

## PRÁCTICA PRIMERA CON R

### MANIPULACIÓN DE DATOS Y ANÁLISIS ESTADÍSTICO EXPLORATORIO CON R

#### Datos:

El archivo de datos proviene del paquete estadístico SPSS. Primero se guardó con formato texto, para recuperarlo desde R.

Nombre del archivo: datoEmpleado.DAT

Nota: En el archivo se usa como separador de columnas el tabulador

#### Manipulación de datos

- 1. Lectura de archivo
- 2. Visualización de los nombres de las variables
- 3. Manipulación de las variables sin referenciar el nombre del data.frame
- 4. Mover columnas a distintas posiciones
- 5. División vertical del data.frame
- 6. Ordenación de datos según los valores de una variable
- 7. Unión de 2 data.frames de distintas variables con información para los mismos casos.
- 8. Selección de parte de los casos de un data frame
  - 8.1. Con operadores lógicos
  - 8.2. Seleccionando filas
- 9. Recodificación y etiquetación de variables. Transformación en factor

#### Análisis estadístico exploratorio

##### 1. Variables cualitativas o factores

- 1.1. Representación gráfica
  - 1.1.1. Gráfico de Barras
  - 1.1.2. Gráfico de sectores
- 1.2. Tablas de frecuencias: absolutas y relativas

##### 2. Variables Cuantitativas

- 2.1. Resumen numérico: tendencia central, dispersión, cuantiles
- 2.2. Gráficos
  - 2.2.1. Gráfico de tallo y hojas
  - 2.2.2. Histograma. Histograma con curva normal
- 2.3. Agrupación en intervalos: Tablas de frecuencias

##### 3. Cuantitativas condicionadas a los valores de los factores: Análisis comparativo de gráficos y de resúmenes numéricos.

- 3.1. Resumen numérico estadístico por subgrupos de un factor

**3.2. Resumen numérico estadístico para subgrupos derivados de combinaciones factores**

**3.3. Representación gráfica conjunta para los subgrupos derivados de los factores**

#### **4. Análisis de tablas de 2 o más dimensiones**

**4.1. Cruce de 2 o más factores**

**4.1.1. Genera una tabla 3-dimensional con la función ftable**

**4.1.2. Tabla marginal 2 dimensiones con ftable**

**4.1.3. Otras organizaciones de la tabla 3-dimensiones con ftable**

**4.1.4. Otros ejemplos con las funciones table y ftable**

**4.1.5. Proceso inverso: del data.frame a las tablas de frecuencias de varias dimensiones.**

**4.2. Cruce de factores y/o variables cuantitativas agrupadas en intervalos**

**4.3. Distribuciones de frecuencias condicionadas, marginales y conjuntas**

**4.3.1. Tabla de frecuencias relativas conjuntas**

**4.3.2. Adición marginales a la tabla**

**4.3.3. Condicionadas**

**4.4. Asociación/independencia de variables: test chi-cuadrado de Pearson**

**4.4.1. Residuos, valores esperados (bajo independencia) y observados**

**4.5. Gráficos de barras**

**4.5.1. Sin apilar**

**4.5.2. Apiladas**

# PRÁCTICA: MANIPULACIÓN DE DATOS Y ANÁLISIS ESTADÍSTICO EXPLORATORIO

## Datos:

### Práctica con archivo datoEmpleado.DAT

## Manipulación de datos

### 1. lectura de archivo

```
>a=read.table("datoEmpleado.DAT",header=T,sep="\t") #Lee archivo y crea un data.frame a
```

Nota: Compruebe que el archivo está en directorio de trabajo de la sesión de R

### 2. Visualice los nombres de las variables:

```
>names(a) #Visualiza los nombres de las variables
```

```
[1] "id"      "sexo"    "fechnac" "educ"    "catlab"  "salario"  
[7] "salini"  "tiempemp" "expprev" "minoría"
```

### 3. Manipulación de las variables sin referenciar el nombre del data.frame

```
> attach(a) #permite trabajar con los nombres de las variables sin hacer referencia al data.frame a
```

Nota: Para deshacer la orden anterior use **detach(a)**

### 4. Mover columnas a distintas posiciones

#Coloca todas las variables cuantitativas a la derecha en el data.frame

```
>b=a[c(1,2,3,5,10,8,4,6,7,9)] #crea una copia de a con las variables cuantitativas a la derecha  
>names(b)
```

```
[1] "id"      "sexo"    "fechnac" "catlab"  "minoría" "tiempemp"  
[7] "educ"    "salario"  "salini"   "expprev"
```

### 5. División vertical del data.frame

Genera dos data frames con todos los casos (todas las filas): uno formado por un subgrupo de columnas o variables cualitativas y otro con las cuantitativas; ambos incluyendo la variable **id** (identificador del empleado). La variable común que representa un identificador de la fila o caso tiene como objeto la posibilidad de poder unirlos, si se desea, sin que la información se desestructure porque haya sido modificado el orden de las filas o casos en los data frames.

```
>b1=b[c(1,2,3,4,5)] #variables cualitativas e identificador de trabajador
```

```
>b1
```

	id	sexo	fechnac	catlab	minoría
1	1	h	2/3/1952	3	0
2	2	h	5/23/1958	1	0
3	3	m	7/26/1929	1	0

```

...
> b2=b[c(1,6,7,8,9,10)] #variables cuantitativas e identificador de trabajador
> names(b2)

[1] "id"    "tiempemp" "educ"   "salario" "salini"  "expprev"

