

Análisis de Series Temporales: Tarea 3

Sergio García Prado

`sergio.garcia.prado@alumnos.uva.es`

31 de diciembre de 2018

- **Archivo:** `weight-loss.csv`
- **Serie:** Frecuencia de búsquedas para la palabra clave “Weight loss” a través del buscador *Google* por meses, desde *Enero de 2004* hasta *Diciembre de 2018*. Los valores han sido estandarizados en el rango $[0, 100]$.

1. Etapa de identificación

Tal y como se indica al comienzo de este documento, en este trabajo se va a trabajar con la serie temporal referida a la frecuencia de búsqueda de la palabra clave “Weight loss” (a nivel mundial) a través del buscador *Google*. Se ha escogido esta serie para el trabajo por su estructura estacional claramente marcada. Se cree que dicha estructura tiene un alto grado de relación con un índice sobre la preocupación de la población por su peso a lo largo del tiempo.

Para evitar problemas de privacidad, los datos se proporcionan estandarizados en el rango $[0, 100]$, lo cual elimina la escala de los mismos y únicamente permite estudiar la estructura estocástica de la serie. Esto no es un problema para el análisis que se realizará en este trabajo, dado que precisamente el objetivo del mismo es el de analizar la estructura de una serie temporal, siguiendo la metodología de *Box-Jenkins*.

En cuanto al particionamiento de los datos, estos se proporcionan en agrupaciones mensuales. Dado que se tiene información desde *Enero de 2004* hasta *Diciembre de 2018*, es decir, un total de *15 años*, lo cual suma $15 * 12 = 180$ observaciones en total. Con esta cantidad de observaciones, se cree que se podrá construir un modelo *SARIMA* (*ARIMA* con estacionalidad) de manera adecuada.

[TODO]

[TODO]

[TODO]

[TODO]

