



Universidad
Internacional
Menéndez Pelayo

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN
INTELIGENCIA ARTIFICIAL

TRABAJO SERIES TEMPORALES

*Datos temporales
y complejos*

Laura Rodríguez Navas

Diciembre de 2020

rodrigueznabas@posgrado.uimp.es

Índice general

1. Análisis de la serie temporal	1
1.1. Introducción	1
1.2. Datos de la serie temporal	2
1.3. Características de la serie temporal	3
1.4. Componentes de la serie temporal	5
1.4.1. Análisis de autocorrelación y autocorrelación parcial	5
1.4.2. Descomposición de la serie	7
2. Búsqueda de outliers	11
3. Método de predicción	14
4. Aplicación del método de predicción	15
A. Apéndice: Título del Apéndice	16
A.1. Primera sección	16
B. Apéndice: Título del Apéndice	17
B.1. Primera sección	17
Bibliografía	18

Capítulo 1

Análisis de la serie temporal

1.1. Introducción

Una empresa tecnológica cuya área de negocio es la inteligencia artificial es contratada por una empresa eléctrica para que diseñe un sistema de recomendación que haga ofertas personalizadas a sus clientes sobre los paquetes energéticos más adecuados a sus consumos. Para ello, la empresa primero debe llevar a cabo un análisis exhaustivo de los consumos energéticos y diseñar una técnica de predicción que sea capaz de predecir dichos consumos con un horizonte temporal dado. La empresa eléctrica suministra a la empresa tecnológica para dicho cometido los consumos eléctricos desde el 01 de enero de 2015 hasta el 31 de diciembre de 2015 medidos con una frecuencia temporal de 10 minutos.

Se pide:

1. Analizar la serie temporal de consumos eléctricos y describir brevemente las principales características de esta.
2. Realizar un estudio para determinar si la serie temporal presenta *outliers*. Describir brevemente el estudio realizado y las conclusiones alcanzadas.
3. Seleccionar un artículo publicado en una revista científica de prestigio internacional en el que se presente un método de predicción. Estudiar de manera detallada el método publicado y describirlo de forma resumida.
4. Aplicar el método seleccionado en el apartado anterior para obtener una predicción del consumo eléctrico con un horizonte temporal de 4 horas para los días de la semana desde el lunes 8 de junio hasta el domingo 14 de junio. Para ello, se debe implementar un método de predicción en el lenguaje de programación que se estime oportuno o bien usar software libre disponible como WEKA, R, KEEL, etc. Una vez obtenidas las predicciones del periodo indicado, visualizar los resultados (predicciones, errores, etc.).

El trabajo de investigación se realiza utilizando el lenguaje de programación **R** y todos los datos y el código necesarios para reproducir el estudio se proporcionan dentro de este documento para garantizar que éste sea completamente reproducible.

Se han utilizado los siguientes paquetes para el análisis:

```
library(tidyverse)
library(ggplot2)
library(lubridate)
library(xts)
library(forecast)
```

1.2. Datos de la serie temporal

Los datos están disponibles para descargar [aquí](#). Una vez descargados, los cargamos en nuestro espacio de trabajo.

```
data <- read.csv("Demanda_2015.csv", header = FALSE, sep = ",")
```

Para su uso posterior, el marco de datos de la serie temporal se formatea.

```
#
colnames(data) <- c("date", "time", "demand")

#
head(data, 10)
```

```
##           date time demand
## 1  01/01/2015 0:00  25459
## 2  01/01/2015 0:10  25591
## 3  01/01/2015 0:20  25531
## 4  01/01/2015 0:30  25453
## 5  01/01/2015 0:40  25329
## 6  01/01/2015 0:50  25247
## 7  01/01/2015 1:00  25093
## 8  01/01/2015 1:10  24853
## 9  01/01/2015 1:20  24678
## 10 01/01/2015 1:30  24391
```

```
#
summary(data)
```

```
##           date           time           demand
## Length:52560      Length:52560      Min.    :17985
## Class :character   Class :character  1st Qu.:24392
## Mode  :character   Mode  :character  Median :28566
##                                     Mean   :28349
##                                     3rd Qu.:31664
##                                     Max.   :40648
```

Tenemos la variable *date* como *character*, la variable *time* como *character* y la variable *demand* como *integer*.

```
sum(is.na(data))
```