```

## **6. Ordenación de datos según los valores de una variable**

Ordena el data.frame **b1** por sexo

```

>b1_ord=b1[order(a$sexo),]
> b1_ord

```

	id	sexo	fechnac	catlab	minoría
1	1	h	2/3/1952	3	0
2	2	h	5/23/1958	1	0
5	5	h	2/9/1955	1	0
...					
469	469	m	6/1/1964	1	0
473	473	m	11/25/1937	1	0
474	474	m	11/5/1968	1	0

Ordena el data.frame a por sexo y selecciona las columnas 4,5 y 6

```

> a[order(a$sexo),4:6]

```

	educ	catlab	salario
1	15	3	57000
2	16	1	40200
5	15	1	45000
6	15	1	32100
7	15	1	36000
12	8	1	28350
...			
473	12	1	21450
474	12	1	29400

## **7. Unión de 2 data.frames de distintas variables con información para los mismos casos.**

Une de nuevo en un solo data.frame (bnuevo) los data.frames b1\_ord y b2. Observa que b1\_ord ha cambiado el orden de las filas, por eso es necesario una variable a modo de clave que permita asociar a cada empleado sus valores en cada una de las variables de los 2 data.frames.

```

>merge(b1_ord, b2, by.x="id", by.y="id") #junta las variables distintas de dos data.frames

```

Nota: **id** representa la variable que identifica la fila o caso. Puede tener distinto nombre en cada data frame, pero en ambos casos identifica el caso o fila con el mismo valor.

## **8. Selección de parte de los casos de un data frame**

### **8.1. Con operadores lógicos**

Forma un data.frame con sólo mujeres

```

> a_muj=a[a$sexo=="m",]

```

> a\_muj

```
id sexo fechnac educ catlab salario salini tiempemp expprev minoría
3 3 m 7/26/1929 12 1 21450 12000 98 381 0
4 4 m 4/15/1947 8 1 21900 13200 98 190 0
...
```

Forma un data.frame que incluya sólo hombres que son directores y minoría étnica

> a[(sexo=="h" & catlab==3 & minoría==1),]

```
341 341 h 1/20/1945 12 3 59400 33750 74 272 1
383 383 h 6/3/1961 17 3 78500 28740 70 67 1
430 430 h 4/29/1956 19 3 66250 34980 67 99 1
446 446 h 8/3/1958 16 3 100000 44100 66 128 1
```

Note que tiene el mismo efecto: a[(a\$sexo=="h" & a\$catlab==3 & a\$minoría==1),], dado que se ha usado previamente attach(a)

id sexo fechnac educ catlab salario salini tiempemp expprev minoría. Observe también que el carácter h va entre comillas, dado que la variable no es numérica.

## 8.2. Seleccionando filas

>a\_muj[1:10,] #Selecciona las 10 primeras filas de a\_muj

## 9. Recodificación y etiquetación de variables. Transformación en factores

>b=a # hace una copia de a

### Recodifica minoría: 0=no 1=sí

Primero declaramos minoría como un factor

>b\$minoría=factor(b\$minoría)

Los niveles 0 y 1 se denominarán no y sí, respectivamente

```
>levels(b$minoría)=c("no", "si")
> b
```

```
id sexo fechnac educ catlab salario salini tiempemp expprev minoría
1 1 h 2/3/1952 15 3 57000 27000 98 144 no
2 2 h 5/23/1958 16 1 40200 18750 98 36 no
3 3 m 7/26/1929 12 1 21450 12000 98 381 no
....
```

Nota: Una instrucción similar es: b\$minoría=factor(b\$minoría, labels=c("no", "si"))

### Recodifica categoría laboral

Modalidades: 1=administrativo; 2=seguridad; 3=director

```
>b$catlab=factor(b$catlab)
>levels(b$catlab)=c("administrativo","Seguridad","director")
> b[1:3,] muestra las 3 primeras filas del data.frame b
```

```

id sexo fechnac educ      catlab salario salini tiempemp expprev minoría
1 1   h 2/3/1952 15    director 57000 27000   98   144   no
2 2   h 5/23/1958 16    administrativo 40200 18750   98   36    no
3 3   m 7/26/1929 12    administrativo 21450 12000   98   381    no

```

## De modo equivalente

```

> b=a # nueva copia de a
> b[1:3,] #visualiza las 3 primeras filas

```

```

id sexo fechnac educ catlab salario salini tiempemp expprev minoría
1 1   h 2/3/1952 15    3 57000 27000   98   144   0
2 2   h 5/23/1958 16    1 40200 18750   98   36    0
3 3   m 7/26/1929 12    1 21450 12000   98   381    0

```

```

> b$catlab=factor(b$catlab,labels=c("Admi","Segu","Dire")) #Declara bcatlab como
# factor y define las nuevas etiquetas de los valores.
> b[1:3,] #Visualiza las 3 primeras filas

```

```

id sexo fechnac educ catlab salario salini tiempemp expprev minoría
1 1   h 2/3/1952 15    Dire 57000 27000   98   144   0
2 2   h 5/23/1958 16    Admi 40200 18750   98   36    0
3 3   m 7/26/1929 12    Admi 21450 12000   98   381    0

