

# PEC2 Tipología de datos - Análisis de dataset

PEC2-GRUPO

03/04/2021

## 1.-Carga de archivo

En el siguiente apartado se muestra la carga de archivo de datos desde el csv extraído.

```
archivo<-"C:/Users/HITSCORPTECH/Documents/valores_historicos_de_commodities_en_la_bolsa_de_valores.csv"
datos<-read.csv(archivo,header=TRUE,',')
head(datos,3)
```

```
##   name      date    open  max    min    close adjclose volume
## 1 Arroz 31/03/2021  12.88  13   12.88  12.945  12.945    319
## 2 Arroz 30/03/2021 1327.00 1330 1317.00 1322.000 1322.000    190
## 3 Arroz 29/03/2021 1308.50 1330 1307.00 1326.500 1326.500    190
```

Se procede a observar los tipos de variables, los cuales consta de 1200 observaciones y 8 variables.

```
str(datos)

## 'data.frame':    1200 obs. of  8 variables:
## $ name      : chr  "Arroz" "Arroz" "Arroz" "Arroz" ...
## $ date       : chr  "31/03/2021" "30/03/2021" "29/03/2021" "26/03/2021" ...
## $ open       : num  12.9 1327 1308.5 1312.5 1331 ...
## $ max        : num  13 1330 1330 1319 1338 ...
## $ min        : num  12.9 1317 1307 1308 1314 ...
## $ close      : num  12.9 1322 1326.5 1308.5 1315 ...
## $ adjclose   : num  12.9 1322 1326.5 1308.5 1315 ...
## $ volume     : int   319 190 190 170 318 405 145 188 201 341 ...
```

Se procede a dar formato a la variable date

```
datos$date<-as.Date(datos$date,"%d/%m/%Y")
```

## 2.-Separación de datos

Se crea un nuevo dataframe para obtener los valores de date,name,open,close

```
library(reshape)
```

```
## Warning: package 'reshape' was built under R version 4.0.5
```

```
data1<-data.frame(datos$date,datos$name,datos$open,datos$close)
data2 = rename(data1, c(datos.date="fecha",datos.name="materia",datos.open='abre',datos.close="cierre"))
head(data2)
```

```
##           fecha materia   abre  cierre
## 1 2021-03-31   Arroz  12.88  12.945
## 2 2021-03-30   Arroz 1327.00 1322.000
## 3 2021-03-29   Arroz 1308.50 1326.500
## 4 2021-03-26   Arroz 1312.50 1308.500
## 5 2021-03-25   Arroz 1331.00 1315.000
## 6 2021-03-24   Arroz 1319.00 1339.500
```

## 2.1-Análisis de Oro

Se procede a extraer toda la información referente al oro para el análisis de regresión lineal

```
oro<-subset(data2,data2$materia=="Oro",select = c(fecha,materia,abre,cierre))
head(oro)
```

```
##           fecha materia   abre cierre
## 701 2021-03-31     Oro 1709.0 1708.7
## 702 2021-03-30     Oro 1711.0 1683.9
## 703 2021-03-29     Oro 1712.1 1712.1
## 704 2021-03-26     Oro 1721.3 1732.2
## 705 2021-03-25     Oro 1732.7 1724.9
## 706 2021-03-24     Oro 1727.5 1732.9
```