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

```
#
data <- cbind(datetime = paste(data$date, data$time), data)
data$datetime <- parse_date_time(data$datetime, "dmY HMS",
                                truncated = 3, tz = "UTC")

#
# data$date <- NULL
data$time <- NULL

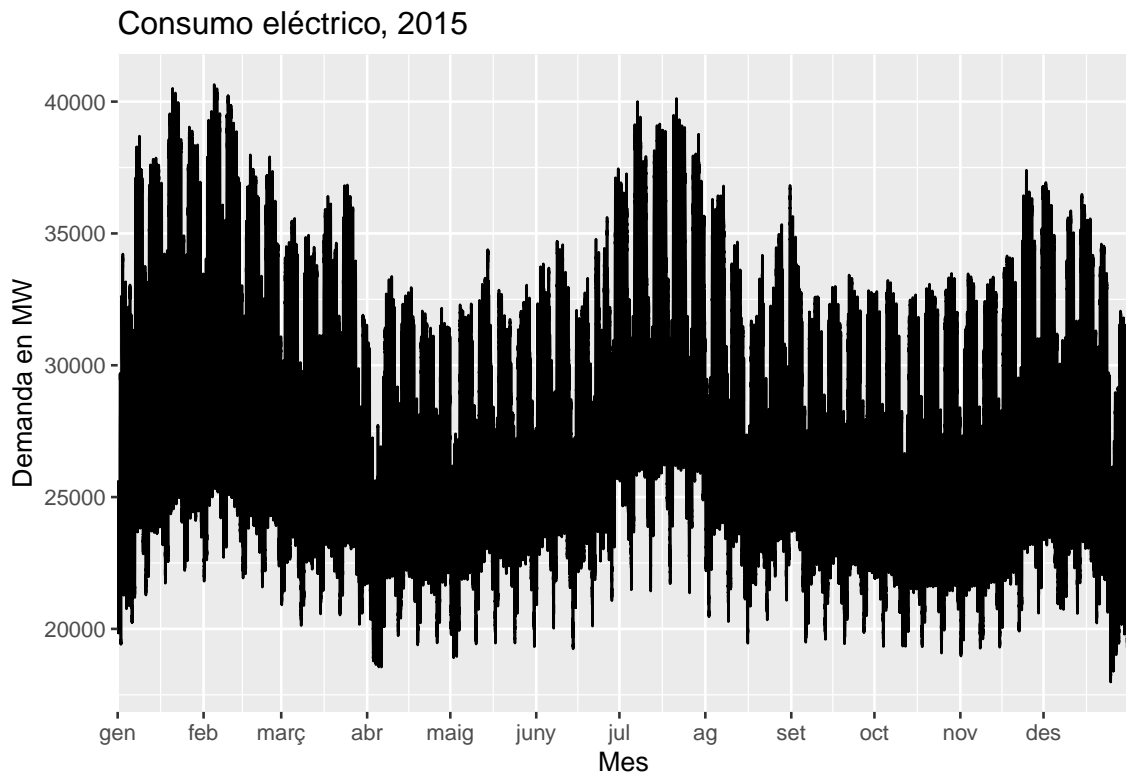
#
head(data, 10)
```

```
##           datetime           date demand
## 1 2015-01-01 00:00:00 01/01/2015 25459
## 2 2015-01-01 00:10:00 01/01/2015 25591
## 3 2015-01-01 00:20:00 01/01/2015 25531
## 4 2015-01-01 00:30:00 01/01/2015 25453
## 5 2015-01-01 00:40:00 01/01/2015 25329
## 6 2015-01-01 00:50:00 01/01/2015 25247
## 7 2015-01-01 01:00:00 01/01/2015 25093
## 8 2015-01-01 01:10:00 01/01/2015 24853
## 9 2015-01-01 01:20:00 01/01/2015 24678
## 10 2015-01-01 01:30:00 01/01/2015 24391
```

1.3. Características de la serie temporal

La siguiente figura ayuda a visualizar la evolución de la demanda a lo largo del año 2015.

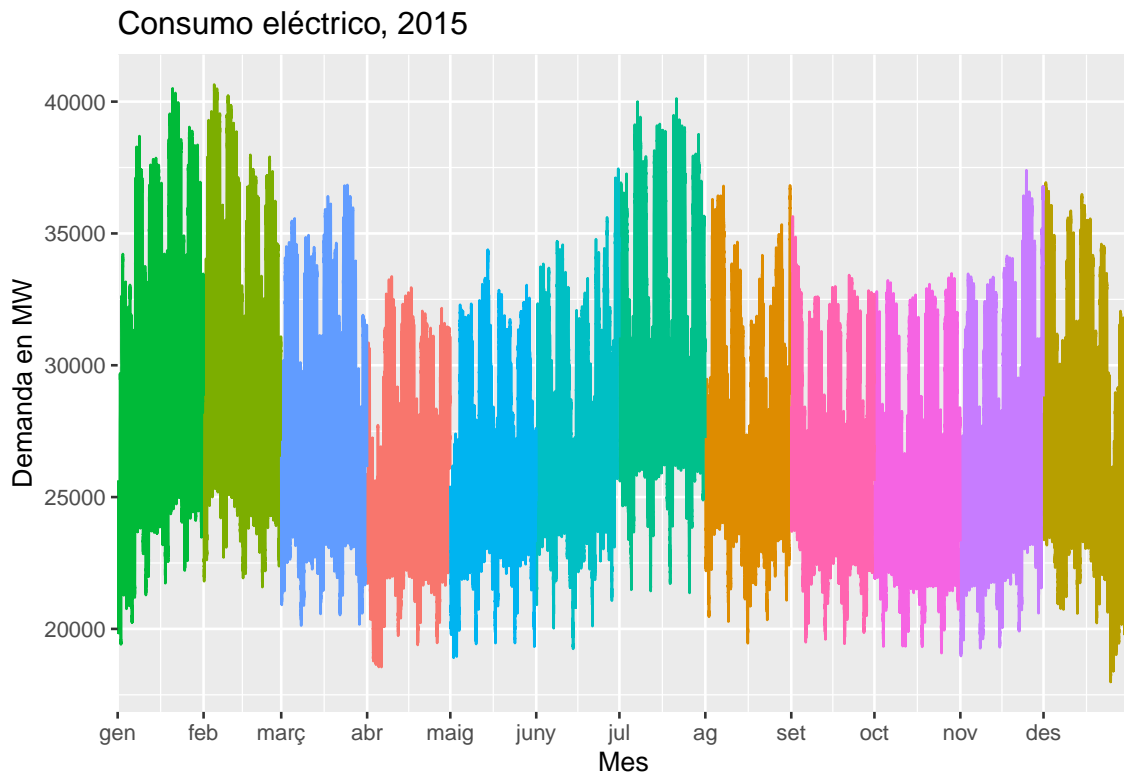
```
ggplot(data = data, aes(x = datetime, y = demand)) +
  geom_line() +
  scale_x_datetime(date_labels = "%b",
                  breaks = "1 month",
                  expand = c(0, 0)) +
  ggtitle("Consumo eléctrico, 2015") +
  xlab("Mes") +
  ylab("Demanda en MW")
```



Observando la figura se puede detectar una dependencia estacional de la demanda, aunque pueden existir otros factores que pueden afectarla, como los días festivos, los fines de semana, ...