```

## Recodifica sexo:

```
> a[1:3,] #muestra las 3 primeras filas del data.frame a
```

```

id sexo fechnac educ catlab salario salini tiempemp expprev minoría
1 1   h 2/3/1952 15    3 57000 27000   98   144   0
2 2   h 5/23/1958 16    1 40200 18750   98   36    0
3 3   m 7/26/1929 12    1 21450 12000   98   381    0

```

Transforma en factor sexo con niveles hombre y mujer.

```
>b$sexo=factor(b$sexo, labels=c("hombre","mujer"))
```

```
> b[1:3,1:4] # Visualiza las 3 primeras filas y 4 primeras columnas
```

```

id sexo fechnac educ
1 1 hombre 2/3/1952 15
2 2 hombre 5/23/1958 16
3 3 mujer 7/26/1929 12

```

# Análisis estadístico

## 1. Cualitativas o factores

### 1.1. Representación gráfica: barras, sectores

#### 1.1.1. Gráfico de Barras categoría laboral

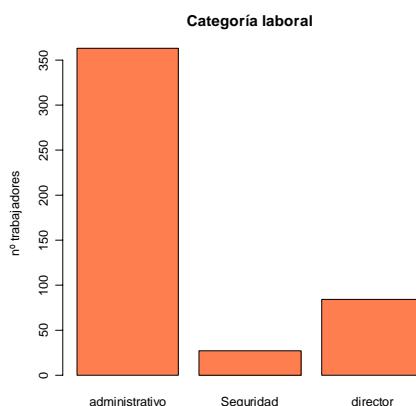
Datos:

```
> table(b$catlab)
```

administrativo	Seguridad	director
363	27	84

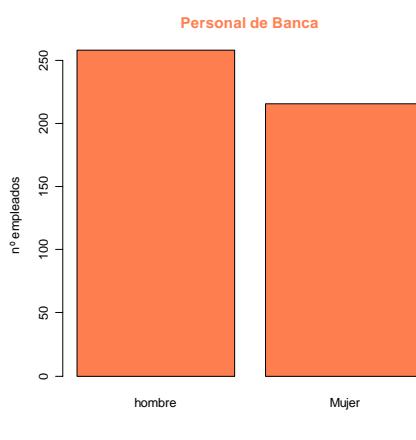
```
> barplot(table(b$catlab),col="coral")
```

```
> barplot(table(b$catlab),col="coral",legend.text=T)#añade leyenda con categorías  
> barplot(table(b$catlab),col="coral",main="Categoría laboral",ylab="nº trabajadores")
```



```
> barplot(table(b$sexo),col="coral",names.arg=c("HOMBRES","MUJERES"),ylab="nº empleados") #cambia los nombres por defecto h y m por hombre y mujer.
```

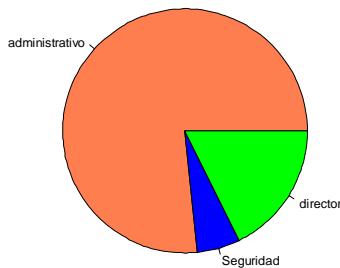
```
> title(main="Personal de Banca",col.main="coral",sub="sexo")
```



### 1.1.2. Gráfico sectores categoría laboral

```
> pie(table(b$catlab),col=c("coral","blue","green"))
> title(main="Personal de Banca",col.main="coral")
```

Personal de Banca



### 1.2. Tablas de frecuencias: absolutas y relativas

Datos:

```
> table(b$catlab)
```

administrativo	Seguridad	director
363	27	84

```
> table(b$minoría)
```

no	si
370	104

```
> table(b$sexo)
```

h	m
258	216

**En frecuencias relativas:**

```
>prop.table(table(b$sexo))
```

h	m
0.5443038	0.4556962

## 2. Variables Cuantitativas

### 2.1 .Resumen numérico: tendencia central, dispersión, cuantiles

```
> summary(data.frame(b$salario,b$salini,b$educ,b$tiempemp,b$expprev))
```

Resumen básico de todas las cuantitativas del data.frame

b.salario	b.salini	b.educ	b.tiempemp
Min. : 15750	Min. : 9000	Min. : 8.00	Min. : 63.00
1st Qu.: 24000	1st Qu.: 12488	1st Qu.: 12.00	1st Qu.: 72.00
Median : 28875	Median : 15000	Median : 12.00	Median : 81.00
Mean : 34420	Mean : 17016	Mean : 13.49	Mean : 81.11

```

3rd Qu.: 36938 3rd Qu.:17490 3rd Qu.:15.00 3rd Qu.:90.00
Max. :135000 Max. :79980 Max. :21.00 Max. :98.00
b.expprev
Min. : 0.00
1st Qu.: 19.25
Median : 55.00
Mean : 95.86
3rd Qu.:138.75
Max. :476.00

```

## 2.2. Gráficos

### 2.2.1. Gráfico de tallo y hojas del salario

```
> stem(salario)
```

The decimal point is 4 digit(s) to the right of the |

```

1 | 6666667777777778888999
2 | 00000000000001111111111111122222222222222222333333333+148
3 | 0000000000000000011111111111111111111112222222222333333333+36
4 | 0000000011222233445555667778999
5 | 01111233445555667778999
6 | 0001122355566777888999
7 | 00134455889
8 | 01346
9 | 1127
10 | 044
11 | 1
12 |
13 | 5

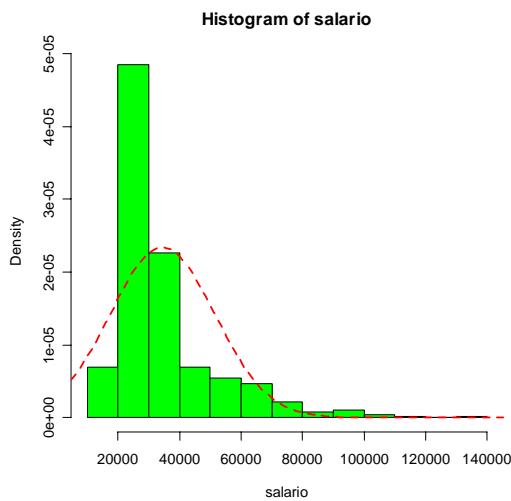
```