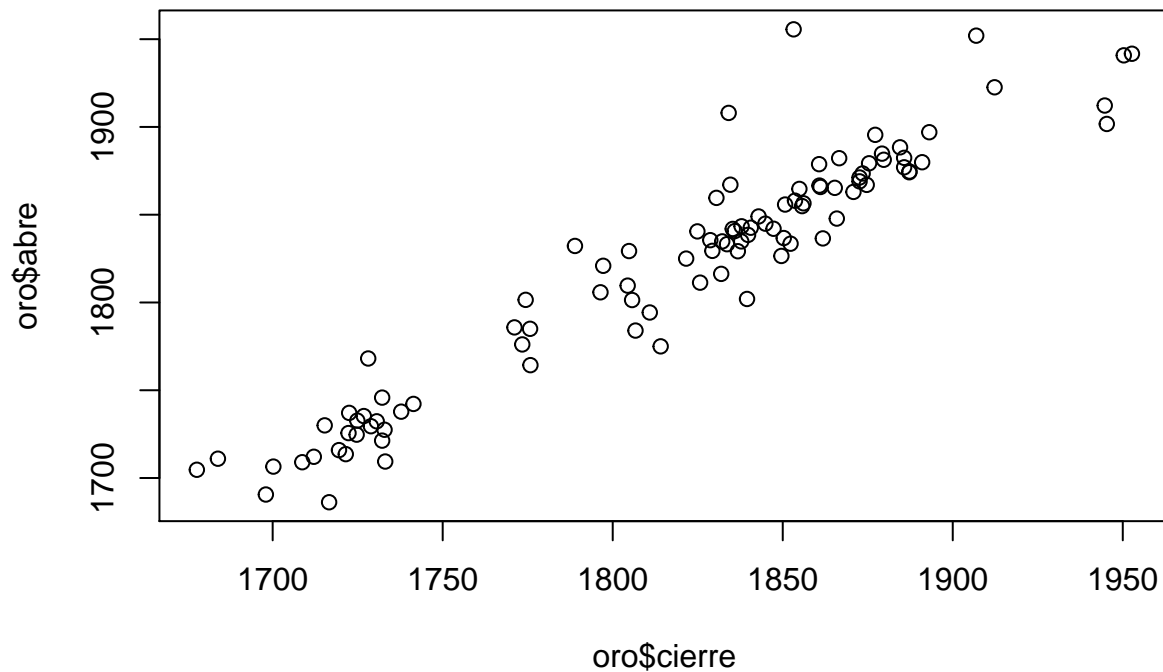
```
tail(oro)
```

```
##           fecha materia   abre cierre
## 795 2020-11-12     Oro 1869.0 1872.6
## 796 2020-11-11     Oro 1878.8 1860.7
## 797 2020-11-10     Oro 1879.3 1875.4
## 798 2020-11-09     Oro 1955.6 1853.2
## 799 2020-11-06     Oro 1940.8 1950.3
## 800 2020-11-05     Oro 1901.7 1945.3
```

### 2.1.1 -Gráfico de dispersión y correlación de Pearson

En el siguiente apartado se puede observar que las variables de apertura y cierre del precio de Oro estan extremadamente relacionadas, como se muestra en la siguiente ilustración:

```
plot(oro$abre~oro$cierre)
```



```
cor.test(oro$abre,oro$cierre)
```

```
##
## Pearson's product-moment correlation
##
## data: oro$abre and oro$cierre
## t = 30.615, df = 96, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9296813 0.9679176
## sample estimates:
##      cor
## 0.9524124
```

La correlación de Pearson, detalla claramente una correlación directamente alta con el 0.95, por lo que se puede rechazar la hipótesis nula y aceptar la correlación en donde destaca que el precio de apertura tiene relación con el precio de cierre.

```
modelooro<-lm(oro$abre~oro$cierre)
summary(modelooro)
```

```
##
## Call:
## lm(formula = oro$abre ~ oro$cierre)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.717  -8.953  -1.388   5.178 101.506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.54381    56.80306   1.453   0.149
## oro$cierre  0.95594     0.03123  30.615 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.5 on 96 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.9071, Adjusted R-squared:  0.9061
## F-statistic: 937.3 on 1 and 96 DF,  p-value: < 2.2e-16
```

En un 90% de R cuadrado ajustado, la variable “y” es explicada con la variable “x”. Se procede a obtener los coeficientes.

```
modelooro$coefficients
```

```
## (Intercept)  oro$cierre
## 82.5438141    0.9559414
```

Según la formula  $y=a+b(x)$ , reemplazando valores seria,  $y=82.54+0.95(x)$ , por lo que si cierra en 1712, es muy probable que abra en:

```
82.5438+(0.95*1712)
```

```
## [1] 1708.944
```

Con este resultado podemos determinar que para el siguiente día el precio del oro estaría abriendo con un valor de 1708.94 dólares.

## 2.2-Analisis de Soja

Se procede a extraer toda la información referente a la Soja para el análisis de regresión lineal

```
soja<-subset(data2,data2$materia=="Soja",select = c(fecha,materia,abre,cierre))
head(soja)
```

```
##      fecha materia  abre  cierre
## 101 2021-03-31   Soja 1368.75 1436.75
## 102 2021-03-30   Soja 1393.00 1366.75
## 103 2021-03-29   Soja 1400.50 1393.00
## 104 2021-03-26   Soja 1414.00 1400.50
## 105 2021-03-25   Soja 1430.75 1414.25
## 106 2021-03-24   Soja 1421.75 1432.75
```

