

Estadística Descriptiva con Datos Cualitativos

Ramon Ceballos

26/1/2021

EJEMPLO FINAL

Vamos a llevar a cabo un análisis completo de un ejemplo con lo que hemos aprendido en esta lección y aprovecharemos para aprender algo nuevo.

El objeto de datos **HairEyeColor** que lleva predefinido R es una tabla de frecuencias absolutas de tres variables cualitativas: color de cabello (**Hair**), color de los ojos (**Eye**) y sexo (**Sex**).

Vamos a extraer de esta tabla una *tabla bidimensional* de frecuencias absolutas de las variables **Eye** y **Hair**, sin distinguir según el sexo. La manera más sencilla de obtener esta tabla es sumando las subtablas de frecuencias para hombres y mujeres, y aplicando **as.table()** al resultado para transformarlo en una **table** por si no lo es.

Chunk para obtener una tabla que considere Eye y Hair únicamente

```
#Se obtienen las frecuencias relativas absolutas  
ftable(HairEyeColor)
```

```
##           Sex Male Female  
## Hair  Eye  
## Black Brown      32      36  
##       Blue      11       9  
##       Hazel      10       5  
##       Green       3       2  
## Brown Brown      53      66  
##       Blue      50      34  
##       Hazel      25      29  
##       Green      15      14  
## Red   Brown      10      16  
##       Blue      10       7  
##       Hazel       7       7  
##       Green       7       7  
## Blond Brown       3       4  
##       Blue      30      64  
##       Hazel       5       5  
##       Green       8       8
```

```
#Extraigo todos los colores de pelo y ojos para los hombres  
male <- HairEyeColor[, , "Male"]
```

```
#Extraigo todos los colores de pelo y ojos para las mujeres
```

```
female <- HairEyeColor[, "Female"]

#Transformo los datos extraidos en una tabla conjunta haciendo la suma
#Es como una matriz así que ocurre lo mismo que con las matrices
#Tabla bidimensional del color de pelo y ojos
data <- as.table(male+female)
data
```

```
##           Eye
## Hair      Brown Blue Hazel Green
## Black      68   20   15    5
## Brown     119   84   54   29
## Red        26   17   14   14
## Blond       7   94   10   16
```

```
#Es la misma tabla obtenida de otra manera
HEC=as.table(HairEyeColor[, , 1]+ HairEyeColor[, , 2])
```

Vamos a traducir al castellano los nombres de las variables de esta tabla y de sus niveles. Esto lo podemos llevar a cabo en un solo paso con la función `dimnames()` que ya usamos sobre data frames. El resultado de aplicar esta función a una table es una **list** cuyas componentes son los niveles de cada variable.

```
#Veo que nombres tiene los niveles
dimnames(HEC)
```

```
## $Hair
## [1] "Black" "Brown" "Red"   "Blond"
##
## $Eye
## [1] "Brown" "Blue"  "Hazel" "Green"
```

Ejercicio. Redefinid dicha **list** para tener los niveles de los factores en castellano

```
#Traduzco al castellano
dimnames(HEC)=list(Cabello=c("Negro", "Marron", "Rojo", "Rubio"), Ojos=c("Marrones", "Azules", "Pardos", "Verdes"),
HEC
```

```
##           Ojos
## Cabello  Marrones Azules Pardos Verdes
## Negro      68    20    15    5
## Marron     119    84    54   29
## Rojo       26    17    14   14
## Rubio       7    94    10   16
```

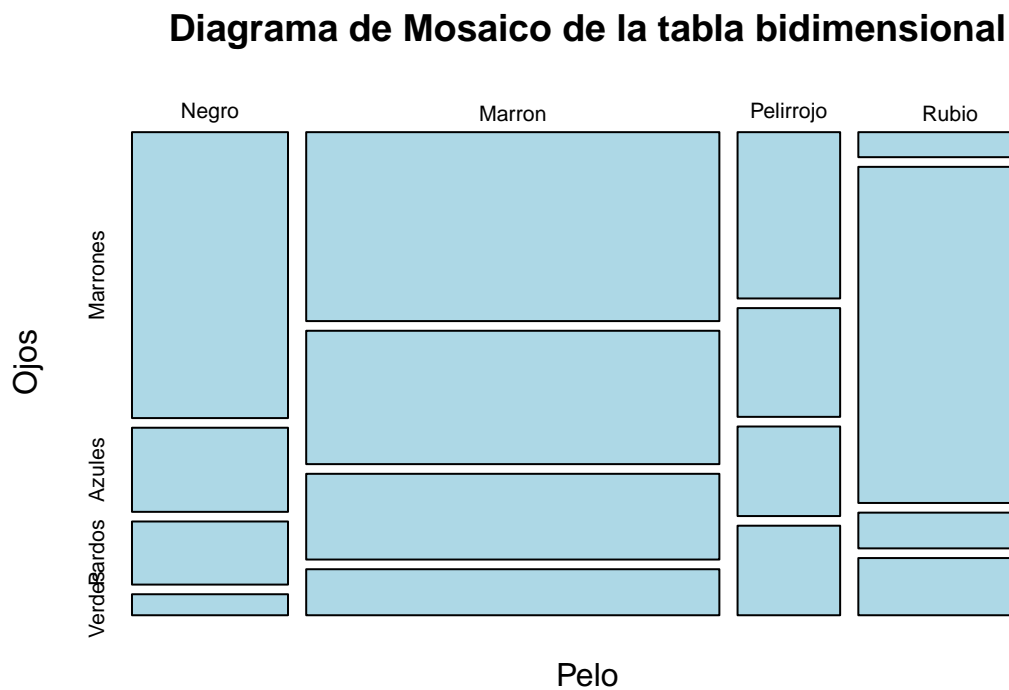
Chunk para manipulación de datos

```
#Para el otro ejercicio
dimnames(data)=list(
  Pelo = c("Negro", "Marron", "Pelirrojo", "Rubio"),
  Ojos= c("Marrones", "Azules", "Pardos", "Verdes")
)
data
```

```
##           Ojos
## Pelo      Marrones Azules Pardos Verdes
## Negro           68    20    15     5
## Marron          119    84    54    29
## Pelirrojo        26    17    14    14
## Rubio            7     94    10    16
```