### 2.2.2. Histograma. Histograma con curva normal

```

>hist(salario, freq = FALSE,col="green")
>curve(dnorm(x,mean(salario), sd(salario)), col = 2, lty = 2, lwd = 2, add = TRUE)

```



### 2.2.3. Agrupación en intervalos: Tablas de frecuencias

cut divide la variable continua en intervalos.

El salario en 5 intervalos:

```
>cut(salario,breaks=5)
```

```

>table(cut(salario,breaks=5,labels=c(1:5)) ) #etiqueta los intervalos con 1 a 5
  1 2 3 4 5
368 67 30 8 1

> table(cut(salario,breaks=5,labels=FALSE) ) #etiqueta los intervalos con 1 a 5
  1 2 3 4 5
368 67 30 8 1

> table(cut(salario,breaks=5,dig.lab=6) ) #con 6 dígitos para las etiquetas de los extremos de intervalos
(15630.8,39528.4] (39528.4,63426.2] (63426.2,87323.9] (87323.9,111222]
 368           67           30           8
(111222,135119]
 1

```

### **En frecuencias relativas:**

```

>ta=table(cut(salario,breaks=5,labels=FALSE) )
>prop.table(ta)

  1      2      3      4      5
0.776371308 0.141350211 0.063291139 0.016877637 0.002109705

```

### **En frecuencias relativas con intervalos**

```

> prop.table(table(cut(salario,breaks=5) ))

(1.56e+04,3.95e+04] (3.95e+04,6.34e+04] (6.34e+04,8.73e+04] (8.73e+04,1.11e+05]
 0.776371308   0.141350211   0.063291139   0.016877637
(1.11e+05,1.35e+05]
 0.002109705

> prop.table(table(cut(salario,breaks=5,dig.lab=6) ))#Con 6 dígitos en las etiquetas de clases

(15630.8,39528.4] (39528.4,63426.2] (63426.2,87323.9] (87323.9,111222]
 0.776371308   0.141350211   0.063291139   0.016877637
(111222,135119]
 0.002109705

> min(salario)
[1] 15750

> max(salario)
[1] 135000

```

## **3.Cuantitativas condicionadas a los valores de los factores:Análisis comparativo de gráficos y de resúmenes numéricos.**

### **3.1. Resumen estadístico por subgrupos de sexo:**

```

> tapply(b$salario,b$sexo,var) #Calcula varianza de salario para grupo de hombres y mujeres

>tapply(b$salario,b$sexo,var)
  h      m
380219336 57123688

>tapply(b$salario,b$sexo,quantile) #calcula cuantiles de salario para h y m

```

```

$h
 0% 25% 50% 75% 100%
19650.0 28050.0 32850.0 50412.5 135000.0

$m
 0% 25% 50% 75% 100%
15750.0 21562.5 24300.0 28500.0 58125.0

> tapply(b$salario,b$minoría,quantile)
$no
 0% 25% 50% 75% 100%
15750.0 24150.0 29925.0 40312.5 135000.0

$si
 0% 25% 50% 75% 100%
16350.0 23662.5 26625.0 30637.5 100000.0

> tapply(b$salario,b$catlab,quantile)

$administrativo
 0% 25% 50% 75% 100%
15750 22800 26550 31200 80000

$Seguridad
 0% 25% 50% 75% 100%
24300 30150 30750 30975 35250

$director
 0% 25% 50% 75% 100%
34410.0 51956.25 60500.00 71281.25 135000.00

```

### **3.2.Resumen para grupos de combinaciones de sexo,minoría y categoría laboral**

```

> tapply(b$salario,list(b$catlab,b$sexo,b$minoría),mean) #calcula media para cada subgroupo
, , no

      h     m
administrativo 32671.64 25471.45
Seguridad    31178.57   NA
director      65683.57 47213.50

, , si

      h     m
administrativo 28952.13 23062.5
Seguridad    30680.77   NA
director      76037.50   NA

> tapply(b$salario,list(b$catlab,b$sexo,b$minoría),var)#Calcula varianza del salario para cada subgroupo
, , no

      h     m
administrativo 73599228 37117002
Seguridad    2751429   NA
director      325061103 72271295

, , si

      h     m
administrativo 32631735 15779712
Seguridad    6568558   NA
director      317622292  NA

> tapply(b$salario,list(b$catlab,b$sexo,b$minoría),median)

, , no

      h     m
administrativo 30825.0 24150.0
Seguridad    30750.0   NA
director      62187.5 45187.5

