

UNIVERSIDAD CATOLICA BOLIVIANA "SAN PABLO"  
MAESTRIA EN CIENCIA DE DATOS, TERCERA VERSION



Materia: ANALISIS ESTADÍSTICO I  
Practica No.1

Maestrante: Ramón Wilder Serdán Cárdenas

Noviembre 2021  
La Paz – Bolivia

## Contenido

1.- Realizar importación de base de datos de Excel a R, con extensión csv, xls(x), txt .....	3
a. Para importar datos con extensión .xls y .xlsx (de Microsoft Excel) .....	3
b. Para importar datos con extensión .csv (Comma Separated Values) .....	3
c. Para importar datos con extensión .txt (Archivo de texto) .....	4
2.- Busque 4 base de datos con la identificación de todas las variables de cada base.....	4
a. Exportaciones según clasificación Tradicional-No tradicional 2020(Base de datos exportaciones.xls) .....	4
b. Informe mundial sobre la felicidad 2019 (Base de datos Felicidad2019.csv). ....	4
c. Atletas mejor pagados (Base de datos forbesathletes.csv” .....	5
d. Inversión Extranjera Directa 2005-2020 (Base de datos IED.xls).....	5
3.- Con la base de datos asignada debe realizar gráficas, calcular la media, la mediana, moda, Rango intercuartílico RIQ, coeficiente de variación, la covarianza y el coeficiente de correlación y la tabla de doble entrada .....	6
3.1 Base de datos asignada “IBD.xls” .....	6
3.2 Base de datos asignada “Broker.xlsx” .....	9

## 1.- Realizar importación de base de datos de Excel a R, con extensión csv, xls(x), txt

Se asume que los archivos que se van a importar se encuentran en la carpeta de trabajo

### a. Para importar datos con extensión .xls y .xlsx (de Microsoft Excel)

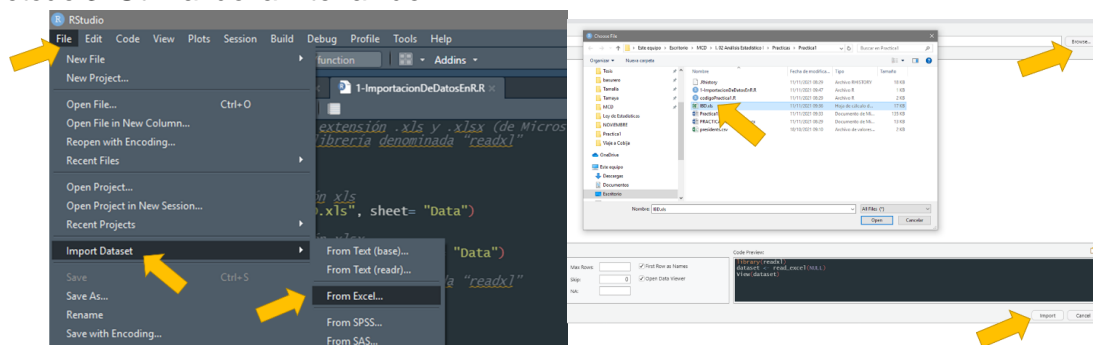
Método 1. Utilizando una librería denominada “readxl”

```
install.packages("readxl")  
library("readxl")  
# para archivos con extensión xls  
misDatos <- read_excel("IBD.xls", sheet="Data")  
View(misDatos)  
# para archivos con extensión xlsx  
misDatos <- read_excel("OtroArchivo.xlsx", sheet="Data")  
View(misDatos)
```

Método 2. Utilizando una librería denominada “xlsx”

```
install.packages("xlsx")  
library("xlsx")  
Datos<-read.xlsx("IBD.xls", 1, header=TRUE)  
View(Datos)
```

Método 3. Utilizando la interfaz de R



### b. Para importar datos con extensión .csv (Comma Separated Values)

Método 1. Utilizando la librería “readr”

```
library(readr)  
presidentes<-read.csv("presidents.csv")  
View(presidentes)
```

c. Para importar datos con extensión .txt (Archivo de texto)

Método 1. Utilizando la librería "readr" y diferentes parámetros

```
library(readr)
consumo<-read.table("consumo.txt", header = TRUE, sep = "",
                    dec = ".")
View(consumo)
```

2.- Busque 4 base de datos con la identificación de todas las variables de cada base

a. Exportaciones según clasificación Tradicional-No tradicional 2020(Base de datos exportaciones.xls)

Información de exportaciones reportados por el Instituto Nacional de Estadística (INE) por la gestión 2020 valorados a valor Free on Board.

