

market activity

Changes in the activity of the active and passive market is uncertain. Established positive trends in various market segments.

Distribution of the securities market key players

INICIO: ABRIL 2018

**CURSO VIRTUAL**

# ▶ PROGRAMACIÓN ESTADÍSTICA

CON



**UNIVERSIDAD PERUANA  
CAYETANO HEREDIA**  
ESCUELA DE POSGRADO

Changes in the activity of the active and passive market is uncertain. Established positive trends in various market segments.

Distribution of the securities market key players

# Programación Estadística con R

UPCH

Abril 2018



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**  
Escuela de Posgrado

# 2.

# Talleres de R y Rstudio



## Ejemplo 1

Se dispone de una muestra de 10 personas para las que se han medido las variables:

- ▶ EDAD (en años)
- ▶ SEXO (codificada como 0: "Hombre"; 1: "Mujer")
- ▶ NIVEL DE ESTUDIOS ( 0: "Sin estudios"; 1: "Estudios Primarios"; 2: "Estudios Secundarios"; 3: "Estudios Superiores")



# Estadística Descriptiva con R (II)

## Ejemplo 1

Los datos están almacenados en los siguientes vectores :

```
1 edad=c(18,19,NA,18,24,17,22,15,22,25)
2 sexo=c(0,1,0,0,1,0,0,1,1,0)
3 estudios=c(1,2,0,1,3,2,3,1,2,3)
```



# Estadística Descriptiva con R (III)

## Ejemplo 1

Con estos datos se desea:

- ▶ Construir tablas de frecuencias unidimensionales para las tres variables.
- ▶ Construir una tabla de frecuencias cruzadas con las variables 'sexo' y 'nivel de estudios'.
- ▶ Construir una tabla de frecuencias de la variable edad agrupada en intervalos.
- ▶ Calcular media y desviación típica de la variable 'edad'.
- ▶ Calcular media y desviación típica de la variable 'edad' según sexo.
- ▶ Representar gráficamente: el nivel de estudios en un diagrama de sectores, el sexo en un diagrama de barras y la edad en un histograma. Construir también un boxplot de la edad en función del sexo.

# Estadística Descriptiva con R (IV)

## Ejemplo 1

Recodificamos como factores el sexo y el nivel de estudios, asignando las etiquetas adecuadas:

```
1 sexo=factor(sexo , levels=c(0,1) , labels=c("Hombre" , "Mujer" ))
2 estudios=factor(estudios , levels=c(0,1,2,3) , labels=c("Sin
    estudios" , "Estudios Primarios" , "Estudios Secundarios" ,
    Estudios Superiores" ))
```



## Ejemplo 1: Análisis de los datos.

### Tablas de frecuencias unidimensionales

- 1 Construir una tablas de frecuencias unidimensionales para las variables **sexo** y **nivel de estudios**.

En R los comandos para la construcción de tablas son:

- ▶ `table(nombre_de_variable)`: tablas de frecuencias absolutas.
- ▶ `prop.table(table(nombre_de_variable))`: tablas de frecuencias relativas. Nótese aquí que la función `prop.table()` no opera directamente sobre la variable, sino sobre la tabla de frecuencias absolutas creada con `table()`!!. Esta es una característica habitual de R: utilizar funciones que operan sobre objetos creados previamente por otras funciones.



# Estadística Descriptiva con R (VI)

## Ejemplo 1: Análisis de los datos.

```
1 table(edad)
```

```
1 prop.table(table(edad))
```



# Estadística Descriptiva con R (VII)

```
script_Unidad2_Parte3.R x
Source on Save Run Source
1 edad=c(18,19,NA,18,24,17,22,15,22,25)
2 sexo=c(0,1,0,0,1,0,0,1,1,0)
3 estudios=c(1,2,0,1,3,2,3,1,2,3)
4
5 sexo=factor(sexo, levels=c(0,1),labels=c("Hombre","Mujer"))
6 estudios=factor(estudios, levels=c(0,1,2,3),
7               labels=c("Sin estudios","Estudios Primarios","Estudios Secundarios","Estudios Superiores"))
8
9 table(edad)
10 prop.table(table(edad))
11

11:1 (Top Level) R Script
Console Terminal
> edad=factor(edad, levels=c(0,1,2,3), labels=c("Hombre","Mujer"))
> estudios=factor(estudios, levels=c(0,1,2,3), labels=c("Sin estudios","Estudios Primarios","Estudios Secundarios","Estudios Superiores"))
> edad=c(18,19,NA,18,24,17,22,15,22,25)
> sexo=c(0,1,0,0,1,0,0,1,1,0)
> estudios=c(1,2,0,1,3,2,3,1,2,3)
> sexo=factor(sexo, levels=c(0,1),labels=c("Hombre","Mujer"))
> estudios=factor(estudios, levels=c(0,1,2,3),
+               labels=c("Sin estudios","Estudios Primarios","Estudios Secundarios","Estudios Superiores"))
+
> table(edad)
edad
15 17 18 19 22 24 25
 1  1  2  1  2  1  1
> prop.table(table(edad))
edad
      15      17      18      19      22      24      25
0.1111111 0.1111111 0.2222222 0.1111111 0.2222222 0.1111111 0.1111111
>
```