```

, , si

h	m
administrativo	27750 23775
Seguridad	30750 NA
director	72375 NA

### 3.3. Representación gráfica conjunta para los subgrupos derivados de los factores

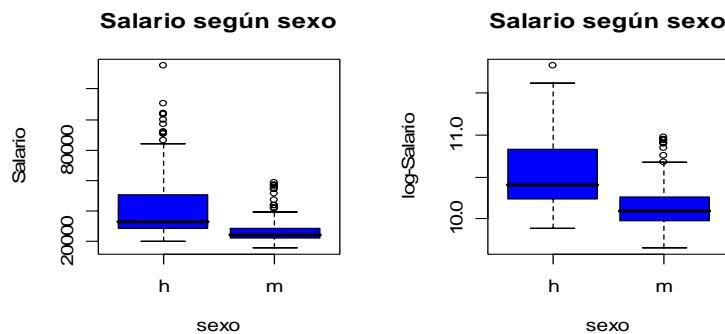
Gráfico caja

```
>par(mfrow=c(1,2),pty="s") #Divide el espacio gráfico en dos áreas: 1 filal y 2 columnas. Las áreas son cuadradas.
```

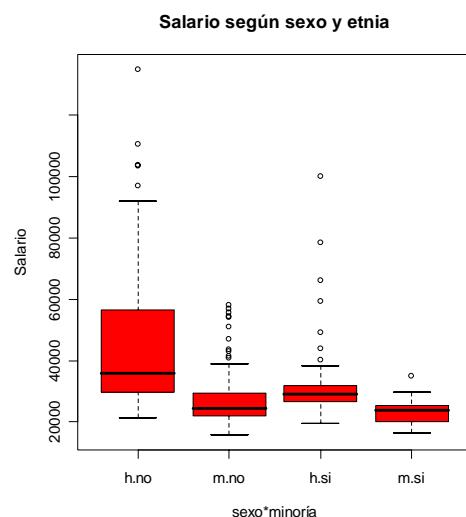
```
> boxplot(salario ~ sexo, data=b, col="blue",main="Salario según sexo",xlab="sexo",ylab="Salario")
```

```
> boxplot(log(salario) ~ sexo, data=b, col="blue",main="Salario según sexo",xlab="sexo",ylab="log-Salario")
```

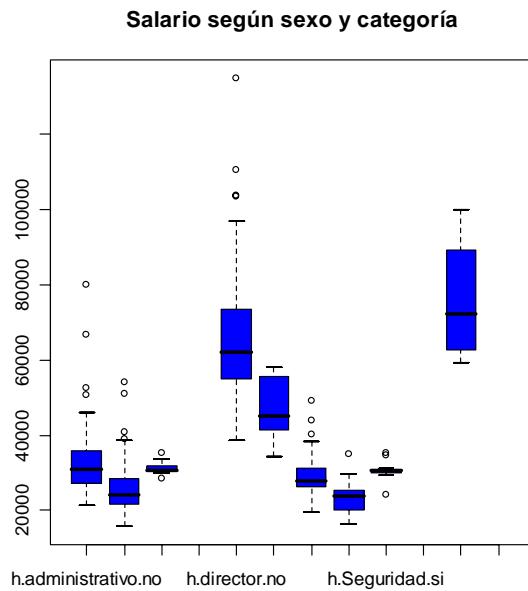
```
par(op)
```



```
>boxplot(salario ~ sexo+minoría, data=b, col="red",main="Salario según sexo y etnia",xlab="sexo*minoría",ylab="Salario")
```



```
>boxplot((salario) ~sexo+catlab+minoría, data=b, col="blue",main="Salario según sexo y categoría")
```



## 4.Análisis de tablas de 2 o más dimensiones

### 4.1.Cruce de 2 o más factores

```
>aa1=data.frame(sexo,catlab,minoría) #crea data.frame aa1 con 3 factores
```

#### 4.1.1.Genera una tabla 3-dimensional con la función ftable

```
> ftable(aa1,col.vars=2) #Incluye todas las variables del data.frame y usa modalidades de la variable 2 para estructurar  
#las frecuencias en columnas
```

	catlab	1	2	3
sexo	minoría			
1	0	110	14	70
	1	47	13	4
2	0	166	0	10
	1	40	0	0

```
> ftable(aa1,col.vars=1)
      sexo 1 2
catlab minoría
1   0    110 166
     1    47  40
2   0    14  0
     1    13  0
3   0    70 10
     1    4  0
```

#### 4.1.2.Tabla marginal 2 dimensiones con ftable

```
> ftable(aa1,col.vars=1,row.vars=2) #Colapsa las variables del data frame (marginal). Variable 2 en filas y
```

```
#modalidades de variable 1 en columnas.
  sexo 1 2
catlab
 1      157 206
 2      27 0
 3      74 10
```

#### **4.1.3. Otras organizaciones de la tabla 3-dimensiones con ftable**

```
> ftable(aa1,col.vars=1:3) #Todas las modalidades de las 3 variables en columnas para estructurar las frecuencias
sexo   1      2
catlab 1      2      3      1      2      3
minoría 0      1      0      1      0      1      0      1      0      1
                           110 47 14 13 70 4 166 40 0 0 10 0

> ftable(aa1,row.vars=1:3)
sexo catlab minoría
 1 1      0      110
          1      47
 2 0      14
          1      13
 3 0      70
          1      4
 2 1      0      166
          1      40
 2 0      0
          1      0
 3 0      10
          1      0
```

Con datos con modalidades de variables etiquetadas:

```
> ftable(b[c("sexo","catlab","minoría")])

               minoría no si
sexo catlab
h administrativo    110 47
  Seguridad        14 13
  director         70 4
m administrativo    166 40
  Seguridad        0 0
  director         10 0
```