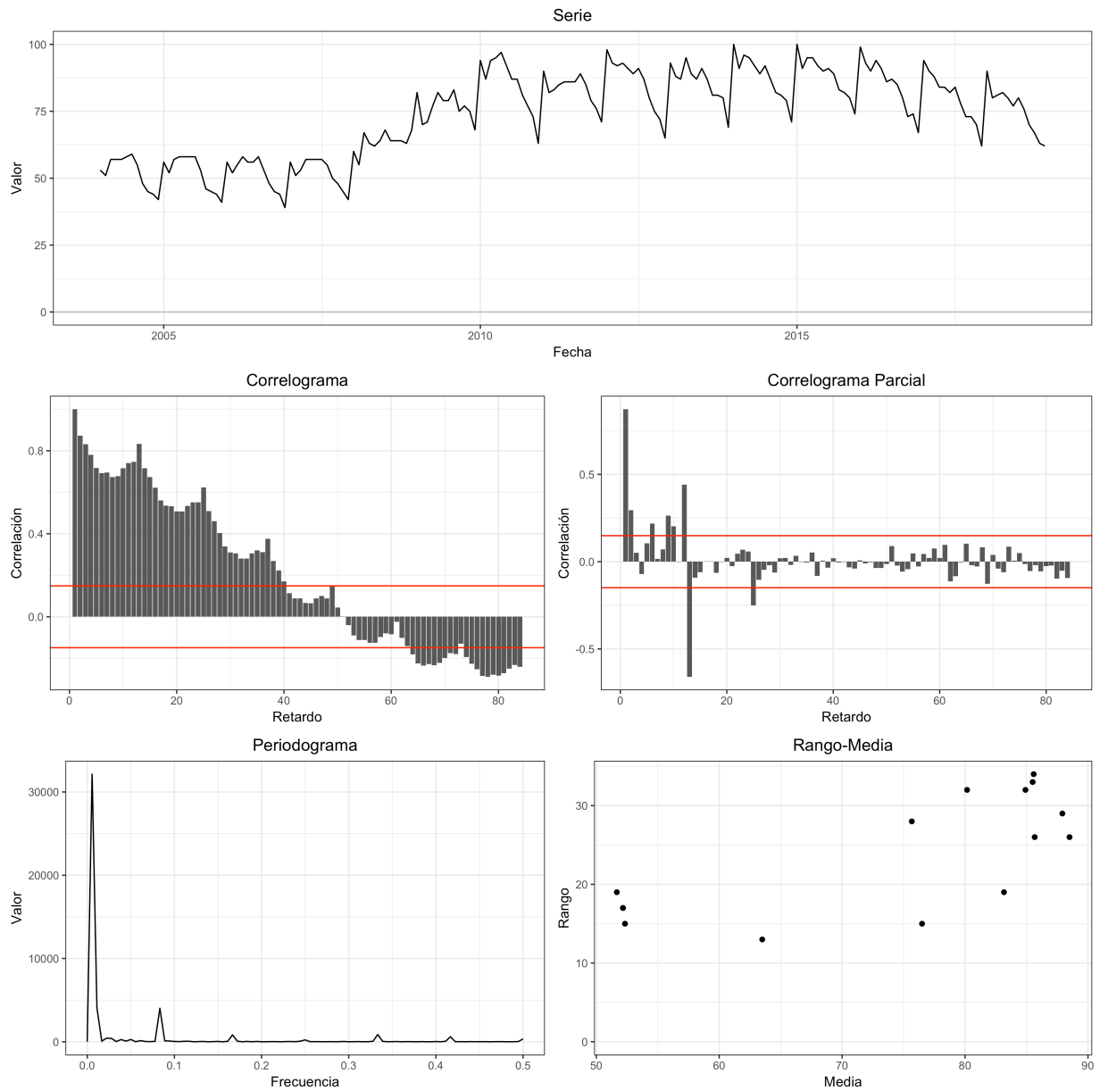


Figura 1: [TODO]