Vamos a dibujar un *diagrama de mosaico* de esta tabla, para visualizar gráficamente sus entradas.

```
plot(data, color = ("lightblue"), main = "Diagrama de Mosaico de la tabla bidimensional")
```



A continuación, vamos a calcular el **número total de individuos** representados en esta tabla:

```
#Suma total de todos los individuos del DataFrame
sum (data)
```

```
## [1] 592
```

Las **tablas de frecuencias absolutas y relativas** de cada variable. El color de ojos va en columnas (2º variable) y el color de pelo va en filas (1º variable).

```
#Frecuencias absolutas para el color de ojos
colSums(data)
```

```
## Marrones  Azules  Pardos  Verdes
##      220     215     93     64
```

```
#Frecuencias absolutas para el color de pelo
rowSums (data)
```

```
##      Negro      Marron Pelirrojo      Rubio
##      108        286          71       127
```

```
#Frecuencias relativa para el color de ojos
#prop.table(colSums(data))
#redondeo a tres decimales
round(prop.table(colSums(data)),3)
```

```
## Marrones      Azules      Pardos      Verdes
##      0.372      0.363      0.157      0.108
```

```
#Frecuencias absolutas para el color de pelo
#prop.table(rowSums(data))
#redondeo a tres decimales
round(prop.table(rowSums(data)),3)
```

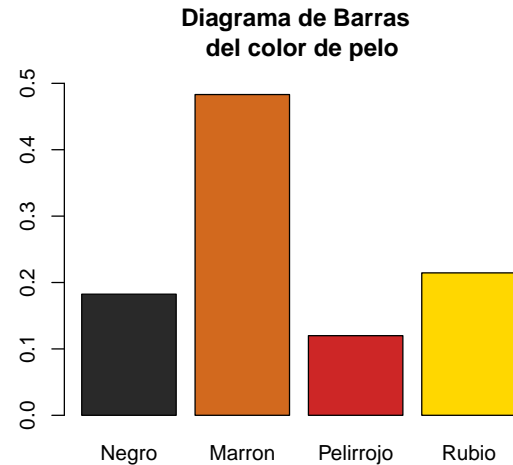
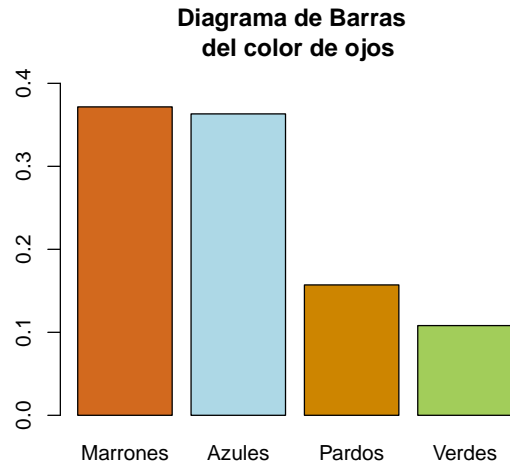
```
##      Negro      Marron Pelirrojo      Rubio
##      0.182      0.483      0.120      0.215
```

Representaremos estas últimas en sendos **diagramas de barras**.

```
#Establezco que habrá dos gráficos en la misma figura
par(mfrow=c(1,2))

#Represento el barplot del color de ojos (FREC. ABS.)
barplot (prop.table(colSums(data)),
        col = c("chocolate","lightblue","orange3","darkolivegreen3"),
        ylim=c(0,0.4),
        main = "Diagrama de Barras \n del color de ojos")

#Represento el barplot del color de pelo (FREC. ABS.)
barplot (prop.table(rowSums(data)),
        col = c("gray16","chocolate","firebrick3","gold1"),
        ylim=c(0,0.5),
        main = "Diagrama de Barras \n del color de pelo")
```



```
#cierro la figura
par(mfrow=c(1,1))
```

En el diagrama anterior vemos que el color dominante de cabellos es el castaño, mientras que en el color de ojos el marrón y el azul están prácticamente empatados.