Idem anterior, pero con modalidades de datos sin etiquetar:

```
> ftable(a[c("sexo","catlab","minoría")])

               minoría 0 1
sexo catlab
h 1      110 47
  2      14 13
  3      70 4
m 1      166 40
  2      0 0
  3      10 0
```

#### **4.1.4. Otros ejemplos con las funciones table, ftable**

```
> table(aa1)
> table(b[c("sexo","catlab","minoría")]) #equivalente anterior

, , minoría = no

  catlab
sexo administrativo Seguridad director
```

```

h      110    14    70
m      166     0    10

, , minoría = si

catlab
sexo administrativo Seguridad director
h      47     13     4
m      40     0     0

```

```
> ftable(b[c("sexo","catlab","minoría")])
```

		minoría	no	si
sexo	catlab			
hombre	administrativo	110	47	
	Seguridad	14	13	
	director	70	4	
mujer	administrativo	166	40	
	Seguridad	0	0	
	director	10	0	

```
> fa=ftable(b[c("sexo","catlab","minoría")])
```

		minoría	no	si
sexo	catlab			
hombre	administrativo	110	47	
	Seguridad	14	13	
	director	70	4	
mujer	administrativo	166	40	
	Seguridad	0	0	
	director	10	0	

```
> as.data.frame(fa)
```

	sexo	catlab	minoría	Freq
1	hombre	administrativo	no	110
2	mujer	administrativo	no	166
3	hombre	Seguridad	no	14
4	mujer	Seguridad	no	0
5	hombre	director	no	70
6	mujer	director	no	10
7	hombre	administrativo	si	47
8	mujer	administrativo	si	40
9	hombre	Seguridad	si	13
10	mujer	Seguridad	si	0
11	hombre	director	si	4
12	mujer	director	si	0

Abrir el data.frame con editor

```
>adf.fa=as.data.frame(fa)
>edit(adf.fa)
```

RGui

Data Editor

	sexo	catlab	minoría	Freq	var5	var6	var7
1	h	administrativo	no	110			
2	m	administrativo	no	166			
3	h	Seguridad	no	14			
4	m	Seguridad	no	0			
5	h	director	no	70			
6	m	director	no	10			
7	h	administrativo	si	47			
8	m	administrativo	si	40			
9	h	Seguridad	si	13			
10	m	Seguridad	si	0			
11	h	director	si	4			
12	m	director	si	0			
13							
14							
15							
16							
17							
18							
19							

> adf.fa

```
sexo      catlab minoría Freq
1 h administrativo no 110
2 m administrativo no 166
3 h Seguridad no 14
4 m Seguridad no 0
5 h director no 70
6 m director no 10
7 h administrativo si 47
8 m administrativo si 40
9 h Seguridad si 13
10 m Seguridad si 0
11 h director si 4
12 m director si 0
```

#### 4.1.5. Proceso inverso: del data.frame a las tablas de frecuencias de varias dimensiones

> xtabs(cbind(Freq) ~ . , data=adf.fa) #Transforma el data.frame anterior en tablas:  
, , minoría = no

```
catlab
sexo administrativo Seguridad director
h      110    14    70
m      166     0    10
```

, , minoría = si

```
catlab
sexo administrativo Seguridad director
h      47     13     4
m      40     0      0
```

> ftable(xtabs(cbind(Freq) ~ . , data=adf.fa))

```
minoría no si
sexo catlab
h administrativo 110 47
      Seguridad 14 13
      director 70 4
m administrativo 166 40
      Seguridad 0 0
      director 10 0
```

> xtabs(Freq ~ sexo + catlab, adf.fa)

```
catlab
sexo administrativo Seguridad director
```

h	157	27	74
m	206	0	10

```
>xtabs(Freq ~ sexo + catlab + minoría, adf.fa)
```

, , minoría = no

catlab			
sexo administrativo	Seguridad	director	
h	110	14	70
m	166	0	10

, , minoría = si

catlab			
sexo administrativo	Seguridad	director	
h	47	13	4
m	40	0	0

```
> ftable(b$sexo ~ b$catlab+b$minoría)
```

b\$catlab	b\$minoría	b\$sexo	h	m
administrativo	no		110	166
	si		47	40
Seguridad	no		14	0
	si		13	0
director	no		70	10
	si		4	0

```
> borra=ftable(b$sexo ~ b$catlab+b$minoría)
> as.data.frame(borra)
```

	b.catlab	b.minoría	b.sexo	Freq
1	administrativo	no	h	110
2	Seguridad	no	h	14
3	director	no	h	70
4	administrativo	si	h	47
5	Seguridad	si	h	13
6	director	si	h	4
7	administrativo	no	m	166
8	Seguridad	no	m	0
9	director	no	m	10
10	administrativo	si	m	40
11	Seguridad	si	m	0
12	director	si	m	0

#### 4.2. Cruce de factores y/o variables cuantitativas agrupadas en intervalos

##### Salario en 3 categorías y catlab (categoría laboral)

```
> table(cut(b$salario,breaks=3,dig.lab=6),b$catlab)
```

administrativo	Seguridad	director	
(15630.8,55460.3]	361	27	29
(55460.3,95289.8]	2	0	49
(95289.8,135119]	0	0	6

##### Salario en 3 categorías y salini (salario inicial) en dos.

```
> table(cut(b$salario,breaks=3,dig.lab=6),cut(b$salini,breaks=2,dig.lab=6))
```