Si queremos comparar mejor la distribución de la demanda de energía para cada mes, la siguiente figura es muy útil.

```
ggplot(data, aes(x = datetime, y = demand)) +
  geom_line(aes(colour = month)) +
  scale_x_datetime(date_labels = "%b",
                   breaks = "1 month",
                   expand = c(0, 0)) +
  ggtitle("Consumo eléctrico, 2015") +
  xlab("Mes") +
  ylab("Demanda en MW") +
  theme(legend.position = "none")
```



Durante el invierno y el verano la demanda es de manera clara mayor salvo, probablemente, en los períodos vacacionales. Como se puede observar en el mes de agosto. Podemos deducir de la figura que en sí mismos los datos siguen un patrón de subidas y caídas en una tendencia que varía sumo en los meses de invierno. Esto se debe a que las personas usan la electricidad para enfriar y calentar sus viviendas respectivamente.

Los meses de menor demanda son abril y octubre. Los meses de mayor demanda son enero-febrero y julio.

1.4. Componentes de la serie temporal

Si queremos que **R** trate un objeto como una serie temporal, tenemos que determinar apropiadamente sus características con el comando *ts*. Para definir la serie correctamente escribimos:

```
ts <- ts(data$demand, frequency = 24*60/10)
```

El argumento *frequency* se utiliza para indicar la periodicidad de la serie (en este caso de 10 minutos).

WORKING HERE

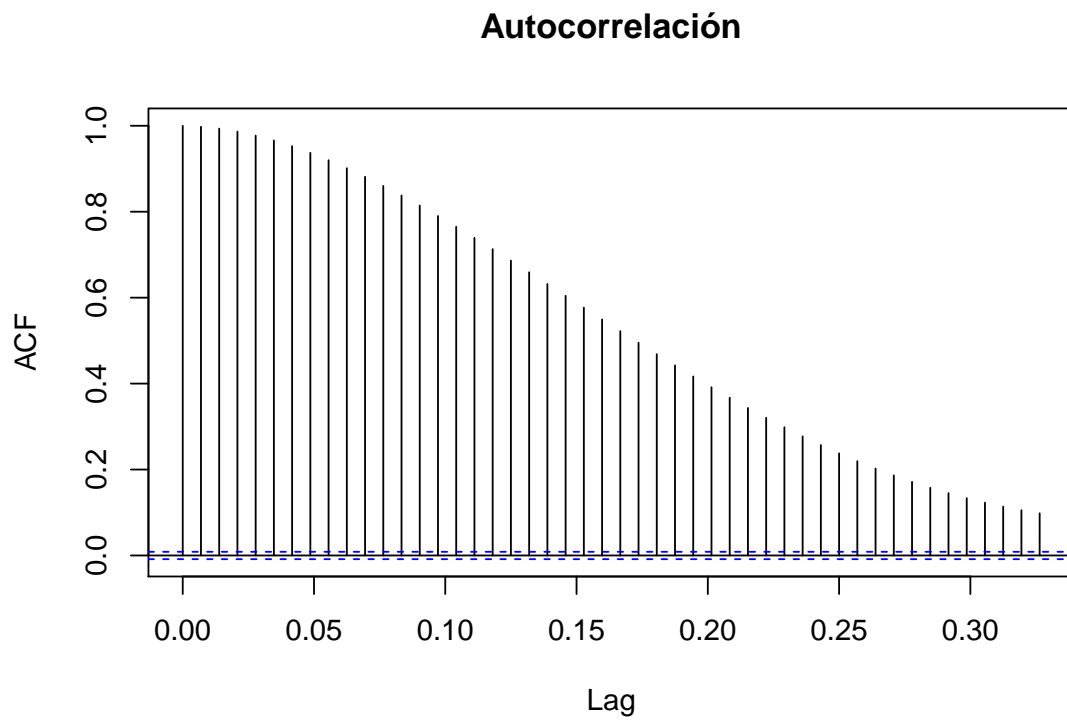
seguir guión de <https://www.diegocalvo.es/analisis-de-series-temporales-en-r-arima/>

Es frecuente analizar las series temporales desde el punto de vista de sus componentes estructurales: *Serie observada = Tendencia + Efecto estacional + Residuos*.