# Estadística Descriptiva con R (VIII)

## Ejemplo 1: Análisis de los datos.

Nótese que, por defecto, R no muestra cuantos valores perdidos hay en los datos. Para ello debemos utilizar la opción `useNA="ifany"`:

```
1 table (edad , useNA= " ifany " )
```

```
## edad
##    15    17    18    19    22    24    25 <NA>
##     1     1     2     1     2     1     1     1
```



## Ejemplo 1: Análisis de los datos.

### **Observacion 1** (Ejercicio)

Construye tablas de frecuencias para las variables sexo y Nivel de estudios



# Estadística Descriptiva con R (X)

## Ejemplo 1: Análisis de los datos.

### Tablas de frecuencias cruzadas

2. Construir una tabla de frecuencias cruzadas con las variables anteriores.

Las tablas de frecuencias cruzadas se construyen también con las funciones `table()` y `prop.table(table())`, especificando ahora las dos variables a cruzar. Para las frecuencias relativas, en `prop.table()` incluimos un 1 si las queremos por filas, un 2 si por columnas, o nada si queremos las frecuencias relativas globales:

```
1 table(estudios , sexo)
2
3 prop.table(table(estudios , sexo))
4
5 prop.table(table(estudios , sexo) ,1)
6
7 prop.table(table(estudios , sexo) ,2)
```

# Estadística Descriptiva con R (XI)

## Ejemplo 1: Análisis de los datos.

Nótese que hemos tenido que escribir mucho para obtener estas tablas.

Cuando un objeto en R se repite muchas veces, resulta conveniente asignarle un nombre que nos permita acceder a él fácilmente. Así, en este caso, podíamos haber construido las tablas anteriores mediante:

```
1 t=table(estudios , sexo)
2 prop.table(t)
3 prop.table(t ,1)
4 prop.table(t ,2)
```

```
> table(estudios,sexo)
      sexo
estudios  Hombre Mujer
Sin estudios      1     0
Estudios Primarios  2     1
Estudios Secundarios 1     2
Estudios Superiores 2     1
> prop.table(table(estudios,sexo))
      sexo
estudios  Hombre Mujer
Sin estudios    0.1    0.0
Estudios Primarios 0.2    0.1
Estudios Secundarios 0.1    0.2
Estudios Superiores 0.2    0.1
> prop.table(table(estudios,sexo),1)
      sexo
estudios  Hombre  Mujer
Sin estudios 1.0000000 0.0000000
Estudios Primarios 0.6666667 0.3333333
Estudios Secundarios 0.3333333 0.6666667
Estudios Superiores 0.6666667 0.3333333
> prop.table(table(estudios,sexo),2)
      sexo
estudios  Hombre  Mujer
Sin estudios 0.1666667 0.0000000
Estudios Primarios 0.3333333 0.2500000
Estudios Secundarios 0.1666667 0.5000000
Estudios Superiores 0.3333333 0.2500000
> |
```

# Estadística Descriptiva con R (XII)

## Ejemplo 1: Análisis de los datos.

3. Tabla de frecuencias de una variable continua agrupada en intervalos.

La función

```
1 # incluimos na.rm=TRUE para que ignore la presencia de
2 # valores perdidos
3 range(edad, na.rm=TRUE)
```

Nos indica que los valores de edad se sitúan en el rango de 15 a 25 años. Para construir los intervalos de edad podemos utilizar la regla de Sturges, que nos proporciona un número adecuado de intervalos en función del rango de los datos (ver `help(nclass.Sturges)`); de modo alternativo podría utilizarse la regla de Scott o la de Freedman-Diaconis).