Pasamos ahora a calcular las tablas de **frecuencias relativas** y dibujar los **dos diagramas de barras de las frecuencias relativas marginales**.

```
#calculo la tabla de FREC. RELAT. de todo el DF
round(prop.table(data),3)
```

```
##           Ojos
## Pelo      Marrones Azules Pardos Verdes
## Negro      0.115  0.034  0.025  0.008
## Marron     0.201  0.142  0.091  0.049
## Pelirrojo  0.044  0.029  0.024  0.024
## Rubio      0.012  0.159  0.017  0.027
```

```
#Frecuencia con la que un color de pelo dado, tiene el color de ojos
#calculo la tabla de FREC. RELAT. MARG. para las filas (MARGIN = 1)
round(prop.table(data, margin = 1),3)
```

```
##           Ojos
## Pelo      Marrones Azules Pardos Verdes
## Negro      0.630  0.185  0.139  0.046
## Marron     0.416  0.294  0.189  0.101
## Pelirrojo  0.366  0.239  0.197  0.197
## Rubio      0.055  0.740  0.079  0.126
```

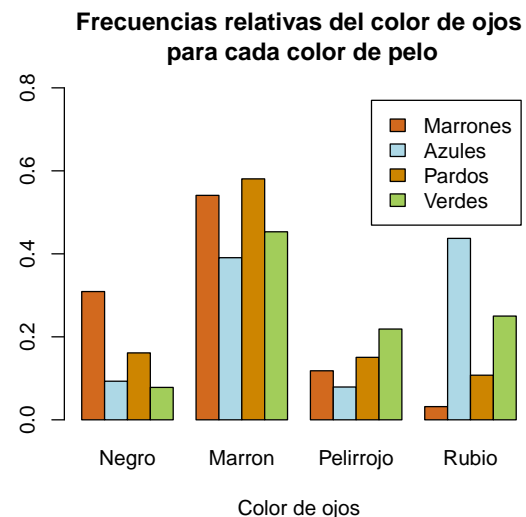
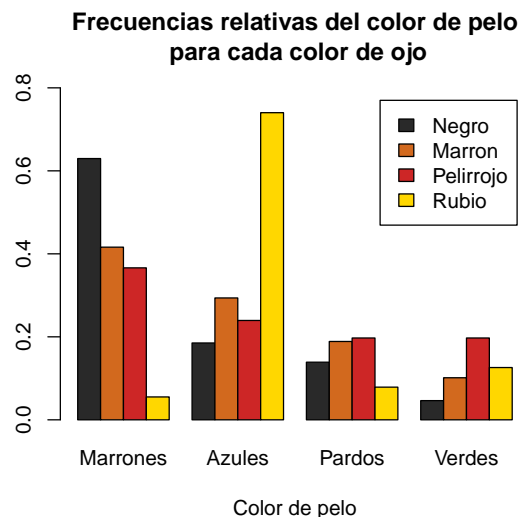
```
#Frecuencia con la que un color de ojos dado, tiene el color de pelo
#calculo la tabla de FREC. RELAT. MARG. para las columnas (MARGIN = 2)
round(prop.table(data, margin = 2),3)
```

```
##           Ojos
## Pelo      Marrones Azules Pardos Verdes
## Negro      0.309  0.093  0.161  0.078
## Marron     0.541  0.391  0.581  0.453
## Pelirrojo  0.118  0.079  0.151  0.219
## Rubio      0.032  0.437  0.108  0.250

#Establezco que habrá dos gráficos en la misma figura
par(mfrow=c(1,2))

#Represento el barplot (FREC. REL. MARG. FILAS)
barplot (prop.table(data, margin = 1),
        beside = TRUE,
        col = c("gray16", "chocolate", "firebrick3", "gold1"),
        ylim=c(0,0.8),
        main = "Frecuencias relativas del color de pelo\n para cada color de ojo",
        legend.text = TRUE,
        xlab = "Color de pelo")

#Represento el barplot (FREC. REL. MARG. COLUMNAS)
#!CUIDADO! Se representa la transpuesta ya que si no lo pintaría al reves (fila vs columna)
barplot (t(prop.table(data, margin = 2)),
        beside = TRUE,
        col = c("chocolate", "lightblue", "orange3", "darkolivegreen3"),
        ylim=c(0,0.8),
        main = "Frecuencias relativas del color de ojos\n para cada color de pelo",
        legend.text = TRUE,
        xlab = "Color de ojos")
```



```
#cierro la figura
par(mfrow=c(1,1))
```

Vemos que entre las personas de ojos azules, los cabellos rubios son los más frecuentes, y que entre las personas castañas el color de ojos más frecuente es el pardo.