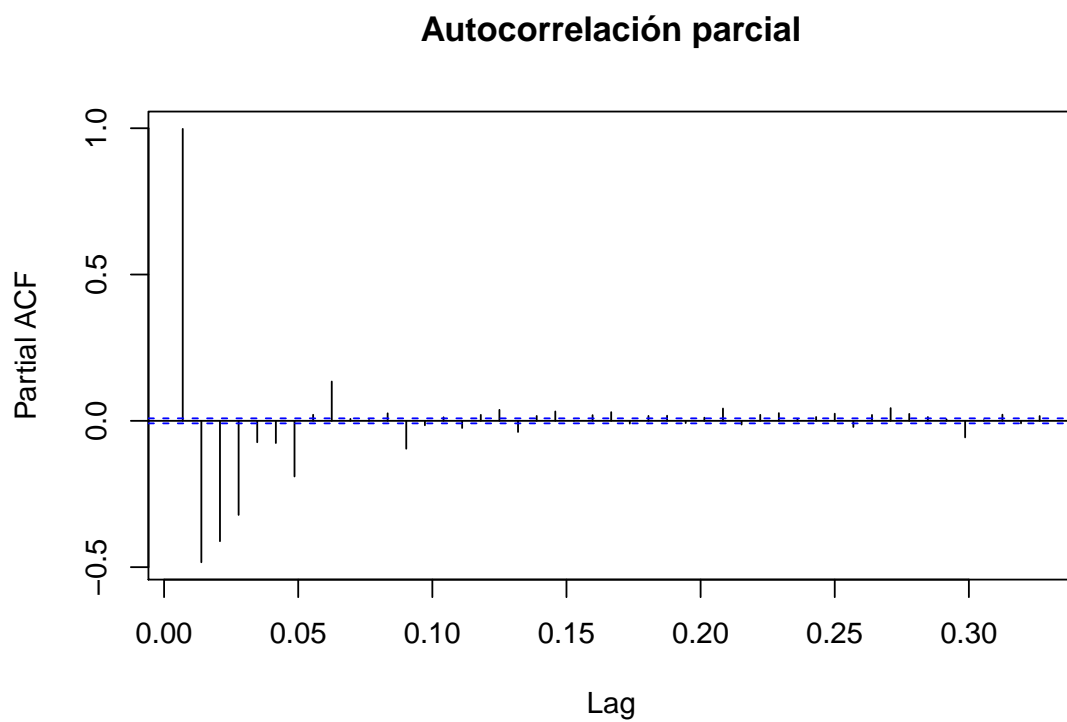
1.4.1. Análisis de autocorrelación y autocorrelación parcial

El análisis de correlación y autocorrelación parcial se usa para examinar la dependencia de la serie.

```
acf(ts, main = "Autocorrelación")
```



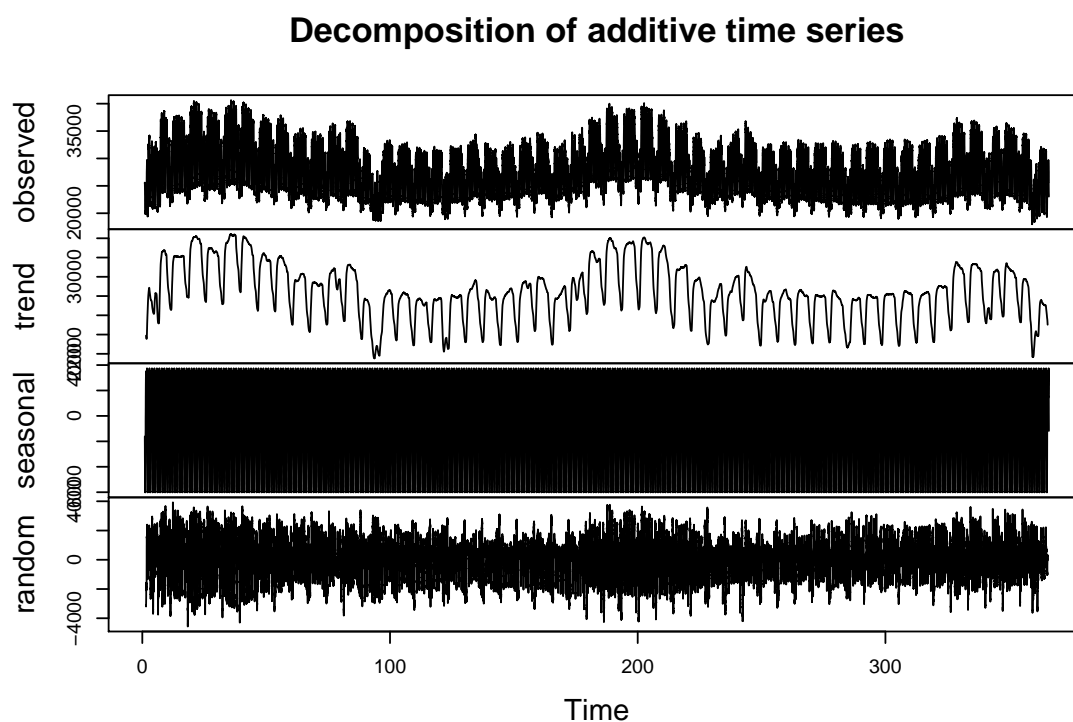
```
pacf(ts, main = "Autocorrelación parcial")
```



