

# 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<- "../data/historical_prices.csv"
datos<-read.csv(archivo,header=TRUE,',')
head(datos,3)
```

```
##   name      date    open      max      min    close adjclose volume
## 1 Soja 12/04/2021 1403.25 1407.25 1392.25 1393.25 1393.25 15220
## 2 Soja 09/04/2021 1416.00 1419.00 1400.75 1403.00 1403.00 98187
## 3 Soja 08/04/2021 1408.75 1422.50 1404.50 1415.25 1415.25 98187
```

Se procede a observar los tipos de variables, los cuales consta de 1100 observaciones y 8 variables.

```
str(datos)
```

```
## 'data.frame':   1100 obs. of  8 variables:
## $ name      : chr  "Soja" "Soja" "Soja" "Soja" ...
## $ date      : chr  "12/04/2021" "09/04/2021" "08/04/2021" "07/04/2021" ...
## $ open      : num  1403 1416 1409 1420 1413 ...
## $ max       : num  1407 1419 1422 1428 1431 ...
## $ min       : num  1392 1401 1404 1404 1412 ...
## $ close     : num  1393 1403 1415 1409 1419 ...
## $ adjclose  : num  1393 1403 1415 1409 1419 ...
## $ volume    : num  15220 98187 98187 84843 81203 ...
```

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)
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="cierres"))
head(data2)
```

```
##      fecha materia  abre cierre
## 1 2021-04-12    Soja 1403.25 1393.25
## 2 2021-04-09    Soja 1416.00 1403.00
## 3 2021-04-08    Soja 1408.75 1415.25
## 4 2021-04-07    Soja 1420.50 1408.75
## 5 2021-04-06    Soja 1412.75 1418.75
```

```
## 6 2021-04-05    Soja 1410.00 1412.75
```

## 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  
## 301 2021-04-12    Oro 1744.5 1741.8  
## 302 2021-04-09    Oro 1750.8 1743.3  
## 303 2021-04-08    Oro 1736.0 1756.8  
## 304 2021-04-07    Oro 1742.7 1740.1  
## 305 2021-04-06    Oro 1726.9 1741.5  
## 306 2021-04-05    Oro 1727.9 1727.0
```

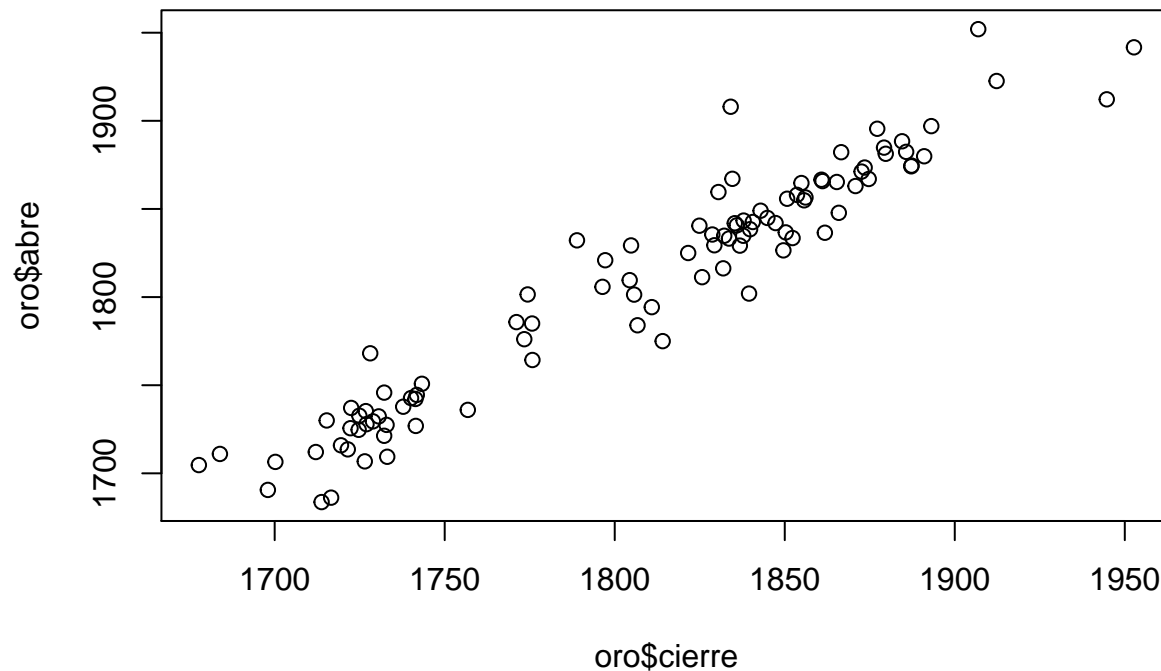
```
tail(oro)
```