Variable	Descripción
Mes	Mes
Clasificacion	Tradicional No Tradicional
Volumen	En Kilogramos Brutos
Valor	En dólares estadounidenses
PaisDestino	País destino de las exportaciones

b. Informe mundial sobre la felicidad 2019 (Base de datos Felicidad2019.csv).

Es una encuesta histórica sobre el estado de la felicidad mundial. Se clasifica a 155 países por sus niveles de felicidad, fue lanzado en las Naciones Unidas en un evento que celebra el Día Internacional de la Felicidad el 20 de marzo. El informe continúa ganando reconocimiento mundial a medida que los gobiernos, las organizaciones y la sociedad civil utilizan cada vez más los indicadores de felicidad para informar sus decisiones de formulación de políticas. Los principales expertos en todos los campos (economía, psicología, análisis de encuestas, estadísticas nacionales, salud, políticas públicas y más) describen cómo las mediciones del bienestar se pueden utilizar de manera eficaz para evaluar el progreso de las naciones.

Variable	Descripción
Overall rank	Puesto en el ranking
Country or region	Nombre del país
Score	Calificación

GDP per capita	La medida en que el PIB contribuye al cálculo del puntaje de felicidad.
Social support	El grado en que la familia contribuye al cálculo del puntaje de felicidad.
Healthy life expectancy	Hasta qué punto la esperanza de vida contribuyó al cálculo de la puntuación de felicidad
Freedom to make life choices	Libertad para tomar decisiones
Generosity	Generosidad
Perceptions of corruption	Grado de Corrupción

c. [Atletas mejor pagados \(Base de datos forbesathletes.csv\)](#)

Lista completa de los atletas mejor pagados del mundo desde la primera lista publicada por Forbes en 1990. En 2002, cambiaron el período del informe del año calendario completo a junio a junio y, en consecuencia, no hay registros de 2001.

Variable	Descripción
Name	Nombre del atleta
Earnings	Ganancias en Dólares
Year	Gestión
Sport	Deporte del atleta

d. [Inversión Extranjera Directa 2005-2020 \(Base de datos IED.xls\)](#)

Encuesta de transacciones realizadas por empresas nacionales con participación accionaria extranjera. Trimestralmente el Banco Central de Bolivia realiza una encuesta a 350 empresas nacionales y éstas reportan datos de movimientos y saldos de activos y pasivos con no residentes en dólares estadounidenses. Se agregan valores por Actividad Económica por confidencialidad.

Variable	Descripción
Gestión	Gestión de la información reportada
Trimestre	Trimestre
País	País del No residente, inversor extranjero
Modalidad	Según categoría funcional
Valor	En dólares estadounidense

3.- Con la base de datos asignada debe realizar gráficas, calcular la media, la mediana, moda, Rango intercuartílico RIQ, coeficiente de variación, la covarianza y el coeficiente de correlación y la tabla de doble entrada

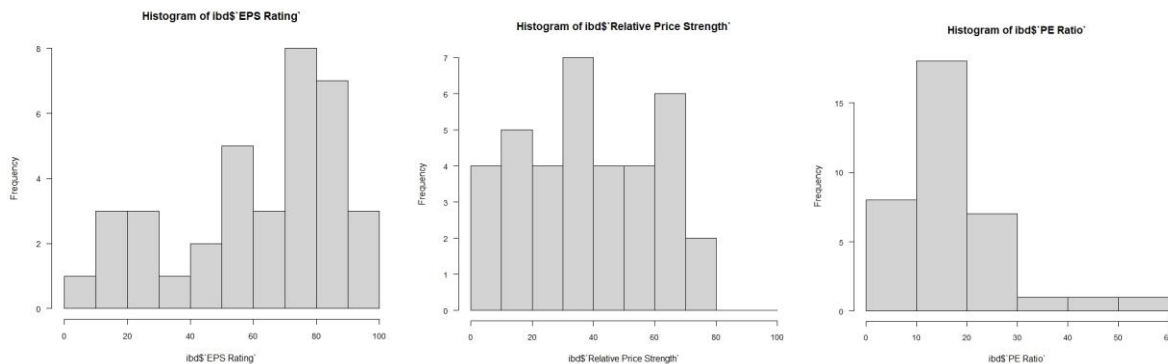
### 3.1 Base de datos asignada “IBD.xls”

La base de datos asignada corresponde a datos financieros de 36 empresas de una muestra cuyas acciones cotizan en la bolsa de valores de Nueva York (Investor's Business Daily, 7 de abril de 2000).