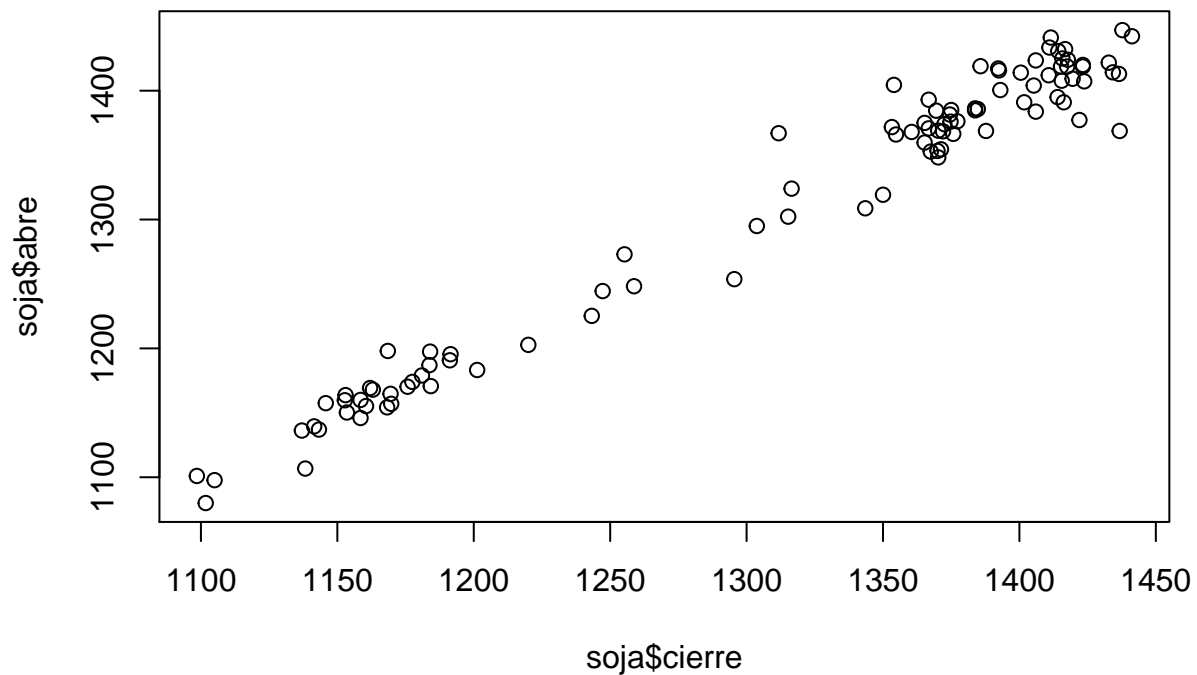
```
tail(soja)
```

```
##      fecha materia  abre  cierre
## 195 2020-11-12   Soja 1136.25 1137.00
## 196 2020-11-11   Soja 1137.00 1143.25
## 197 2020-11-10   Soja 1106.75 1138.25
## 198 2020-11-09   Soja 1097.75 1105.00
## 199 2020-11-06   Soja 1101.00 1098.50
## 200 2020-11-05   Soja 1080.00 1101.75
```

### 2.2.1 -Gráfico de dispersión y correlación de Pearson

En el siguiente apartado se puede observar que las variables de apertura y cierre del precio de Soja estan extremadamente relacionadas, como se muestra en la siguiente ilustración:

```
plot(soja$abre~soja$cierre)
```



```
cor.test(soja$abre,soja$cierre)
```

```
##
## Pearson's product-moment correlation
##
## data: soja$abre and soja$cierre
```

```
## t = 58.286, df = 96, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9793838 0.9907239
## sample estimates:
##      cor
## 0.9861634
```

La correlación de Pearson, detalla claramente una correlación directamente alta con el 0.98 bastante cerca a 1, por lo que se puede rechazar la hipótesis nula y aceptar la correlación en donde destaca que el precio de apertura tiene relación con el precio de cierre.

```
modelosoja<-lm(soja$abre~soja$cierre)
summary(modelosoja)
```

```
##
## Call:
## lm(formula = soja$abre ~ soja$cierre)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -66.778 -10.556   0.320   9.344  56.893
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.06767    22.67377  -0.268    0.79
## soja$cierre  1.00337     0.01721  58.286 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.52 on 96 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.9725, Adjusted R-squared:  0.9722
## F-statistic: 3397 on 1 and 96 DF,  p-value: < 2.2e-16
```

En un 97% de R cuadrado ajustado, la variable “y” es explicada con la variable “x”. Se procede a obtener los coeficientes.

```
modelosoja$coefficients
```

```
## (Intercept) soja$cierre
##      -6.067670      1.003373
```

Según la fórmula  $y=a+b(x)$ , reemplazando valores sería  $y=-6.0676+1.003(x)$ , por lo que si cierra en 1420 para el siguiente día, es muy probable que abra en:

```
-6.0676+(1.003*1420)
```

```
## [1] 1418.192
```

Con este resultado podemos determinar que para el siguiente día el precio de la Soja estaría abriendo con un valor de 1418.19 dólares.