```
##      fecha materia  abre cierre  
## 395 2020-11-23    Oro 1834.8 1837.8  
## 396 2020-11-20    Oro 1871.2 1872.6  
## 397 2020-11-19    Oro 1865.8 1861.1  
## 398 2020-11-18    Oro 1873.5 1873.5  
## 399 2020-11-17    Oro 1888.4 1884.5  
## 400 2020-11-16    Oro 1874.6 1887.3
```

### 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 = 35.333, df = 96, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9461172 0.9755297
## sample estimates:
## cor
## 0.9636354
```

La correlación de Pearson, detalla claramente una correlación directamente alta con el 0.96, 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
## -40.070 -10.149   0.096   5.923  73.341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.22527   50.12750   0.763   0.448
## oro$cierre  0.97946    0.02772  35.333 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.84 on 96 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.9286, Adjusted R-squared:  0.9278
## F-statistic: 1248 on 1 and 96 DF, p-value: < 2.2e-16
```

En un 92% de R cuadrado ajustado, la variable “y” es explicada con la variable “x”. Se procede a obtener los coeficientes.

```
modelooro$coefficients
```

```
## (Intercept) oro$cierre
## 38.2252718 0.9794634
```

Según la formula  $y=a+b(x)$ , reemplazando valores seria,  $y=38.23+0.98(x)$ , por lo que si cierra en 1712, es muy probable que abra en:

```
38.23+(0.98*1712)
```

```
## [1] 1715.99
```

Con este resultado podemos determinar que para el siguiente día el precio del oro estaría abriendo con un valor de 1715.99 dólares.

## 2.2-Análisis 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  
## 1 2021-04-12    Soja 1403.25 1393.25  
## 2 2021-04-09    Soja 1416.00 1403.00  
## 3 2021-04-08    Soja 1408.75 1415.25  
## 4 2021-04-07    Soja 1420.50 1408.75  
## 5 2021-04-06    Soja 1412.75 1418.75  
## 6 2021-04-05    Soja 1410.00 1412.75
```

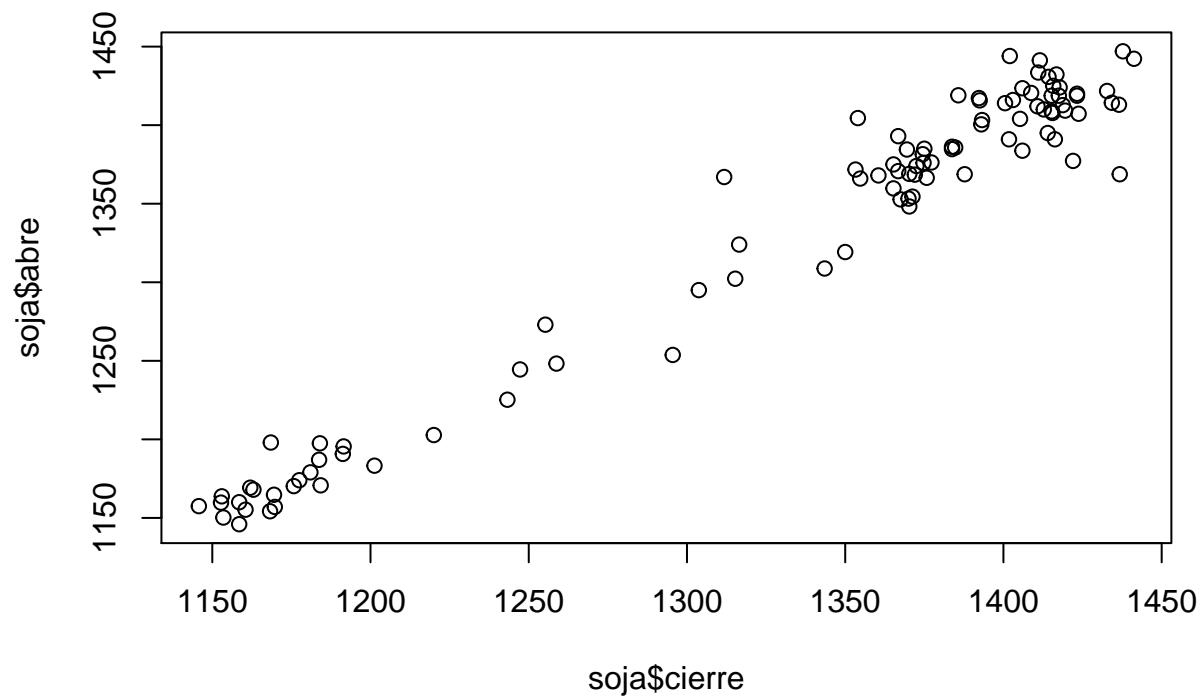
```
tail(soja)
```

```
##      fecha materia  abre  cierre  
## 95 2020-11-23    Soja 1195.50 1191.50  
## 96 2020-11-20    Soja 1179.00 1181.00  
## 97 2020-11-19    Soja 1174.00 1177.50  
## 98 2020-11-18    Soja 1170.25 1175.75  
## 99 2020-11-17    Soja 1157.00 1169.75  
## 100 2020-11-16    Soja 1150.25 1153.50
```

