

Estadística Descriptiva con Datos Cualitativos

Ramon Ceballos

27/1/2021

Estudio de bebés recién nacidos

Instalad y cargad el paquete **MASS**. Encontraréis una tabla de datos llamada **birthwt** sobre factores que pueden incidir en el peso de los niños al nacer.

PREGUNTAS

1. Con `str()` y `head()`, explorad la estructura, y con `help()`, mirad el significado de cada variable.

```
#carga la librería MASS
library(MASS)

#Exploro la tabla de datos
str(birthwt)
```

```
## 'data.frame':    189 obs. of  10 variables:
## $ low  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ age  : int  19 33 20 21 18 21 22 17 29 26 ...
## $ lwt  : int  182 155 105 108 107 124 118 103 123 113 ...
## $ race : int  2 3 1 1 1 3 1 3 1 1 ...
## $ smoke: int  0 0 1 1 1 0 0 0 1 1 ...
## $ ptl  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ ht   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ ui   : int  1 0 0 1 1 0 0 0 0 0 ...
## $ ftv  : int  0 3 1 2 0 0 1 1 1 0 ...
## $ bwt  : int  2523 2551 2557 2594 2600 2622 2637 2637 2663 2665 ...
```

```
#Se observan 10 variables numéricas (int)

#Observo las 4 primeras filas
head(birthwt,4)
```

```
##    low age lwt race smoke ptl ht ui ftv  bwt
## 85    0  19 182    2     0   0  0  1   0 2523
## 86    0  33 155    3     0   0  0  0   3 2551
## 87    0  20 105    1     1   0  0  0   1 2557
## 88    0  21 108    1     1   0  0  1   2 2594
```

```
tail(birthwt)
```

```
##      low age lwt race smoke ptl ht ui ftv  bwt
## 78   1  14 101   3     1   1  0  0   0 2466
## 79   1  28  95   1     1   0  0  0   2 2466
## 81   1  14 100   3     0   0  0  0   2 2495
## 82   1  23  94   3     1   0  0  0   0 2495
## 83   1  17 142   2     0   0  1  0   0 2495
## 84   1  21 130   1     1   0  1  0   3 2495
```

2. Calcula una tabla de frecuencias relativas marginales de los pares (raza de la madre, peso inferior a 2.5 kg o no) que permita ver si la raza de la madre influye en el peso del bebé.

```
#Guardo el data frame birthwt en el parámetro data
data = birthwt

#Frecuencias relativas de la raza para los recién nacidos de peso mayor o menor a 2.5 kg
Race_Low = prop.table(table(data$race, data$low),margin = 1)

#Cambio el nombre de las filas y columnas
colnames(Race_Low) = c("peso > 2.5 kgs", "peso < 2.5 kgs")
rownames(Race_Low) = c("white", "black", "other")

round(Race_Low,2)
```

```
##
##      peso > 2.5 kgs peso < 2.5 kgs
## white           0.76           0.24
## black           0.58           0.42
## other           0.63           0.37
```

#En la raza blanca el 76% de los recién nacidos pesan más de 2.5 kgs mientras que el 24% pesan menos de 2.5 kgs

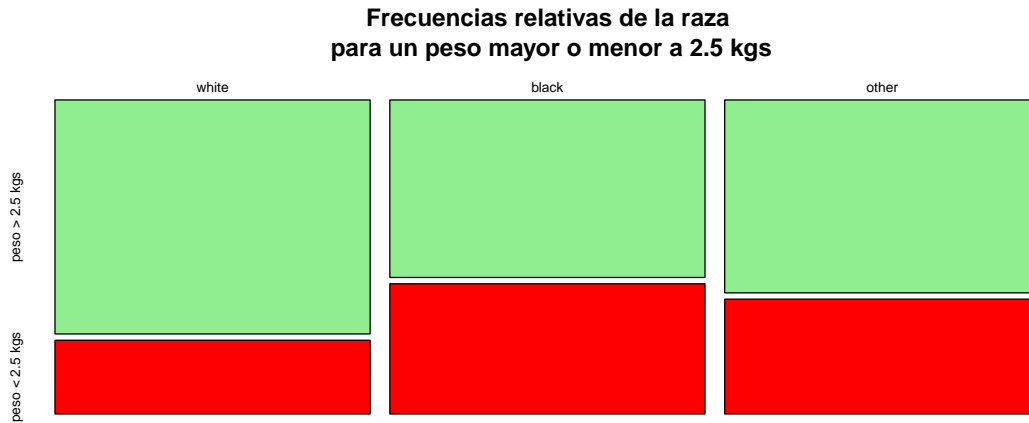
#En la raza negra el 58% de los recién nacidos pesan más de 2.5 kgs mientras que el 42% pesan menos de 2.5 kgs