	(8929.02,44490]	(44490,80051]
(15630.8,55460.3]	417	0
(55460.3,95289.8]	48	3
(95289.8,135119]	3	3

### 4.3. Distribuciones de frecuencias condicionadas, marginales y conjuntas

Datos: salario (categorías: altos y bajos) y sexo.

```
> summary(salario) #vemos un resumen de la variable cuantitativa
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15750	24000	28880	34420	36940	135000

Dividimos el salario en bajo=menor a 28880 y alto = mayor a 28880

```
> salar=salario<28880
> salar=factor(salar,labels=c("BAJO","ALTO"))
> tsal=table(salar,sexo) #crea objeto tabla en frecuencias absolutas
> tsal
```

sexos	
salar	h m
BAJO	185 52
ALTO	73 164

#### 4.3.1. Tabla de frecuencias relativas conjuntas

```
> prop.table(tsal,margin=NULL) #Tabla en frecuencias relativas conjuntas
```

sexos	
salar	h m
BAJO	0.3902954 0.1097046
ALTO	0.1540084 0.3459916

#### 4.3.2. Adición de marginales a la tabla:

```
> addmargins(table(salar,b$sexo)) #marginales en frecuencias absolutas
```

salar	h m Sum
BAJO	73 164 237
ALTO	185 52 237
Sum	258 216 474

```
> addmargins(prop.table(tsal,margin=NULL)) #marginales en frecuencias relativas
```

sexos	
salar	h m Sum
BAJO	0.3902954 0.1097046 0.5000000
ALTO	0.1540084 0.3459916 0.5000000
Sum	0.5443038 0.4556962 1.0000000

```
> addmargins(tsal) #equivalente con el objeto table previamente creado
```

sexos	
salar	h m Sum
BAJO	185 52 237
ALTO	73 164 237
Sum	258 216 474

#### 4.3.3. Condicionadas

```
> prop.table(tsal,margin=1) #Condicionadas de Sexo a Salar (Salario)
```

```

sexos
salar   h   m
BAJO 0.7805907 0.2194093
ALTO 0.3080169 0.6919831

> addmargins(prop.table(tsal,margin=1),2) #Añade la suma de frecuencias

sexos
salar   h   m   Sum
BAJO 0.7805907 0.2194093 1.0000000
ALTO 0.3080169 0.6919831 1.0000000
>

> prop.table(tsal,margin=2) #Condicionadas de salario a sexo

salar   h   m
BAJO 0.2829457 0.7592593
ALTO 0.7170543 0.2407407

```

```
> addmargins(prop.table(tsal,margin=2),1)
```

```

sexos
salar   h   m
BAJO 0.7170543 0.2407407
ALTO 0.2829457 0.7592593
Sum 1.0000000 1.0000000

```

#### **4.4. Asociación/independencia de variables: test chi-cuadrado de Pearson**

##### **Test de independencia Chi-cuadrado sexo, minoría**

```
> chisq.test(table(b$sexo, b$minoría))
```

```
Pearson's Chi-squared test with Yates' continuity correction

data: table(b$sexo, b$minoría)
X-squared = 2.3592, df = 1, p-value = 0.1245
```

Nota: Similar resultado se obtiene sin tabular previamente:

```
> chisq.test(b$sexo, b$minoría)
```

```
Pearson's Chi-squared test with Yates' continuity correction

data: b$sexo and b$minoría
X-squared = 2.3592, df = 1, p-value = 0.1245
```

##### **4.4.1. Residuos, valores esperados (bajo independencia) y observados**

```
> chisq.test(b$sexo, b$minoría)$residuals
```

```
b$minoría
b$sexo   no   si
h -0.5209118 0.9825358
m 0.5693079 -1.0738197
```

```
> chisq.test(b$sexo, b$minoría)$expected
```

```
b$minoría
b$sexo   no   si
h 201.3924 56.60759
m 168.6076 47.39241
```

```
> chisq.test(b$sexo, b$minoría)$observed
```

```

b$minoría
b$sexo no si
  h 194 64
  m 176 40

```

### Test Chi-cuadrado sexo, salario(bajos, altos)

Dividimos la variable salario en bajos y altos:

```
> summary(salario) #vemos un resumen de la variable cuantitativa
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
15750 24000 28880 34420 36940 135000
```

Dividimos el salario en bajo=menor a 28880 y alto = mayor a 28880

```
> chisq.test(salario > 28880, b$sexo)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: salario > 28880 and b$sexo
X-squared = 104.7975, df = 1, p-value < 2.2e-16
```

```
> chisq.test(salar, b$sexo) #equivalente con la variable salar
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: salar and b$sexo
X-squared = 104.7975, df = 1, p-value < 2.2e-16
```

```
> table(salar,b$sexo)
```

	h	m
BAJO	73	164
ALTO	185	52

```
> chisq.test(salar, b$sexo)$expected #Las frecuencias esperadas bajo independencia:
```

b\$sexo	salar	h	m
BAJO	129	108	
ALTO	129	108	

### Nota: El test Chi-cuadrado sin corrección es igual a

```
> summary(table(salar,b$sexo))
```

```
Number of cases in table: 474
Number of factors: 2
Test for independence of all factors:
  Chisq = 106.69, df = 1, p-value = 5.194e-25
```

Ambos permiten rechazar la hipótesis de independencia de salario y sexo a un nivel de significación bajísimo.

### Test chi-cuadrado categoría laboral y sexo