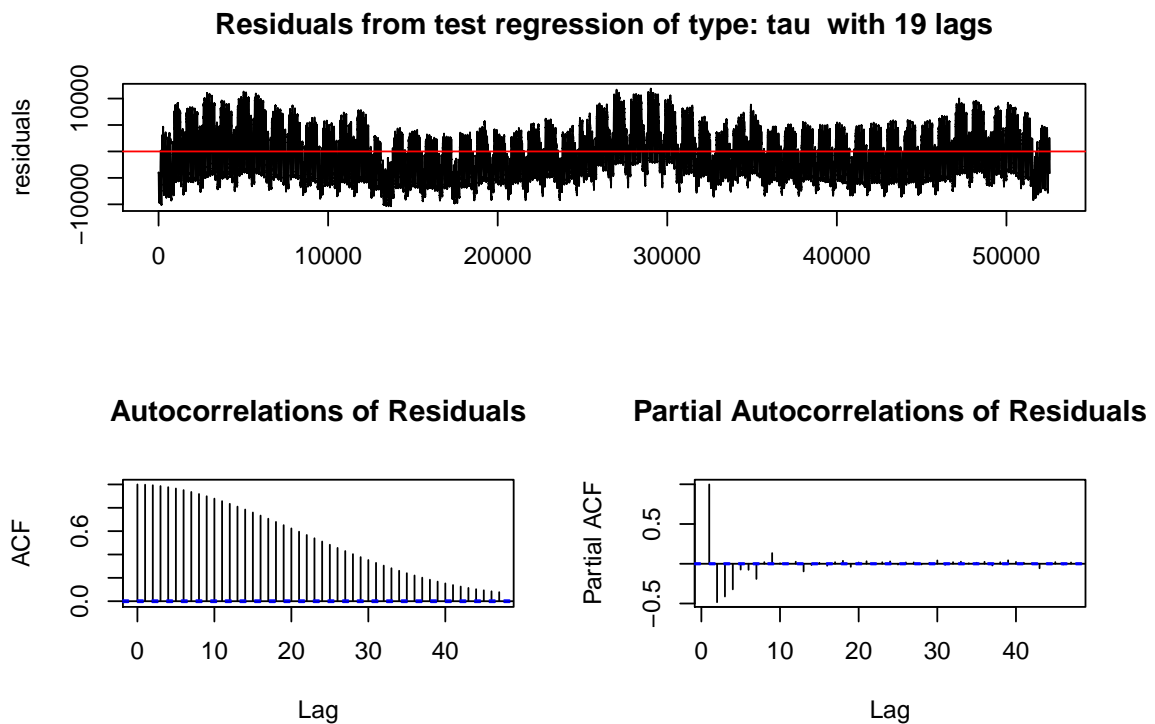
La función de autocorrelación muestra un proceso no estacionario altamente autocorrelacionado, y destaca la estacionalidad. La función de autocorrelación parcial es más difícil de interpretar, pero debido a la forma de *tail-off*, donde los valores cada vez se van haciendo más pequeños, parece un comportamiento típico de los modelos **ARIMA**.

1.4.2. Descomposición de la serie

```
componentes.ts <- decompose(ts)
plot(componentes.ts)
```



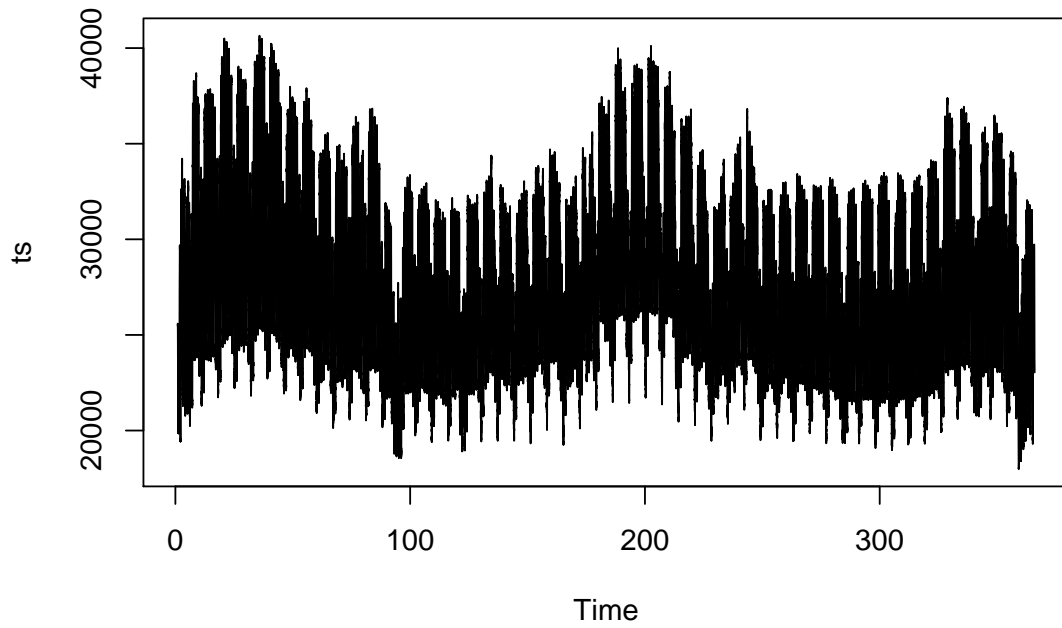
```
library("fUnitRoots")
urkpssTest(ts, type = c("tau"), lags = c("short"), use.lag = NULL, doplot = TRUE)
```