#En otras razas el 63% de los recién nacidos pesan más de 2.5 kgs mientras que el 37% pesan menos de 2.5 kgs

#Debido a la variabilidad de porcentajes según la raza podemos concluir que la raza puede llegar a influir en el peso del bebé

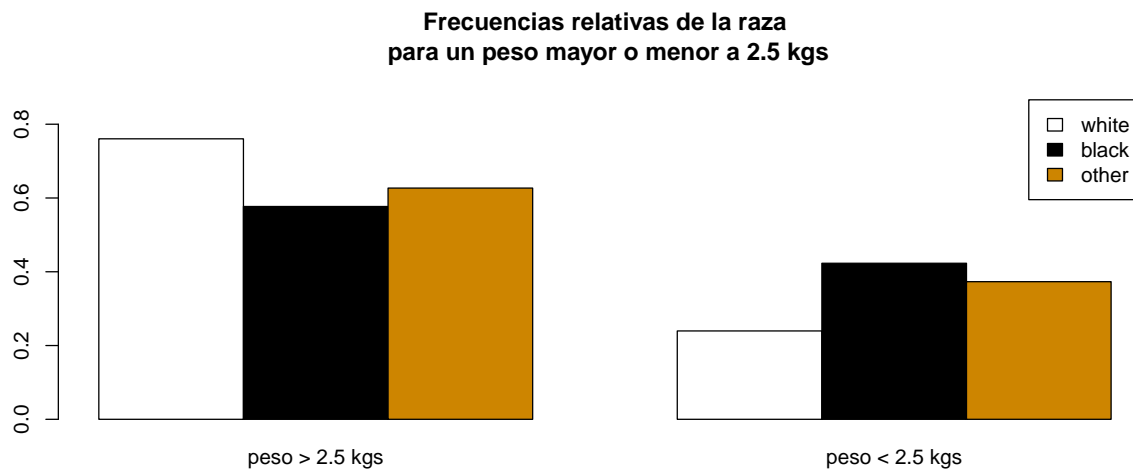
3. Dibuja un diagrama de mosaico de la tabla del ejercicio anterior.

```
plot(Race_Low,
     color =c("lightgreen" , "red"),
     main = "Frecuencias relativas de la raza\n para un peso mayor o menor a 2.5 kgs")
```



4. Dibuja un diagrama bidimensional de barras, con las barras organizadas en bloques, que permita visualizar la información de los ejercicios anteriores. Aprovecha para nombres adecuados a los bloques, colores a las barras, y añadir una leyenda que explique qué representa cada barra.

```
barplot(Race_Low,
        beside = TRUE,
        col = c("white", "black", "orange3"),
        ylim = c(0, 0.9),
        legend.text = TRUE,
        main = "Frecuencias relativas de la raza\n para un peso mayor o menor a 2.5 kgs")
```



5. ¿Se puede obtener alguna conclusión de la tabla y el diagrama de barras anterior? Argumenta tu respuesta

Si se pueden obtener conclusiones a partir de las gráficas de los ejercicios 3 y 4. Más del 50% de los recién nacidos pesan más de 2.5 kgs con independencia de la raza, no obstante, los recién nacidos de madres de raza blanca presentan una mayor probabilidad respecto a la raza negra y otras razas