```
> chisq.test(b$catlab, b$sexo)
```

```
Pearson's Chi-squared test
```

```
data: b$catlab and b$sexo
X-squared = 79.2771, df = 2, p-value < 2.2e-16
```

```
> summary(table(b$catlab,b$sexo))
```

Number of cases in table: 474  
 Number of factors: 2  
 Test for independence of all factors:  
 Chisq = 79.28, df = 2, p-value = 6.098e-18

### Test Chi-cuadrado en tabla obtenida mediante agrupamiento del salario en clases según cuantiles y sexo

```
> quantile(salario) # Para agrupar el salario en clases o intervalos
```

	0%	25%	50%	75%	100%
h	15750.0	24000.0	28875.0	36937.5	135000.0

```
> table(cut(salario,quantile(salario)),b$sexo) # Tabla de contingencia sexo y Grupos salariales (definidos por los cuartiles)
```

	h	m
(1.58e+04,2.4e+04]	16	103
(2.4e+04,2.89e+04]	57	60
(2.89e+04,3.69e+04]	82	36
(3.69e+04,1.35e+05]	103	16

```
> summary(table(cut(salario,quantile(salario)),b$sexo)) #Resumen de Chi-cuadrado
```

Number of cases in table: 473  
 Number of factors: 2  
 Test for independence of all factors:  
 Chisq = 142.49, df = 3, p-value = 1.099e-30

### 4.5. Gráficos

#### Para Sexo y Categoría laboral

Datos:

```
> mt=table(sexo,b$catlab)
> mt
```

	sexu administrativo	Seguridad	director
h	157	27	74
m	206	0	10

```
> nmt=as.matrix(mt)

> dimnames(nmt)
$sexo
[1] "h" "m"

[[2]]
[1] "administrativo" "Seguridad"     "director"
```

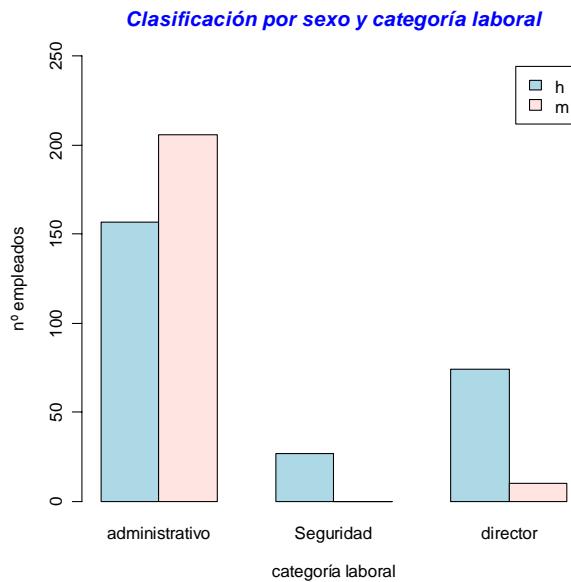
#### 4.5.1. Gráfico barras sin apilar

```
> barplot(nmt, beside = TRUE,
  col = c("lightblue", "mistyrose"),
  legend = rownames(nmt), ylim = c(0, 250), xlab = "categoría laboral", ylab = "nº empleados")

> title(main = list("Clasificación por sexo y categoría laboral", font = 4, col = "blue"))
```

Nota: una opción equivalente para título es:

```
>title(main = " total empleados por sexo y categoría", font.main = 2, col.main="red")
```



Otra forma de ver el gráfico:

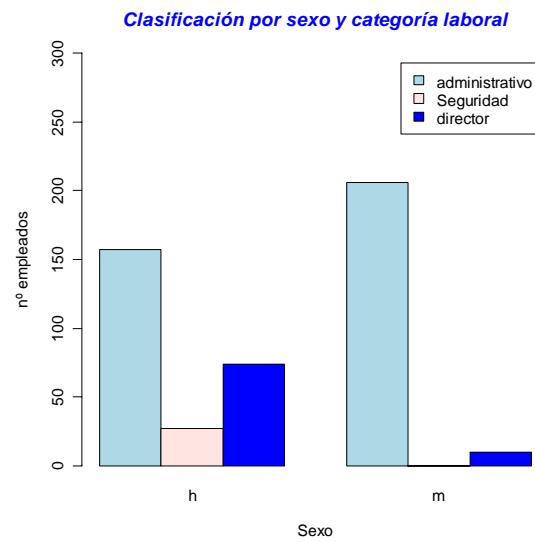
Datos:

```
> nmt=t(mt) #Transposición o cambio de filas por columnas
```

```
> nmt
```

sexo	h	m
administrativo	157	206
Seguridad	27	0
director	74	10

```
>barplot(nmt, beside = TRUE,  
         col = c("lightblue", "mistyrose","blue"),  
         legend = rownames(nmt), ylim = c(0, 300),xlab="Sexo",ylab="nº empleados")  
>title(main = list("Clasificación por sexo y categoría laboral", font = 4, col="blue"))
```



#### **4.5.2.Barras apiladas:**

```
>barplot(nmt, beside=FALSE,
  col = c("lightblue", "mistyrose", "blue"),
  legend = rownames(nmt), ylim = c(0, 300), xlab="Sexo", ylab="nº empleados")
> title(main = list("Clasificación por sexo y categoría laboral", font = 4, col="blue"))
```

Nota: Observe que se ha sustituido FALSE por TRUE en reside