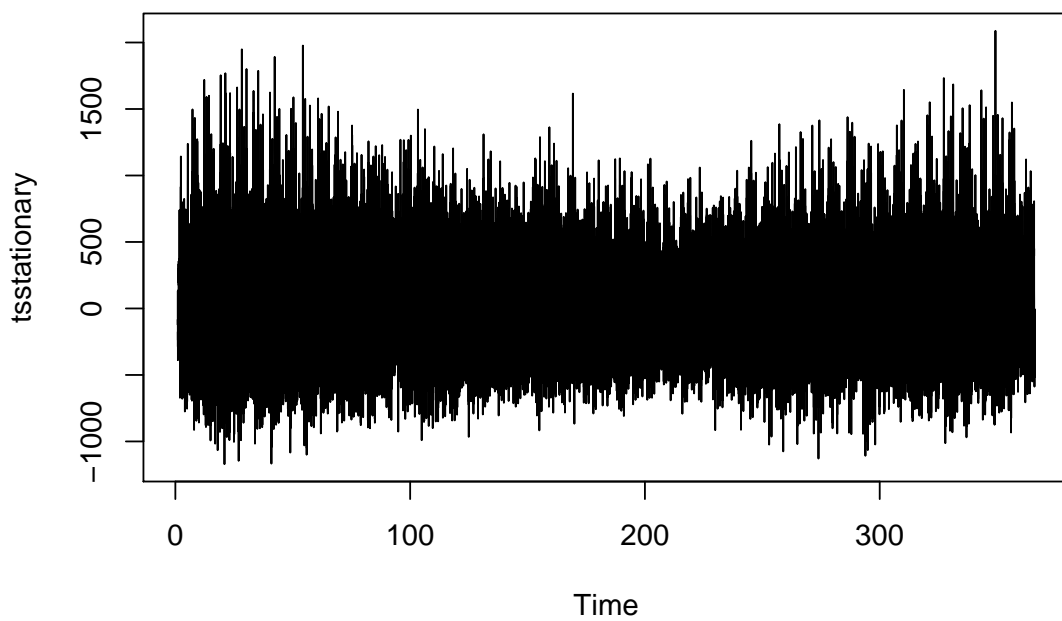
```
##
## Title:
## KPSS Unit Root Test
##
## Test Results:
## NA
##
## Description:
## Tue Nov 24 21:23:23 2020 by user: Laura
```

```
tsstationary = diff(ts, differences=1)
```

```
plot(ts)
```