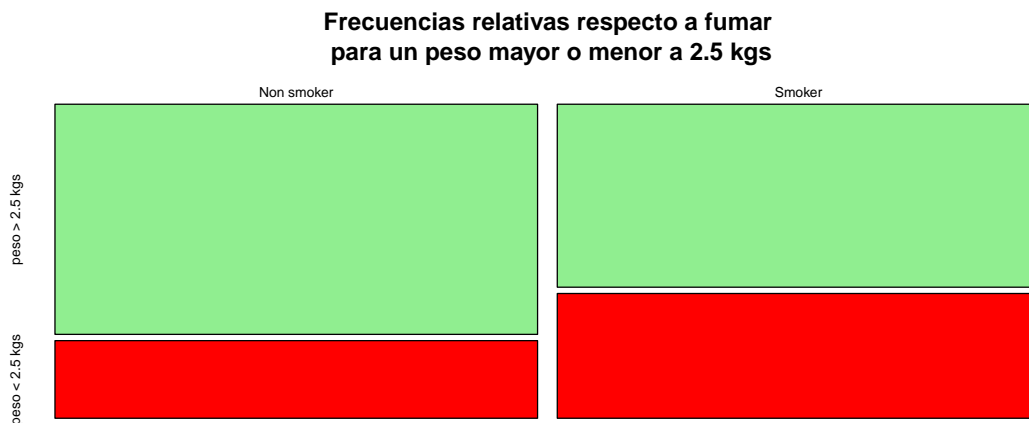
6. Repite los cuatro ejercicios anteriores para los pares (madre fumadora o no, peso inferior a 2.5 kg o no) y para los pares (madre hipertensa o no, peso inferior a 2.5 kg o no).

```
Smoke_Low = prop.table(table(data$smoke, data$low),margin = 1)
#Cambio el nombre de las filas y columnas
colnames(Smoke_Low) = c("peso > 2.5 kgs", "peso < 2.5 kgs")
#0 es no fumadora y 1 es fumadora
rownames(Smoke_Low) = c("Non smoker", "Smoker")

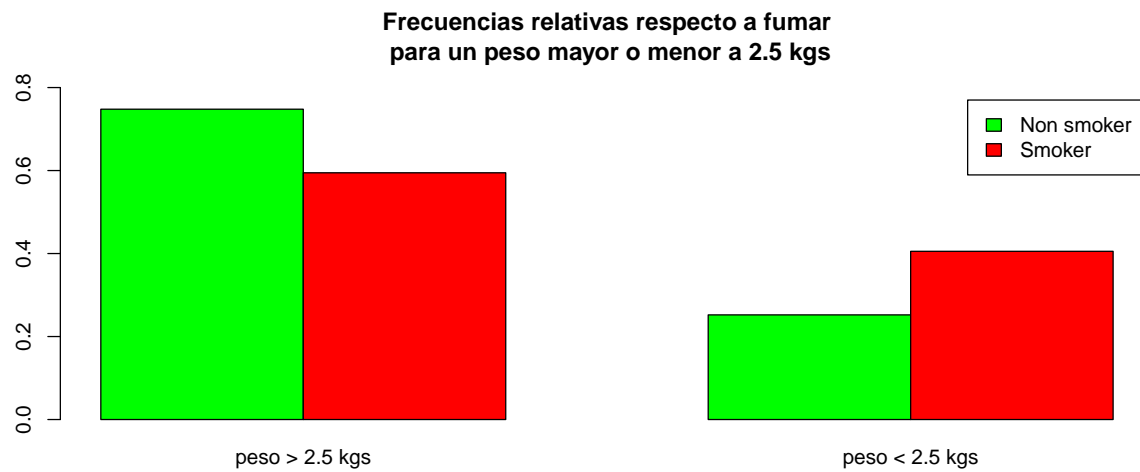
round(Smoke_Low,2)
```

```
##
##           peso > 2.5 kgs  peso < 2.5 kgs
## Non smoker           0.75           0.25
## Smoker                0.59           0.41
```

```
#Gráfica de mosaico
plot(Smoke_Low,
     color =c("lightgreen" , "red"),
     main = "Frecuencias relativas respecto a fumar\n para un peso mayor o menor a 2.5 kgs")
```



```
#Gráfica de barras
barplot(Smoke_Low,
       beside =TRUE,
       col =c("green","red"),
       ylim = c(0,0.8),
       legend.text = TRUE,
       main = "Frecuencias relativas respecto a fumar\n para un peso mayor o menor a 2.5 kgs")
```

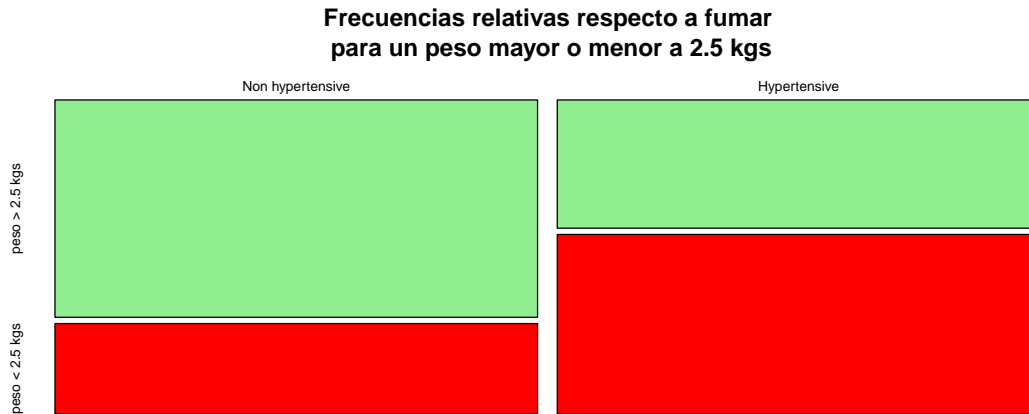


#Se observa que en las mujeres que fumaron durante el embarazo tienen una mayor probabilidad (41%) de q