La descripción de los campos es la siguiente:

- **EPS Rating (Calificación EPS)** es una medida del crecimiento por acción.
- **Relative Price Strength (Fuerza relativa del precio)**
- **Relative Strength (Fuerza relativa del grupo de industrias)**
- **Sales/Margins/ROE (Ventas/margen/ROE)** son evaluaciones financieras compuestas que se basan en la tasa de crecimiento de las ventas de una empresa, su margen de ganancia y su rendimiento de los activos (ROE return on capital employed)
- **PE Ratio (price-to-earnings ratio, Razón precio-beneficio)** es la relación entre el precio o valor y los beneficios que se usa en el análisis fundamental de las empresas, especialmente en aquellas que cotizan en bolsa. Su valor indica cuántas veces se está pagando el beneficio neto anual de una empresa determinada al comprar una acción de esta. Un PER más elevado implica que los inversores están pagando más por cada unidad de beneficio

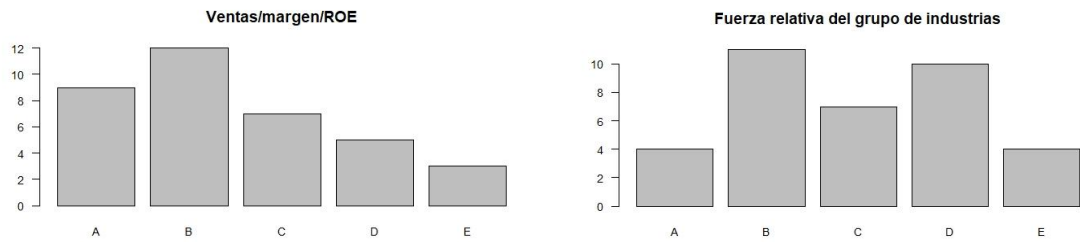
### Histogramas de las variables cuantitativas



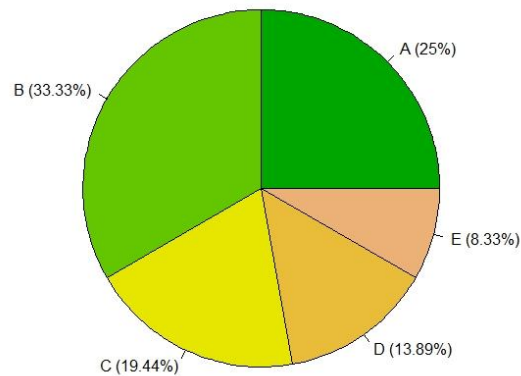
### Gráficos de barras de las variables cualitativas

De acuerdo a las evaluaciones financieras basadas en la tasa de crecimiento de las ventas de una empresa, su margen de ganancia y su rendimiento de los activos, la mayor parte de las empresas de la muestra se encuentran en la categoría C. Por otro lado, de acuerdo al ranking

de la fuerza relativa del grupo de industrias, las empresas de las muestra se encuentran en los grupos B y D



