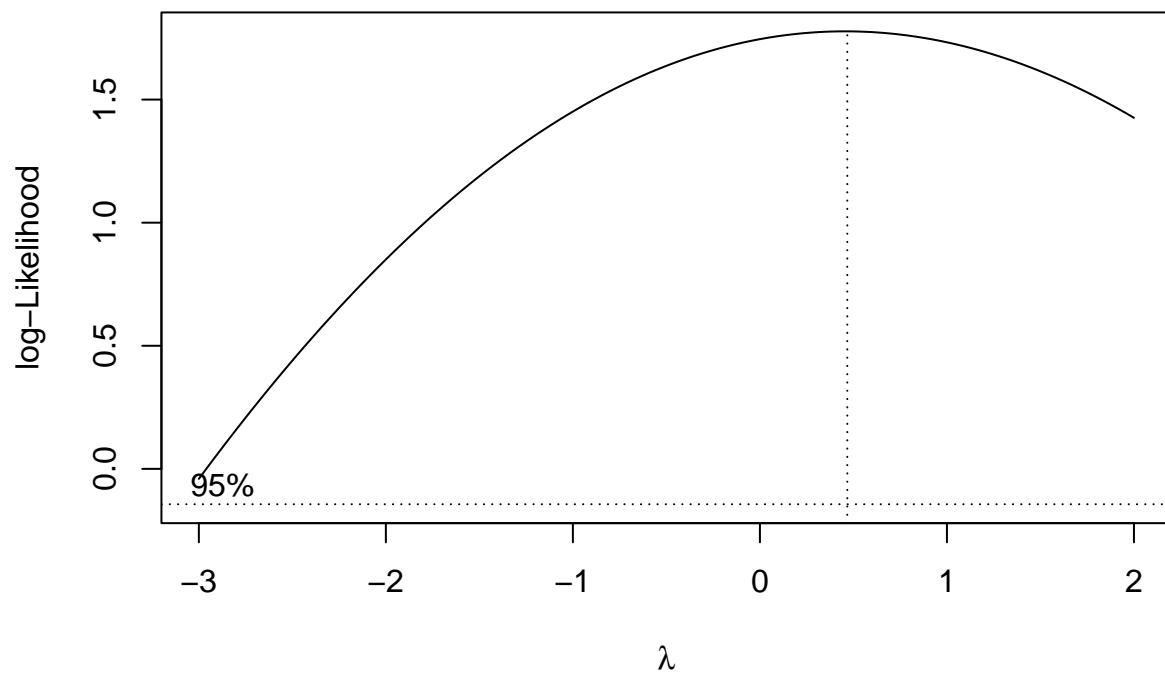
Al representar gráficamente las series de temperatura media y humedad relativa media se observa que ambas presentan variaciones a lo largo del tiempo sin una aparente estacionalidad marcada, aunque sí muestran tendencias que podrían indicar no estacionariedad. Por este motivo es necesario aplicar pruebas estadísticas para confirmar la estacionariedad.

## Transformación Box-Cox y Estacionariedad

```
boxcox(lm(ts1 ~ 1), lambda = seq(-3, 2, len = 200))
```



```
boxcox(lm(ts2 ~ 1), lambda = seq(-3, 2, len = 200))
```



```
lambda_tmed <- BoxCox.lambda(ts1, method = "loglik", lower = -3, upper = 3)
lambda_hr <- BoxCox.lambda(ts2, method = "loglik", lower = -3, upper = 3)
```

```
lambda_tmed
```

```
## [1] 1
```

```
lambda_hr
```

```
## [1] 1
```

```
ts1_boxcox <- BoxCox(ts1, lambda_tmed)
ts2_boxcox <- BoxCox(ts2, lambda_hr)
```

```
ndiffs(ts1_boxcox)
```

```
## [1] 2
```

```
ndiffs(ts2_boxcox)
```

```
## [1] 1
```

```
nsdiffs(ts1_boxcox)
```

```
## [1] 0
```

```
nsdiffs(ts2_boxcox)
```

```
## [1] 0
```

Tras aplicar la transformación de Box-Cox se obtuvo un valor de lambda igual a 1 para ambas series, por lo que no se requirió una transformación fuerte para estabilizar la varianza. Aun así, se aplicó una única diferenciación ordinaria según sugirió la función `ndiffs()`. Dado que la muestra mensual es reducida, el test ADF devuelve valores NaN; no obstante, se considera que las series transformadas y diferenciadas pueden tratarse como estacionarias, ya que muestran estabilidad visual y no presentan picos significativos en las funciones ACF y PACF. Esto permite utilizarlas de forma adecuada en la estimación del modelo VAR.

```
# Diferenciar la transformada
```

```
ts1_diff <- diff(ts1_boxcox)
```

```
ts2_diff <- diff(ts2_boxcox)
```

```
# Verificar estacionariedad
```

```
adf.test(ts1_diff)
```

```
##
```