### Ejemplo 1: Análisis de los datos.

Una vez que hemos decidido el número de intervalos, generamos una secuencia de valores con los límites de cada intervalo y utilizamos la función `cut()` para que nos construya una nueva variable que recodifica la edad en dichos intervalos:

```
1 # Numero de intervalos
2 nclass.Sturges(edad)
3
4 # Limites de los intervalos
5 seq(15,25,length=nclass.Sturges(edad))
```





# Estadística Descriptiva con R (XIV)

## Ejemplo 1: Análisis de los datos.

Construcción de los intervalos mediante la función cut():

```
1 intervalosEdad=cut(edad,breaks=seq(15,25,length=nclass.Sturges(edad)),include.lowest=TRUE)
```

Se muestran los intervalos de edad, uno correspondiente a cada edad observada

```
1 intervalosEdad
```

Por último construimos la tabla de frecuencias de esta variable:

```
1 table(intervalosEdad)
```

# Estadística Descriptiva con R (XV)

```
> range(edad,na.rm=TRUE)
[1] 15 25
> nclass.Sturges(edad)
[1] 5
> seq(15,25,length=nclass.Sturges(edad))
[1] 15.0 17.5 20.0 22.5 25.0
> intervalosEdad=cut(edad,breaks=seq(15,25,length=nclass.Sturges(edad)),include.lowest=TRUE)
> table(intervalosEdad)
intervalosEdad
[15,17.5] [17.5,20] [20,22.5] [22.5,25]
      2         3         2         2
> |
```



# Estadística Descriptiva con R (XVI)

## Ejemplo 1: Análisis de los datos.

4. Calcular media y desviación estándar de la variable 'edad'.

Las funciones disponibles en R para calcular estos estadísticos descriptivos son, respectivamente, `mean(nombre_de_variable)` y `sd(nombre_de_variable)`:

```
1 mean(edad)
2 sd(edad)
```



## Ejemplo 1: Análisis de los datos.

R dispone de una función genérica `summary()`, que cuando se aplica a una variable presenta un pequeño resumen descriptivo. Si la variable es numérica, dicho resumen incluye el mínimo, máximo, mediana, primer y tercer cuartiles, media y número de valores perdidos. Si la variable es de tipo factor, `summary()` muestra el número de observaciones en cada nivel del factor:

```
1 summary(edad)
2 summary(estudios)
3 summary(sexo)
```



# Estadística Descriptiva con R (XVIII)

## Ejemplo 1: Análisis de los datos.

Si las variables están contenidas en un `data.frame` la aplicación de `summary()` lleva a cabo un resumen descriptivo de todas las variables que contiene:

```
1 misDatos=data.frame(edad, estudios , sexo)
2 summary(misDatos)
```



# Estadística Descriptiva con R (XIX)

## Ejemplo 1: Análisis de los datos.

5. Calcular media y desviación típica de la variable edad por sexo.

Para ello utilizamos la función `aggregate()`:

```
1 aggregate(edad, by= list(sexo), mean)
```

En el caso de los hombres nos encontramos nuevamente con el problema de los valores perdidos. Para obtener el resultado prescindiendo de dichos valores, es preciso utilizar la opción `na.rm=TRUE` (`na.rm` es un acrónimo de `na remove`):

```
1 aggregate(edad, by= list(sexo), mean, na.rm=TRUE)
```

## Ejemplo 1: Análisis de los datos.

Existen algunas librerías o paquetes que contienen funciones más elaboradas para la obtención de estadísticos descriptivos de manera sencilla. Por ejemplo, la librería `psych` contiene las funciones `describe()` y `describe.by()` que no sólo calculan la media sino que añaden otros estadísticos de interés.



## Ejemplo 1: Análisis de los datos.

6. Representar gráficamente: el nivel de estudios en un diagrama de sectores, el sexo en un diagrama de barras y la edad en un histograma. Construir también un boxplot de la edad en función del sexo.

Esta tarea puede llevarse a cabo de manera muy simple mediante los comandos que se muestran a continuación. Nótese que tanto para el diagrama de sectores (`pie()`) como para el diagrama de barras (`barplot()`) el argumento que se requiere es la tabla de frecuencias obtenida mediante `table()`:

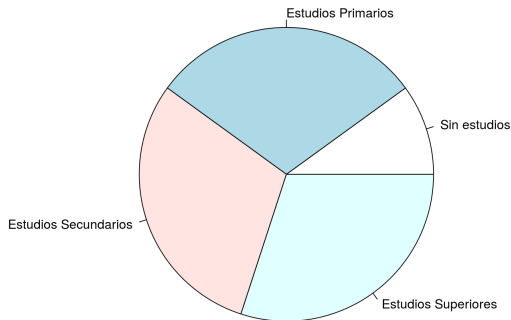
```
1 pie(table(estudios))
```





# Estadística Descriptiva con R (XXII)

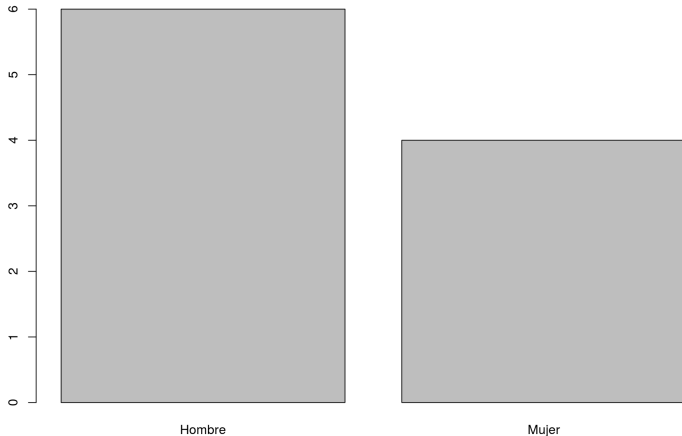
## Ejemplo 1: Análisis de los datos.



# Estadística Descriptiva con R (XXIII)

## Ejemplo 1: Análisis de los datos.

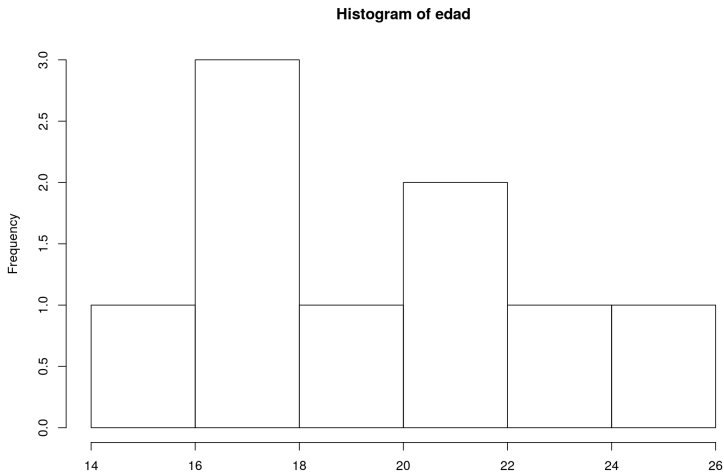
```
1 barplot(table(sexo))
```



# Estadística Descriptiva con R (XIV)

## Ejemplo 1: Análisis de los datos.

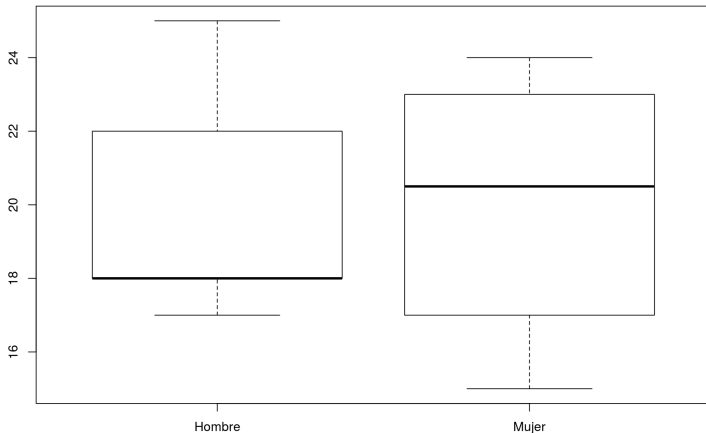
```
1 hist(edad)
```



# Estadística Descriptiva con R (XV)

## Ejemplo 1: Análisis de los datos.

```
1 boxplot(edad ~ sexo)
```



# Estadística Descriptiva con R (XIV)

## Ejemplo 1: Análisis de los datos.

### Observacion 2 (EJERCICIOS)

1. Ejecuta `help(pie)` y averigua cómo se puede poner un título al gráfico de sectores. Repite este gráfico, esta vez con el título **Distribución del nivel de estudios**.
2. Colorea los sectores de modo diferente: utiliza la opción `col=c("red", "blue", "green", "yellow")`.
3. Teclea `colors()` para ver una lista de los colores disponibles para los gráficos.
4. Dibuja un diagrama de barras para sexo con las barras horizontales y con un color distinto para cada sexo. Ayúdate de `help(barplot)`.
5. Cambia el título del histograma de edades, de forma que diga **Distribución por edades**. Explora la ayuda del comando `hist()` para que el histograma se construya en frecuencias relativas, y las barras se levanten sobre los intervalos (14,17), (17,20), (20,23) y (23,26).