

Datos cualitativos 7 - Ejemplo final

Moises

28/2/2020

Un 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.

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.

Ejercicio 1

Proceso para juntar el color de ojos y pelo sin tener en cuenta el sexo, osea llevarla de tridimensional a bidimensional

```
HEC = HairEyeColor #Guardamos HairEyeColor en una variable

dimnames(HEC) #Aqui se puede ver cada variable con sus niveles
```

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

```
male_table = HEC[,,"Male"] #Filtro solo los datos de los hombres y los guardo
female_table = HEC[,,"Female"] #Filtro solo los datos de las mujeres y los guardo
```

```
HEC_table = male_table+female_table #Junto las dos tablas y listo
```

```
HEC_table
```

```
##           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
```

Ejercicio 2.

Redefinid dicha list para tener los niveles de los factores en castellano

```
dimnames(HEC_table) #Con dimnames podremos generar una lista de HEC_table dimnames con sus factores y n
```

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

```
dimnames(HEC_table) = list(
  Pelo = c("Negro", "Marron", "Pelirrojo", "Rubio"),
  Ojos = c("Marrones", "Azules", "Cremas", "Verdes")
) #Por lo que podemos redefinir dicha lista con los nuevos valores
```

```
dimnames(HEC_table) # Y listo
```

```
## $Pelo
## [1] "Negro"      "Marron"      "Pelirrojo" "Rubio"
##
## $Ojos
## [1] "Marrones" "Azules"      "Cremas"     "Verdes"
```

Ejercicio 3

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

```
sum(HEC_table) #Numero total de individuos, esto se puede hacer con este dataframe porque es un table y
```

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

Ejercicio 4

Las tablas de frecuencias absolutas y relativas de cada variable,

```
colSums(HEC_table) #Sumando todas las columnas tenemos la cantidad de individuos por ojos
```

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

```
rowSums(HEC_table) #Sumando todas las filas tenemos la cantidad de individuos por pelo
```

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

```
prop.table(colSums(HEC_table)) #Para obtener la frecuencia relativa global
```

```
## Marrones      Azules      Cremas      Verdes  
## 0.3716216 0.3631757 0.1570946 0.1081081
```

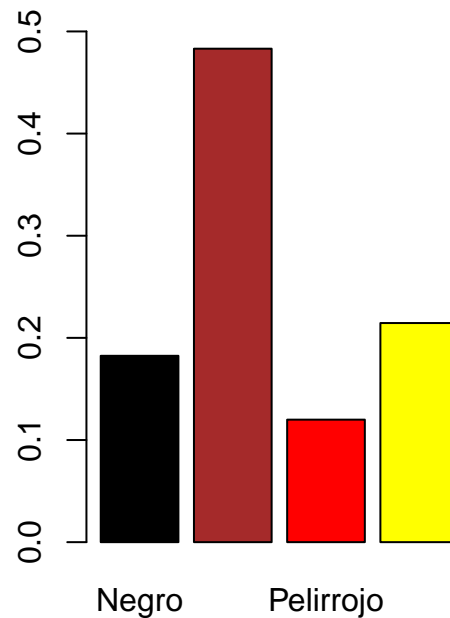
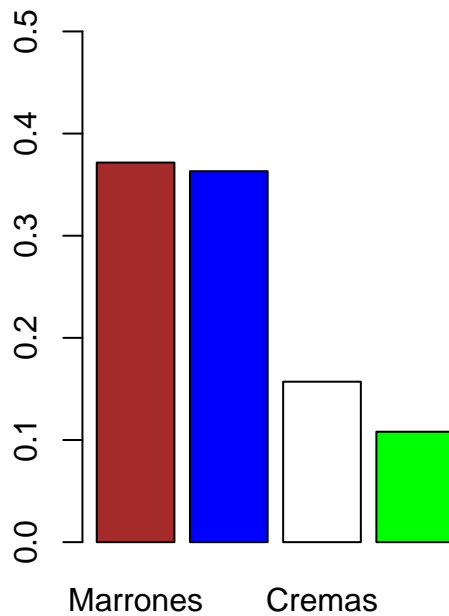
```
prop.table(rowSums(HEC_table)) #Para obtener la frecuencia relativa global
```

```
##      Negro      Marron Pelirrojo      Rubio  
## 0.1824324 0.4831081 0.1199324 0.2145270
```

Ejercicio 5

Representaremos estas últimas en sendos diagramas de barras.

```
par(mfrow=c(1,2)) #Para poder poner 2 o mas graficos uno al lado del otro  
barplot(prop.table(colSums(HEC_table)), col=c("brown","blue","white","green"), ylim=c(0,0.5))  
barplot(prop.table(rowSums(HEC_table)), col=c("black","brown","red","yellow"),ylim=c(0,0.5))
```



Ejercicio 6

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.

```
round(prop.table(HEC_table),3) # Frecuencia relativa global
```

```
##           Ojos
## Pelo      Marrones Azules Cremas 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
```

```
round(prop.table(HEC_table, margin=1),3) # Frecuencia relativa marginal por ojos
```

```
##           Ojos
## Pelo      Marrones Azules Cremas 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
```

```
round(prop.table(HEC_table, margin=2),3) # Frecuencia relativa marginal por pelo
```

```
##           Ojos
## Pelo      Marrones Azules Cremas 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
```

Ejercicio 7

Haz 2 graficos con las frecuencias relativas marginales con colores, titulo y leyendas

```
par(mfrow=c(1,2))
barplot(round(prop.table(HEC_table, margin=1),3), col=c("black","brown","red","yellow"),ylim=c(0,0.90),l
barplot(round(prop.table(HEC_table, margin=2),3), col=c("brown","blue","white","green"),ylim=c(0,0.90),l
```

relativas de colores de cabello en cs relativas de colores de ojos en ca