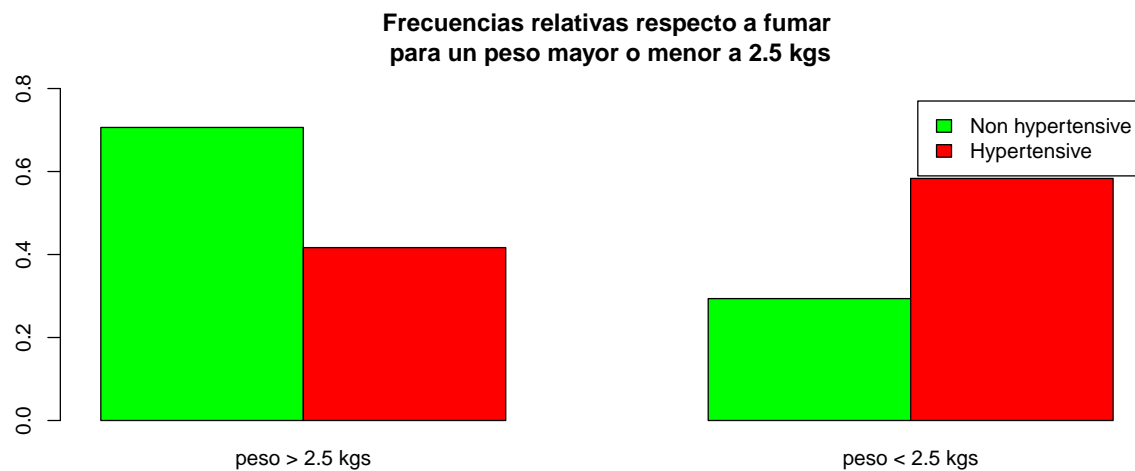
```
Ht_Low = prop.table(table(data$ht, data$low),margin = 1)
#Cambio el nombre de las filas y columnas
colnames(Ht_Low) = c("peso > 2.5 kgs", "peso < 2.5 kgs")
#0 es no fumadora y 1 es fumadora
rownames(Ht_Low) = c("Non hypertensive", "Hypertensive")
round(Ht_Low,2)
```

```
##
##           peso > 2.5 kgs peso < 2.5 kgs
## Non hypertensive      0.71      0.29
## Hypertensive          0.42      0.58
```

```
#Gráfica de mosaico
plot(Ht_Low,
     color =c("lightgreen" , "red"),
     main = "Frecuencias relativas respecto a fumar\n para un peso mayor o menor a 2.5 kgs")
```



```
#Gráfica de barras
barplot(Ht_Low,
  beside =TRUE,
  col =c("green","red"),
  ylim = c(0,0.8),
  legend.text = TRUE,
  main = "Frecuencias relativas respecto a fumar\n para un peso mayor o menor a 2.5 kgs")
```



#Se observa que en las mujeres hipertensas tienen una mayor probabilidad (58%) de que el recién nacido

7. Calcula una tabla de frecuencias relativas marginales de las ternas (raza de la madre, madre fumadora o no, peso inferior a 2.5 kg o no) que permita ver si la raza de la madre y su condición de fumadora o no fumadora influyen en el peso del bebé.

```

#Convierto en factores las tres variables implicadas en el estudio y le cambio el nombre a los niveles
#race
data$race =as.factor(data$race)
levels(data$race) = c("White","Black","Others")
#smoke
data$smoke =as.factor(data$smoke)
levels(data$smoke) = c("Non smoker","Smoker")
#low
data$low =as.factor(data$low)
levels(data$low) = c("peso > 2.5 kgs", "peso < 2.5 kgs")

#Guardo la tabla de frecuencia marginales de la raza y fumadores respecto al peso mayor o no de 2.5kgs
Race_Smoke_Low = prop.table(table(data$race,data$smoke,data$low), margin = c(1,2))

round(ftable(Race_Smoke_Low),2)

```

```

##                peso > 2.5 kgs peso < 2.5 kgs
##
## White  Non smoker          0.91          0.09
##         Smoker            0.63          0.37
## Black  Non smoker          0.69          0.31
##         Smoker            0.40          0.60
## Others Non smoker          0.64          0.36
##         Smoker            0.58          0.42

```

8. Dibuja un diagrama de mosaico de la tabla del ejercicio anterior.

```

plot(Race_Smoke_Low, col=c("green","red"),
     main = "Gráfico mosaico para las frecuencias relativas de la raza y fumar \n para el peso mayor o no de 2.5kgs")

```

Gráfico mosaico para las frecuencias relativas de la raza y fumar
para el peso mayor o menor de 2.5 kg