### 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 = 51.135, df = 96, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9734057 0.9880142
## sample estimates:
## cor
## 0.9821334
```

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
## -67.365 -10.386   0.387  10.384  55.517
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.6035     26.0588   0.138   0.89
## soja$cierre   0.9970      0.0195  51.135 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.78 on 96 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.9646, Adjusted R-squared:  0.9642
## F-statistic: 2615 on 1 and 96 DF, p-value: < 2.2e-16
```

En un 96% de R cuadrado ajustado, la variable “y” es explicada con la variable “x”. Se procede a obtener los coeficientes.

```
modelosoja$coefficients
```

```
## (Intercept) soja$cierre
## 3.6034601 0.9970497
```

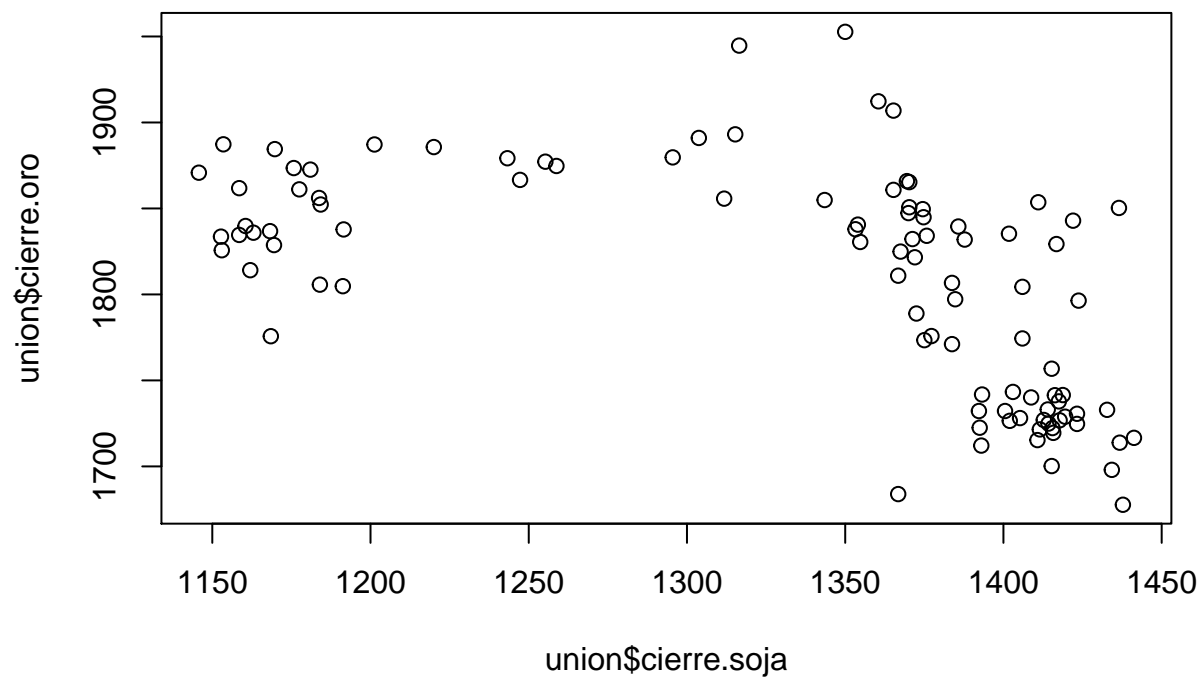
Según la fórmula  $y=a+b(x)$ , reemplazando valores sería  $y=3.60+0.997(x)$ , por lo que si cierra en 1420 para el siguiente día, es muy probable que abra en:

```
3.60+(0.997*1420)
```

```
## [1] 1419.34
```

Con este resultado podemos determinar que para el siguiente día el precio de la Soja estaría abriendo con un valor de 1419.34 dólares.

```
union = merge(x = oro, y = soja, by = "fecha", suffixes = c(".oro", ".soja"))
plot(union$cierre.oro~union$cierre.soja)
```



```
cor.test(union$cierre.oro, union$cierre.soja)
```

```
##
## Pearson's product-moment correlation
##
## data: union$cierre.oro and union$cierre.soja
## t = -6.5487, df = 96, p-value = 2.864e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.6792116 -0.4015310
## sample estimates:
## cor
## -0.5556796
```

En este caso, la correlación de Pearson nos muestra una correlación negativa baja (-0.55), por lo que en este caso no se puede rechazar la hipótesis nula.

Este análisis se podría realizar con todas las materias primas, y lo dejamos para un estudio posterior.