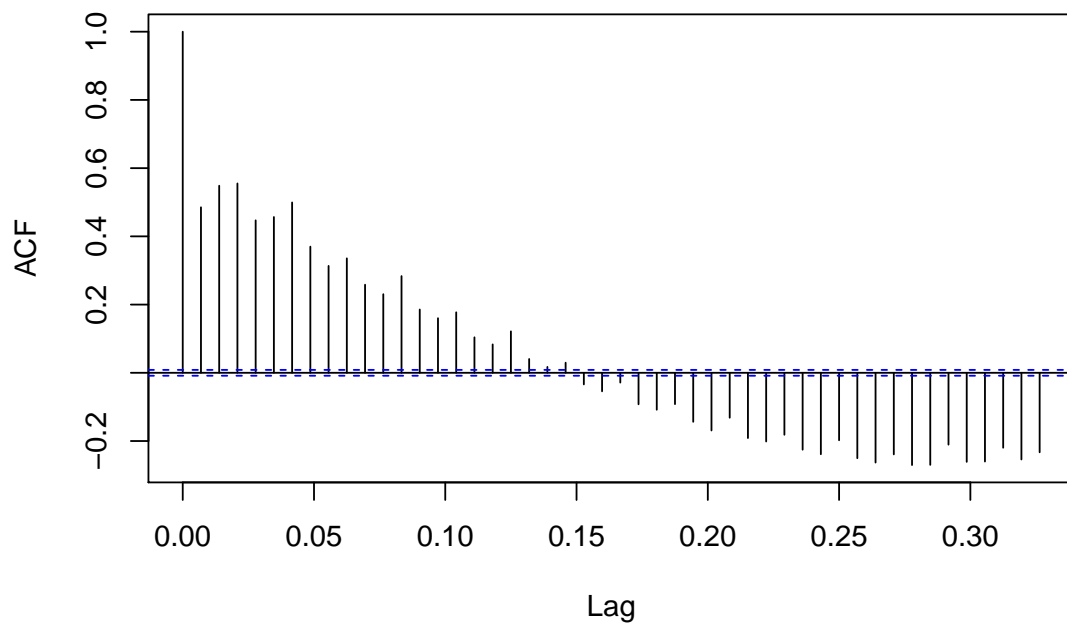


```
plot(tsstationary)
```



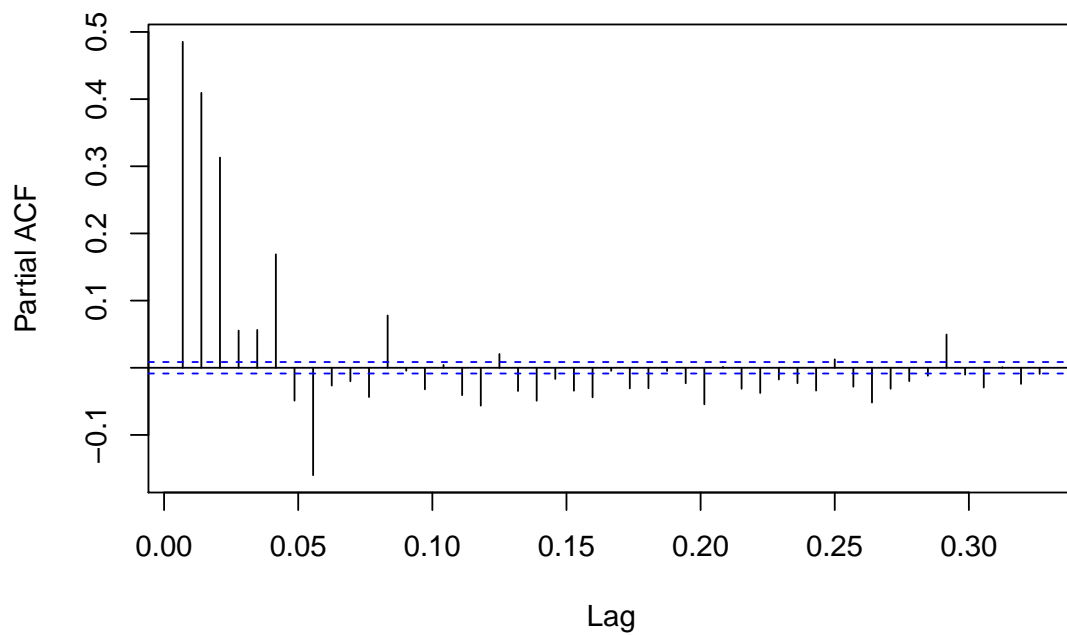
```
acf(tsstationary, main = "Autocorrelación")
```

Autocorrelación



```
pacf(tsstationary, main = "Autocorrelación parcial")
```

Autocorrelación parcial



Capítulo 2

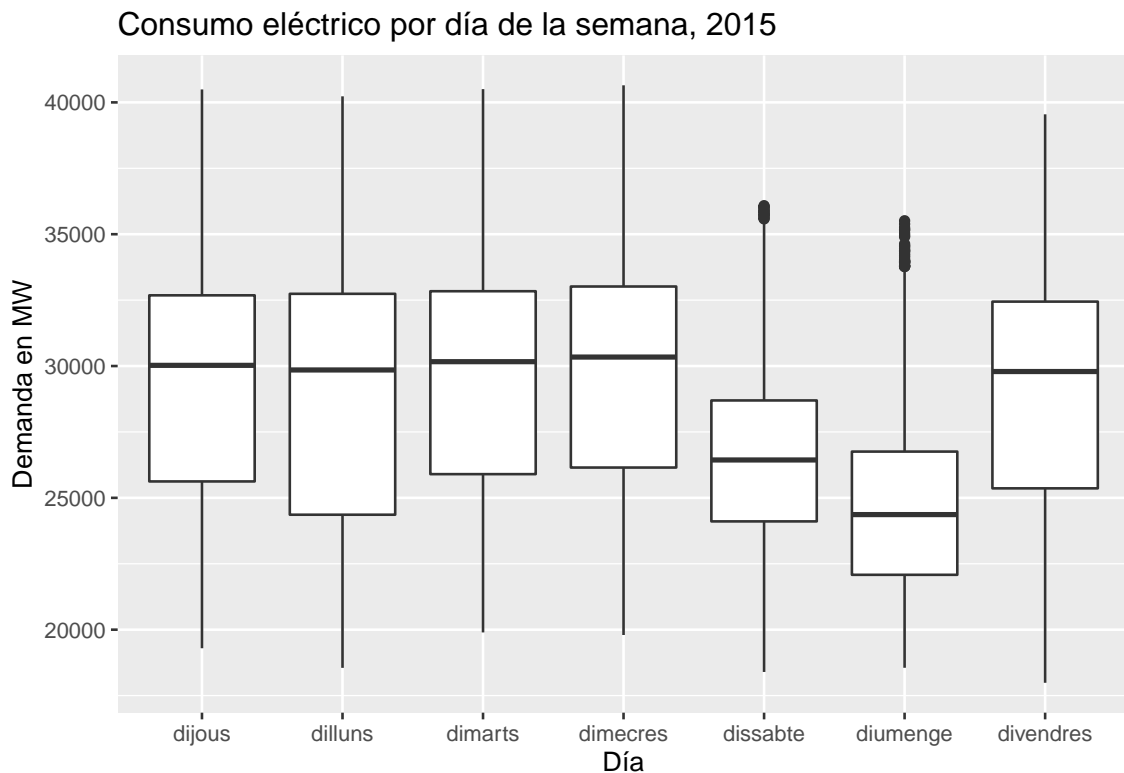
Búsqueda de outliers

Se han utilizado los siguientes paquetes para la búsqueda de outliers:

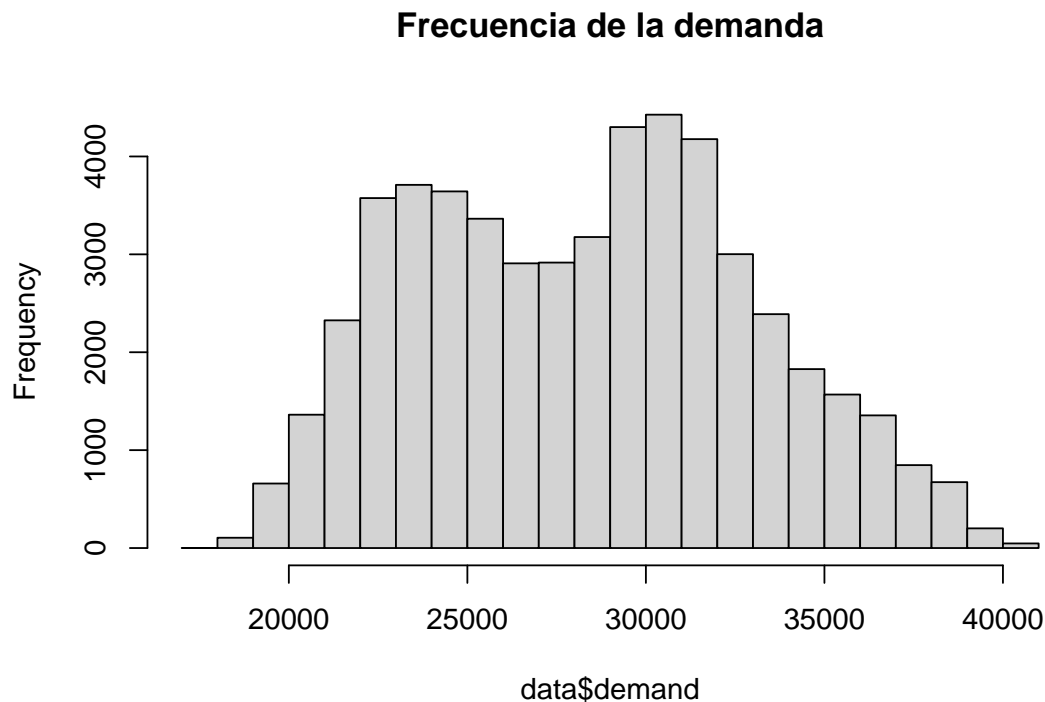
```
library(tidyverse)
library(tsoutliers)
```

Durante los fines de semana, la demanda disminuye considerablemente en comparación con el resto de días de la semana, además observamos la presencia de *outliers*.

```
ggplot(data, aes(day, demand)) +
  geom_boxplot() +
  ggtitle("Consumo eléctrico por día de la semana, 2015") +
  xlab("Día") +
  ylab("Demanda en MW")
```



```
hist(data$demand, main="Frecuencia de la demanda")
```



```
breakPoints <- c(0, 22000, 33000, Inf)
categories <- c("Low", "Medium", "High")
data$demand.F <- cut(data$demand, breaks = breakPoints, labels = categories)
demand_low <- data[data$demand.F == "Low", ]
demand_low2 <- demand_low[demand_low$day != "dissabte" & demand_low$day != "diumenge", ]
dates <- as.Date(demand_low2$date, format = '%d/%m/%Y')
dates_unique <- unique(dates)
dates_unique
```

```
## [1] "2015-01-01" "2015-01-02" "2015-01-05" "2015-01-06" "2015-01-07"
## [6] "2015-01-12" "2015-03-02" "2015-03-09" "2015-03-16" "2015-03-23"
## [11] "2015-03-30" "2015-03-31" "2015-04-01" "2015-04-02" "2015-04-03"
## [16] "2015-04-06" "2015-04-07" "2015-04-08" "2015-04-09" "2015-04-10"
## [21] "2015-04-13" "2015-04-14" "2015-04-16" "2015-04-17" "2015-04-20"
## [26] "2015-04-21" "2015-04-22" "2015-04-23" "2015-04-24" "2015-04-27"
## [31] "2015-04-28" "2015-04-29" "2015-05-01" "2015-05-04" "2015-05-05"
## [36] "2015-05-06" "2015-05-07" "2015-05-11" "2015-05-18" "2015-05-25"
## [41] "2015-06-01" "2015-06-08" "2015-06-15" "2015-06-22" "2015-08-10"
## [46] "2015-08-17" "2015-08-18" "2015-08-19" "2015-08-24" "2015-08-26"
## [51] "2015-09-07" "2015-09-08" "2015-09-14" "2015-09-15" "2015-09-18"
## [56] "2015-09-21" "2015-09-28" "2015-10-01" "2015-10-02" "2015-10-05"
## [61] "2015-10-08" "2015-10-09" "2015-10-12" "2015-10-13" "2015-10-14"
## [66] "2015-10-15" "2015-10-16" "2015-10-19" "2015-10-20" "2015-10-21"
## [71] "2015-10-22" "2015-10-23" "2015-10-26" "2015-10-27" "2015-10-28"
## [76] "2015-10-29" "2015-10-30" "2015-11-02" "2015-11-03" "2015-11-04"
```

```
## [81] "2015-11-05" "2015-11-06" "2015-11-09" "2015-11-10" "2015-11-11"
## [86] "2015-11-12" "2015-11-13" "2015-11-16" "2015-11-17" "2015-11-18"
## [91] "2015-11-19" "2015-11-23" "2015-11-30" "2015-12-07" "2015-12-08"
## [96] "2015-12-09" "2015-12-14" "2015-12-21" "2015-12-23" "2015-12-24"
## [101] "2015-12-25" "2015-12-28" "2015-12-29" "2015-12-30" "2015-12-31"
```

los high pasamos de explicar que date son. ya hay suficientes ejemplos. explicar quartiles y reemplazo de outliers en R.

Capítulo 3

Método de predicción

Capítulo 4

Aplicación del método de predicción

Se aplica el método de predicción descrito en el capítulo anterior para obtener una predicción del consumo eléctrico con un horizonte temporal de 4 horas para los días de la semana desde el lunes 8 de junio hasta el domingo 14 de junio.

Apéndice A

Apéndice: Título del Apéndice

A.1. Primera sección

Apéndice B

Apéndice: Título del Apéndice

B.1. Primera sección

Bibliografía

Pedro L. Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.