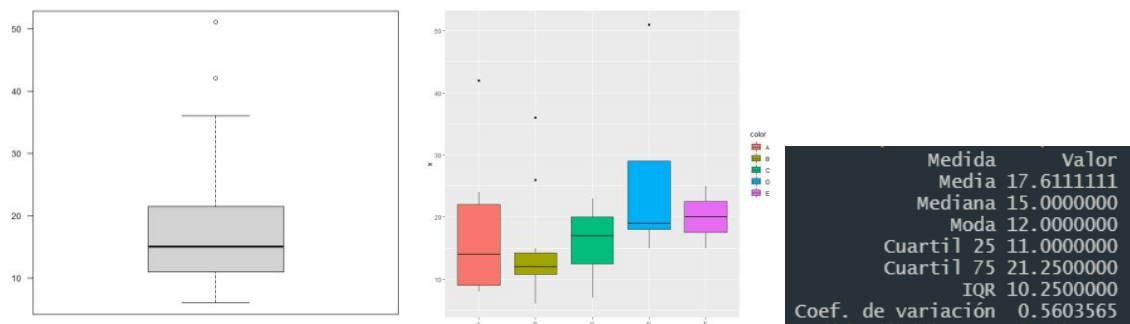
**Sales/Margins/ROE**



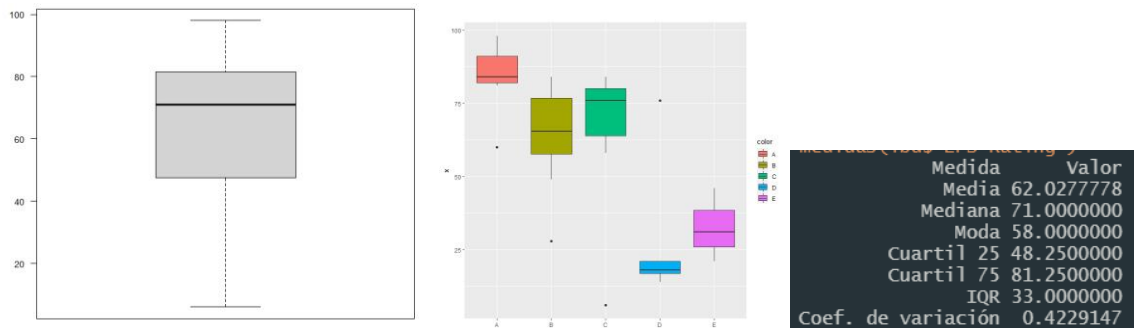
### Boxplot de las tres variable cuantitativas:

La variable “PE Ratio” muestra una proporción media de 17,6 y un coeficiente de variación alto debido a 2 valores atípicos. La variable “EPS Rating” tiene una media de crecimiento por acción de 17,6 y no presenta valores atípicos.

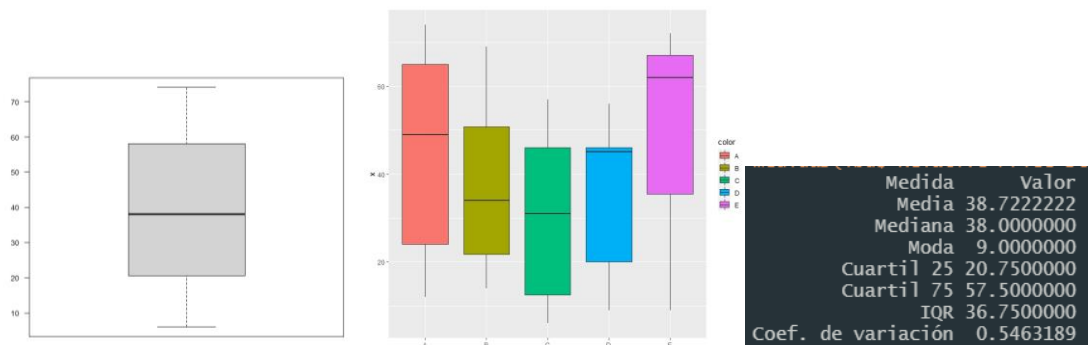
### PE Ratio



## EPS Rating



## Relative Price Strength



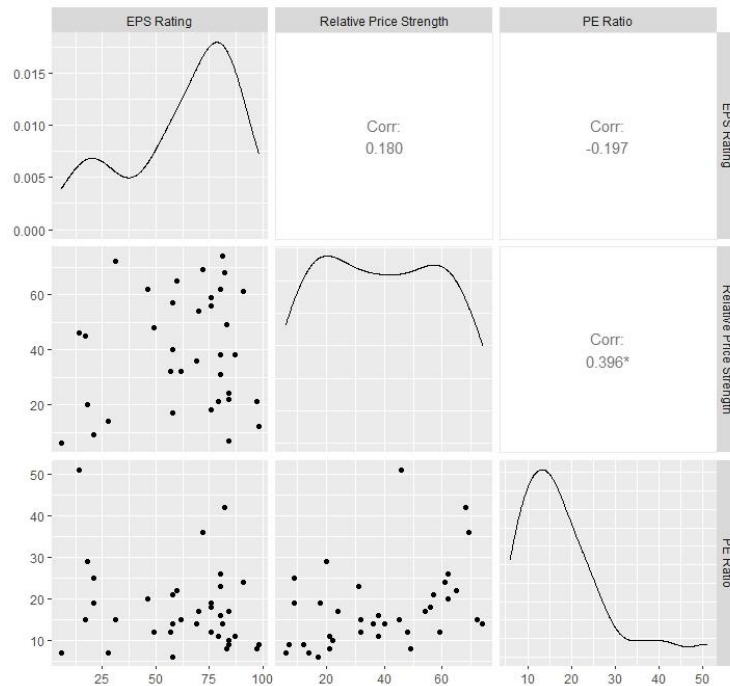
## Covarianza y el coeficiente de correlación

Come se puede apreciar en el cuadro siguiente, las variables “Fuerza relativa del precio” y “Razón precio-beneficio” muestran un mayor grado de asociación, seguido del par “Fuerza relativa del precio” y “Calificación EPS”. Por otro lado, la variable “Calificación EPS” se correlacionan negativamente con la variable y “Razón precio-beneficio”.

```
> asociacion(ibd$`EPS Rating`,ibd$`PE Ratio`)
Medida Valor
1 Covarianza -50.9031746
2 Coef. de Correlación -0.1966322
> asociacion(ibd$`Relative Price Strength`,ibd$`EPS Rating`)
Medida Valor
1 Covarianza 100.0650794
2 Coef. de Correlación 0.1803171
> asociacion(ibd$`Relative Price Strength`,ibd$`PE Ratio`)
Medida Valor
1 Covarianza 82.7746032
2 Coef. de Correlación 0.3964965
>
```



Asimismo, se corroboraron los resultados con el siguiente gráfico utilizando la librería GGally en el que también se muestran los gráficos de dispersión entre las variables cuantitativas.