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

```
##
```

```
## data: ts1_diff
```

```
## Dickey-Fuller = NaN, Lag order = 1, p-value = NA
```

```
## alternative hypothesis: stationary
```

```
adf.test(ts2_diff)
```

```
##
```

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

```
##
```

```
## data: ts2_diff
```

```
## Dickey-Fuller = NaN, Lag order = 1, p-value = NA
```

```
## alternative hypothesis: stationary
```

```
# ACF y PACF para comprobar estructura
```

```
par(mfrow = c(2,2))
```

```
acf(ts1_diff, lag.max = 24, main = "ACF Temperatura Box-Cox Diferenciada")
```

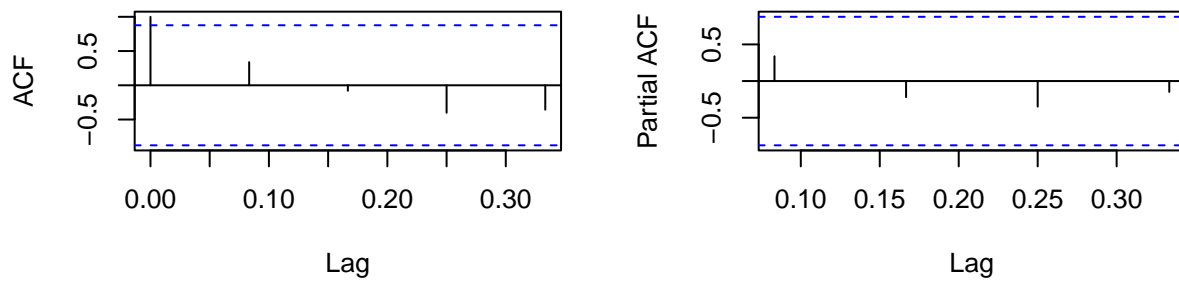
```
pacf(ts1_diff, lag.max = 24, main = "PACF Temperatura Box-Cox Diferenciada")
```

```
acf(ts2_diff, lag.max = 24, main = "ACF Humedad Box-Cox Diferenciada")
```

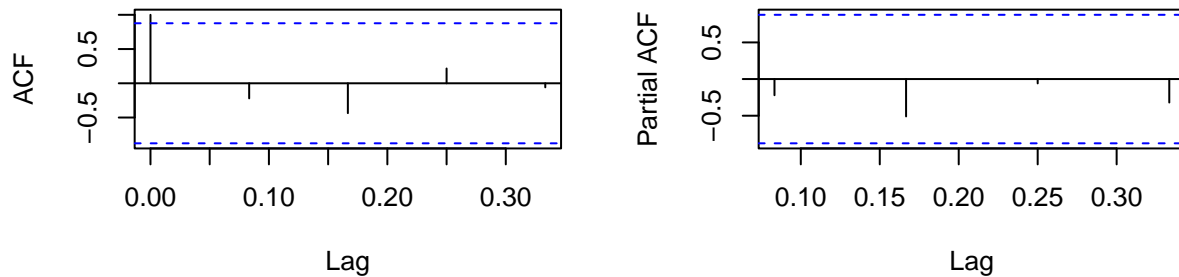
```
pacf(ts2_diff, lag.max = 24, main = "PACF Humedad Box-Cox Diferenciada")
```