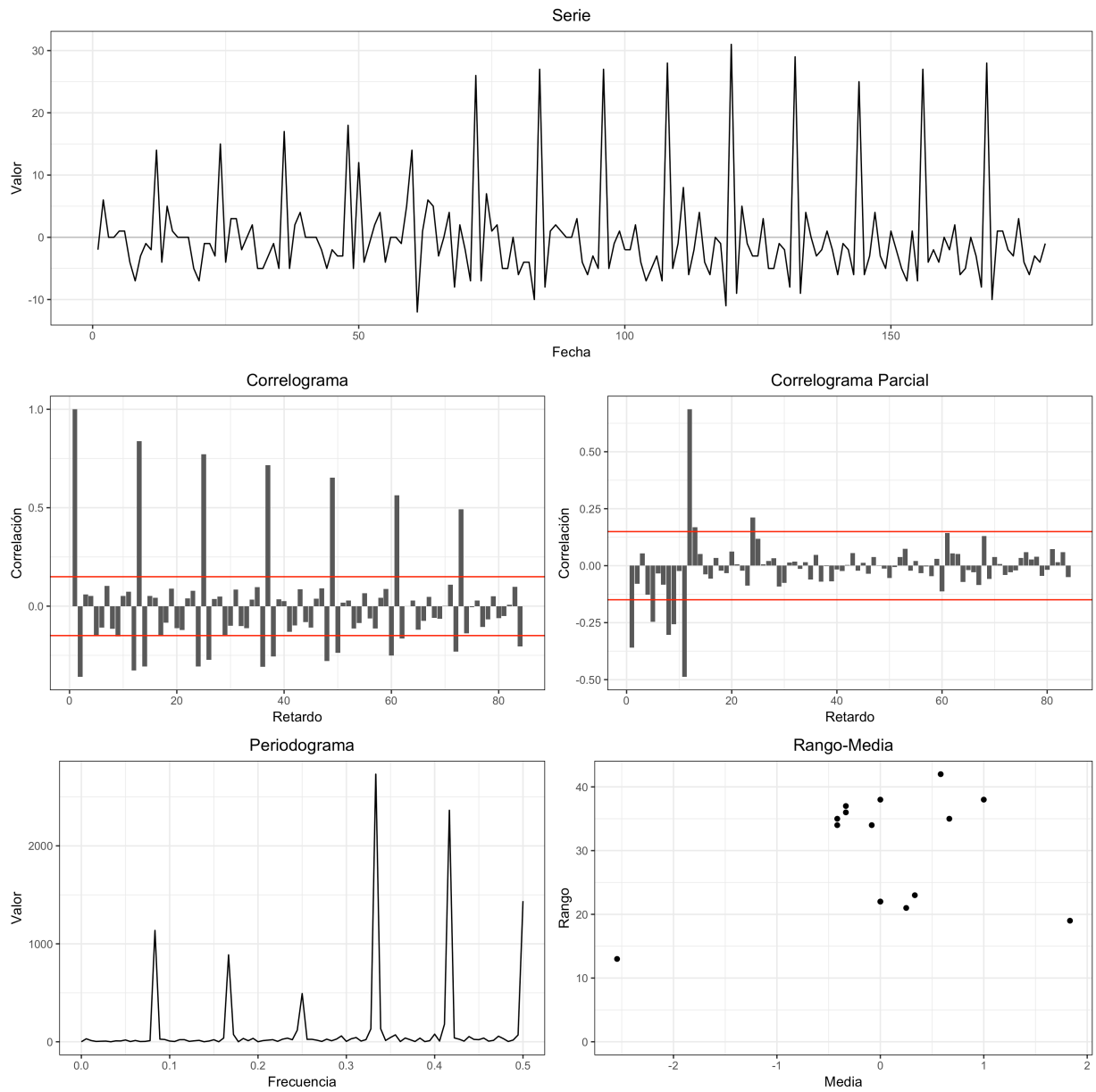


Figura 2: [TODO]