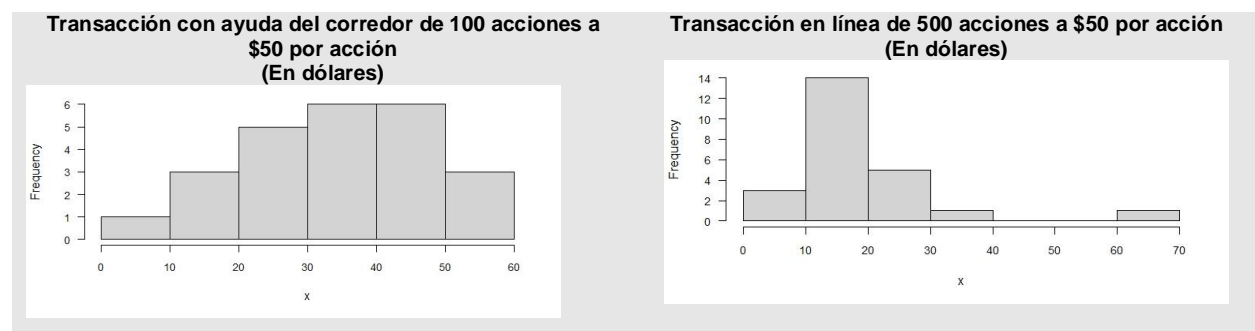
### 3.2 Base de datos asignada “Broker.xlsx”

La base de datos asignada corresponde a comisiones de 24 corredores de bolsa que cobran los corredores de bolsa con descuento por dos tipos de transacciones (AII Journal, enero de 2003):

- Transacción con ayuda del corredor de 100 acciones a \$50 por acción
- Transacción en línea de 500 acciones a \$50 por acción

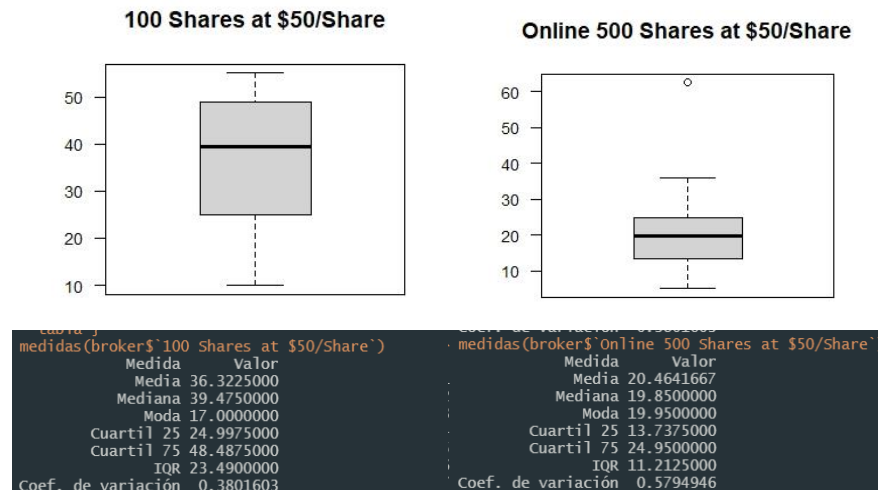
### Histogramas

Las transacciones con ayuda del corredor muestran un comportamiento con distribución normal. En contraposición, las transacciones on-line suponen un costo relativamente bajo, además el volumen transado es muy superior a las primeras. Se deduce que el esfuerzo para llevar a cabo las transacciones online es menor.



## Boxplot

Se observa que la media en las transacciones “manuales” es mayor, es decir, tienen un costo mayor en promedio en comparación a las transacciones online. Esto también se verifica por el valor alto de la mediana. Asimismo, las transacciones online muestran un solo valor atípico.



## Covarianza y el coeficiente de correlación

Según el coeficiente de correlación se deduce un grado de relación importante. El valor atípico en las transacciones online puede afectar a esta conclusión.

Medida	Valor
Covarianza	76.003911
Coef. de Correlación	0.464141