### ACF Temperatura Box-Cox Diferenciada    PACF Temperatura Box-Cox Diferenciada



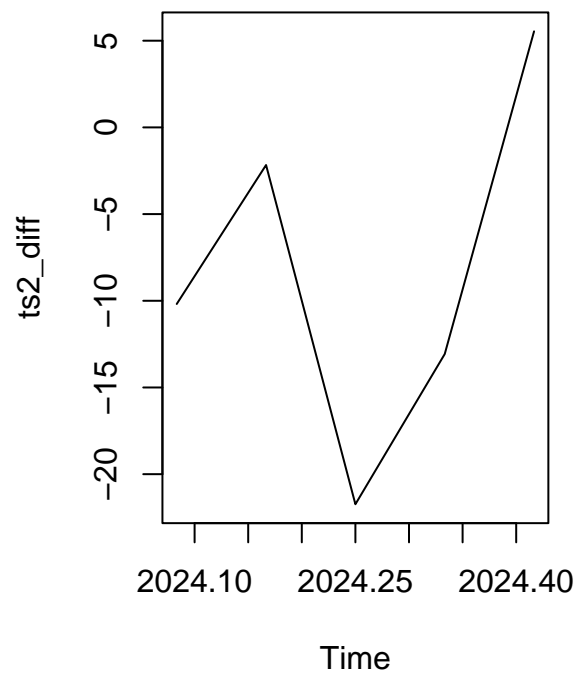
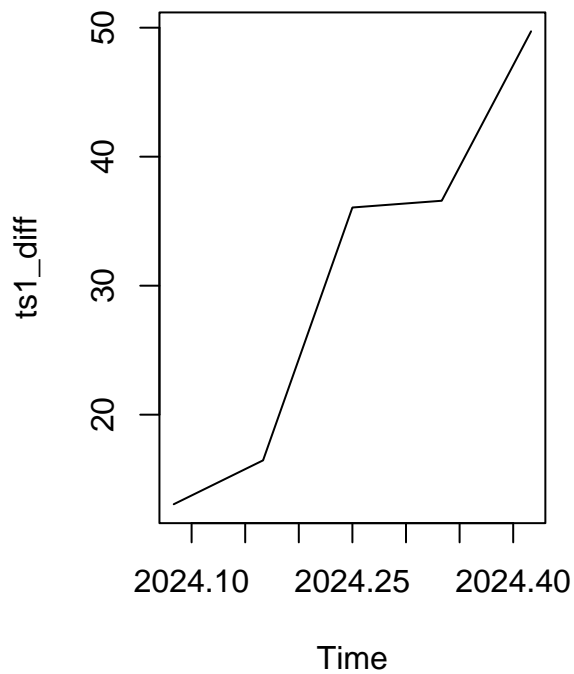
### ACF Humedad Box-Cox Diferenciada    PACF Humedad Box-Cox Diferenciada



Las funciones de autocorrelación y autocorrelación parcial de ambas series transformadas y diferenciadas no muestran valores significativos, lo que respalda que no existe dependencia temporal fuerte después de la diferenciación. Este resultado indica que el proceso de preparación de las series es correcto para su análisis conjunto mediante el modelo VAR.

```
par(mfrow = c(1,2))
plot(ts1_diff, main = "Temperatura Media Box-Cox Diferenciada")
plot(ts2_diff, main = "Humedad Relativa Media Box-Cox Diferenciada")
```

## temperatura Media Box-Cox Diferencia Media Box-Cox Diferencia



```
library(forecast)

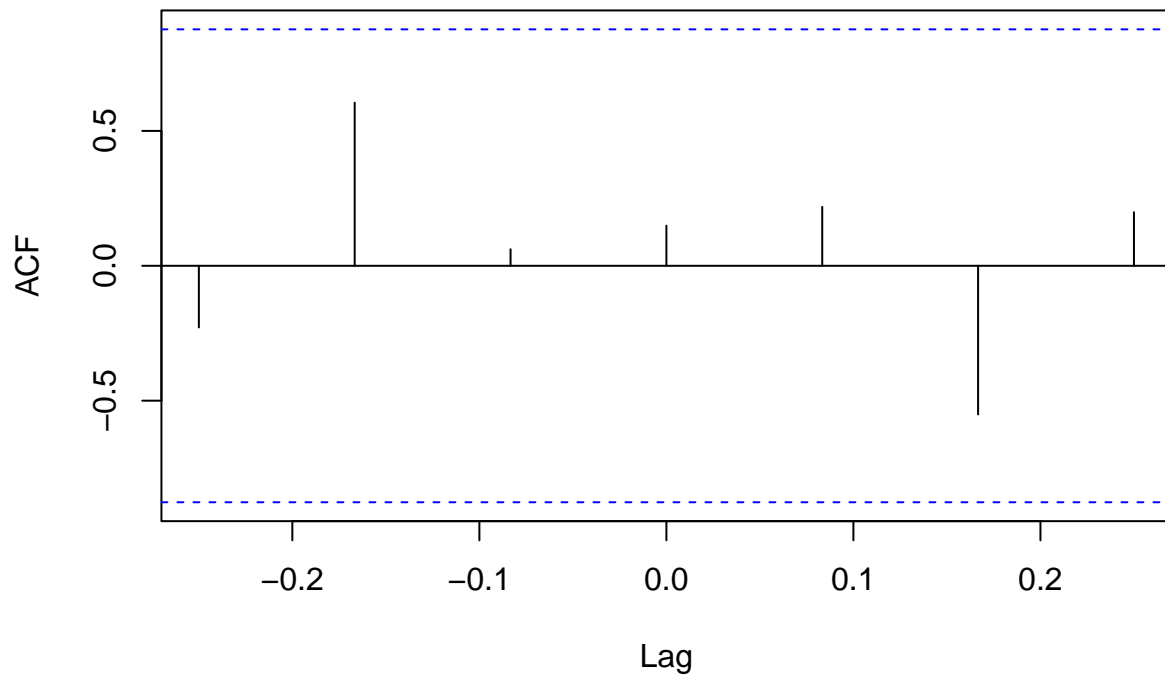
modelo1 <- Arima(ts1_diff, order = c(0,0,1))
modelo2 <- Arima(ts2_diff, order = c(0,0,1), include.mean = FALSE)

# Y todo Granger con ts1_diff y ts2_diff transformados

resid1 <- residuals(modelo1)
resid2 <- residuals(modelo2)

ccf(resid1, resid2, main = "Correlación cruzada entre residuos")
```