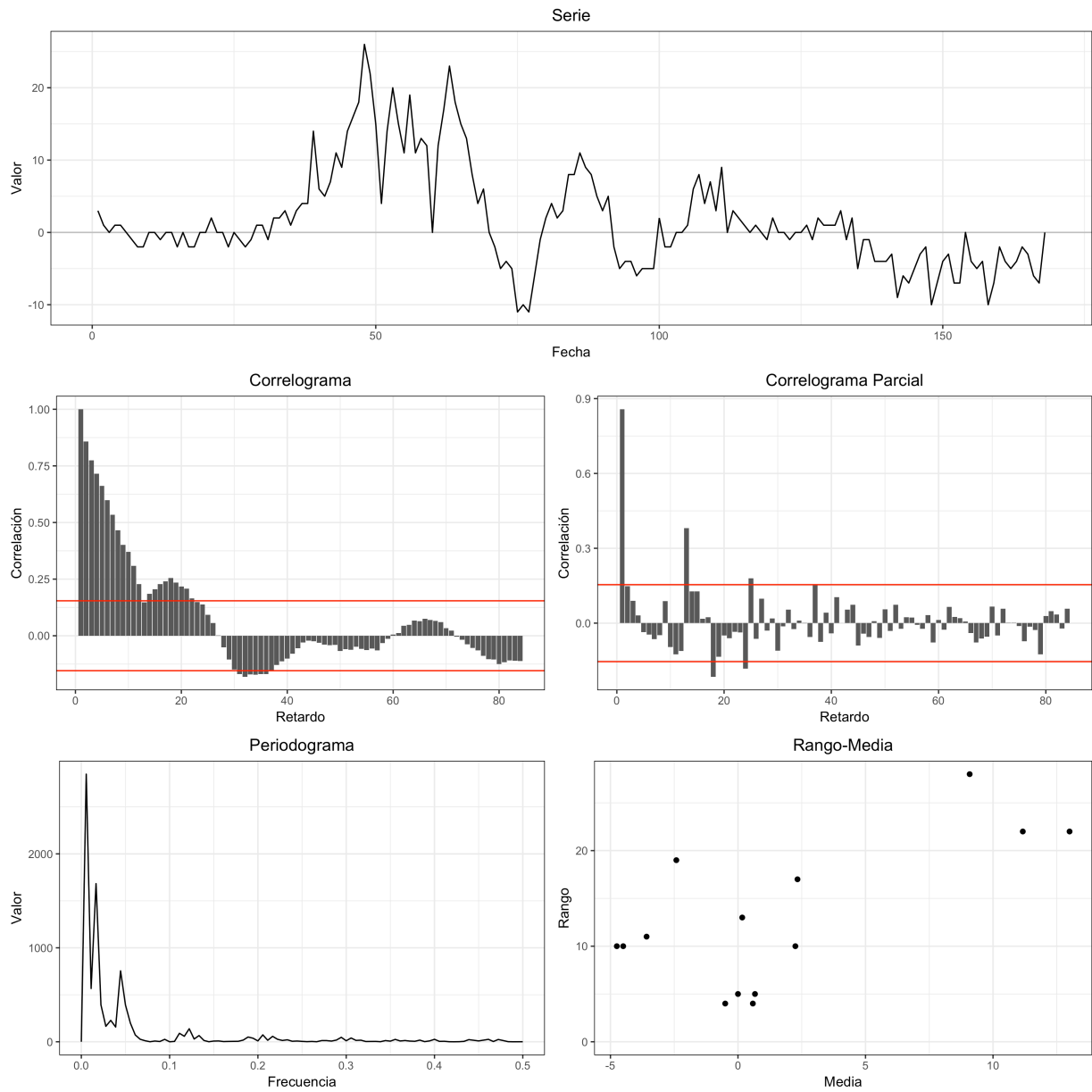


Figura 3: [TODO]