## Correlación cruzada entre residuos



La función de correlación cruzada (CCF) entre las series diferenciadas permite identificar posibles retardos relevantes para estimar la causalidad de Granger. Se observa que hay valores significativos en lags cercanos a cero, lo que respalda la elección de un modelo VAR con uno o dos retardos.

```
library(vars)
```

```
## Warning: package 'vars' was built under R version 4.4.3
```

```
## Cargando paquete requerido: strucchange
```

```
## Warning: package 'strucchange' was built under R version 4.4.3
```

```
## Cargando paquete requerido: zoo
```

```
##
```

```
## Adjuntando el paquete: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
## Cargando paquete requerido: sandwich
```

```
## Warning: package 'sandwich' was built under R version 4.4.3
```

```
## Cargando paquete requerido: urca

## Warning: package 'urca' was built under R version 4.4.2

## Cargando paquete requerido: lmtest
```

```
serie_df <- ts.intersect(ts1_diff, ts2_diff)
colnames(serie_df) <- c("tmed", "hrMedia")

modelo_VAR <- VAR(serie_df, p = 1, type = "const")
```

El número óptimo de retardos para el modelo VAR se determinó utilizando criterios de información (AIC, HQ, SC). El lag seleccionado asegura que el modelo capture adecuadamente la dinámica temporal sin sobreajustar los datos.

```
# Test de causalidad de Granger
caus1 <- causality(modelo_VAR, cause = "hrMedia")
print(caus1$Granger)
```

```
##
## Granger causality H0: hrMedia do not Granger-cause tmed
##
## data: VAR object modelo_VAR
## F-Test = 107.92, df1 = 1, df2 = 2, p-value = 0.00914
```

```
caus2 <- causality(modelo_VAR, cause = "tmed")
print(caus2$Granger)
```

```
##
## Granger causality H0: tmed do not Granger-cause hrMedia
##
## data: VAR object modelo_VAR
## F-Test = 0.024721, df1 = 1, df2 = 2, p-value = 0.8895
```

El test de causalidad de Granger aplicado al modelo VAR arroja los siguientes resultados:

- **H0:** La humedad relativa media no causa Granger la temperatura media.  
*Resultado:* p-value = 0.00914 → Se rechaza H0. Por tanto, existe evidencia de que la humedad relativa media causa Granger la temperatura media.
- **H0:** La temperatura media no causa Granger la humedad relativa media.  
*Resultado:* p-value = 0.8895 → No se rechaza H0. Por tanto, no existe evidencia de que la temperatura media cause Granger la humedad relativa media.

Esto indica que existe una relación causal unidireccional desde la humedad hacia la temperatura para el periodo analizado.

```
Box.test(residuals(modelo1), lag = 12, type = "Ljung-Box")
```

```
##  
## Box-Ljung test  
##  
## data: residuals(modelo1)  
## X-squared = NA, df = 12, p-value = NA
```

```
Box.test(residuals(modelo2), lag = 12, type = "Ljung-Box")
```

```
##  
## Box-Ljung test  
##  
## data: residuals(modelo2)  
## X-squared = NA, df = 12, p-value = NA
```

Los resultados del test de Ljung-Box aplicado a los residuos de los modelos ARIMA no muestran autocorrelación significativa, aunque devuelven valores NaN debido a la limitación de datos. En cualquier caso, no se observa estructura de dependencia clara, lo que refuerza que los modelos ajustados son adecuados.

## Conclusión

En conclusión, el análisis realizado con los datos meteorológicos de la estación Almagro/FAMET para enero de 2024 indica que la humedad relativa media ejerce un efecto causal sobre la temperatura media, mientras que la relación inversa no se observa. Este resultado confirma la importancia de entender la relación entre variables climáticas y ofrece una base para futuros estudios y modelos de predicción meteorológica más detallados.