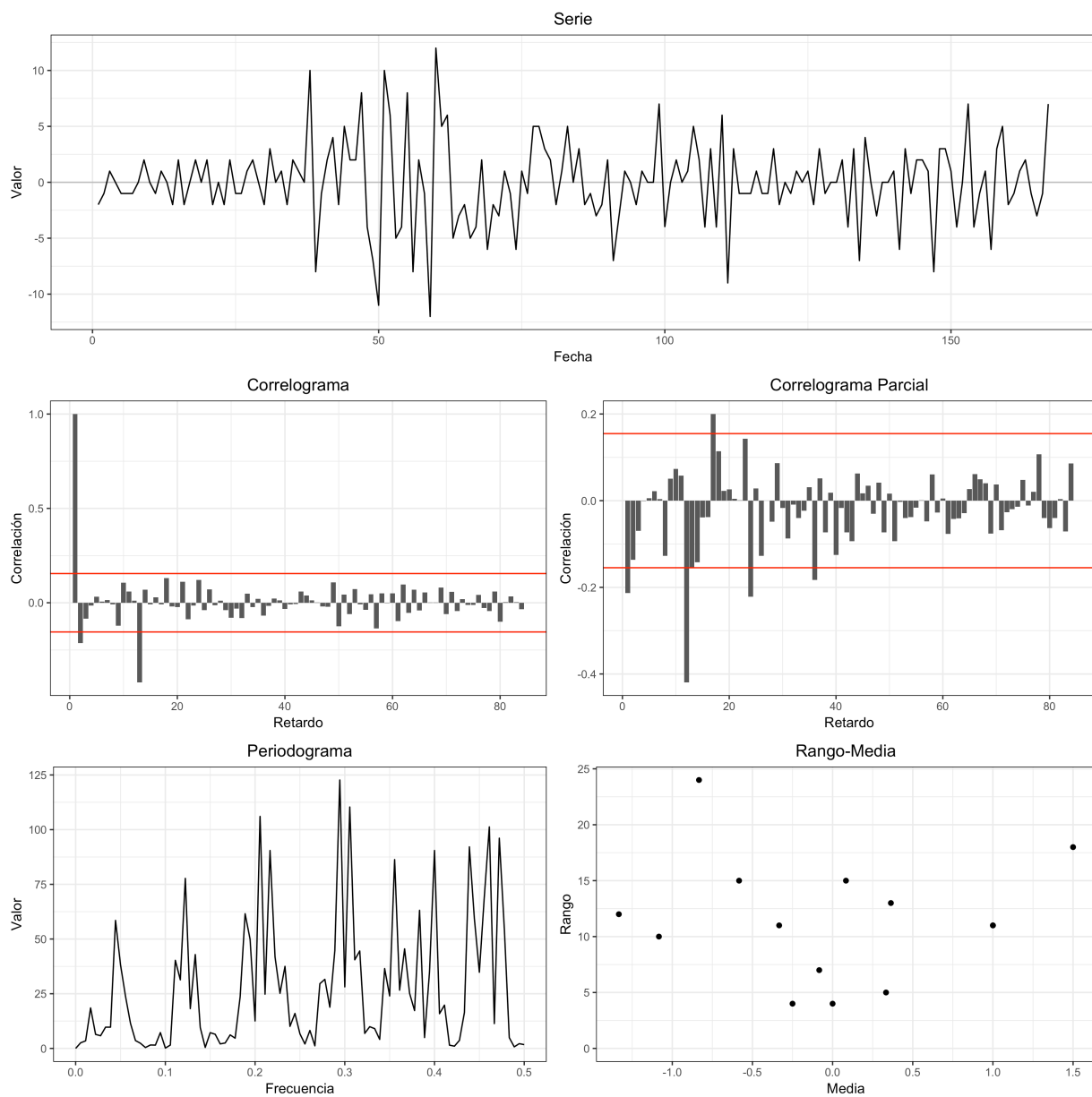


Figura 4: [TODO]

2. Etapa de estimación y validación

[TODO]

3. Comparación de modelos

[TODO]

4. Predicción

[TODO]

A. Código Fuente

[TODO]

```

## Author: Sergio García Prado
## Title: Time Series - Weight Loss - EDA

rm(list = ls())

library(magrittr)
library(dplyr)
library(ggplot2)
library(latex2exp)
require(reshape2)
library(forecast)
library(cowplot)
library(lubridate)

RangeMean <- function(x, seasonality) {
  n <- length(x)
  seq(1, n, by=seasonality) %>%
  sapply(function(i){
    a <- x[i:(i + seasonality - 1)]
    c(mean=mean(a, na.rm=TRUE), range=diff(range(a, na.rm = TRUE)))
  }) %>%
  t() %>%
  as.data.frame()
}

Correlogram <- function(x, n = length(x) - 1) {
  result <- acf(x, lag.max=n, plot=FALSE)$acf[0:n]
  data.frame(lag = 1:length(result), values = result)
}

PartialCorrelogram <- function(x, n = length(x) - 1) {
  result <- pacf(x, lag.max=n, plot=FALSE)$acf
  data.frame(lag = 1:length(result), values = result)
}

Periodogram <- function(x) {
  result <- TSA::periodogram(x, plot=FALSE)
  data.frame(freq = c(0, result$freq), spec = c(0, result$spec))
}

PlotTimeSeries <- function(df, seasonality, armonics = c(), lags = MAX_LAG){
  p.a <- ggplot(df) +
    aes(x = index, y = values) +
    xlab("Fecha") +
    ylab("Valor") +
    geom_hline(yintercept = 0, color = "gray") +
    geom_line() +
    theme_bw() +
    # scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggtitle('Serie')

  p.b <- ggplot(RangeMean(df$values, seasonality)) +
    aes(x = mean, y = range) +
    geom_point() +
    xlab("Media") +
    ylab("Rango") +
    expand_limits(y=0) +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggtitle('Rango-Media')

  p.c <- ggplot(Correlogram(df$values, lags)) +
    aes(x = lag, y = values) +
    xlab("Retardo") +
    ylab("Correlación") +
    geom_bar(stat="identity") +
    geom_hline(yintercept = 2/sqrt(nrow(df)), color = "red") +
    geom_hline(yintercept = -2/sqrt(nrow(df)), color = "red") +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggtitle('Correlograma')

  p.partial.correlogram <- ggplot(PartialCorrelogram(df$values, lags)) +
    aes(x = lag, y = values) +
    xlab("Retardo") +
    ylab("Correlación") +
    geom_bar(stat="identity") +
    geom_hline(yintercept = 2/sqrt(nrow(df)), color = "red") +
    geom_hline(yintercept = -2/sqrt(nrow(df)), color = "red") +

```