

Capitulo 2 Introducción a la Estadística Descriptiva

Dereck Amesquita

6/4/2021

¿Que es la estadística descriptiva?

Es la rama de la estadística que nos permite resumir un gran conjunto de datos. “Deducir” un elemento según un gran conjunto de elementos similares. Con summary podremos obtener los principales datos descriptivos de un dataframe

```
prueba=iris
summary(prueba)
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.       :4.300    Min.       :2.000    Min.       :1.000    Min.       :0.100
## 1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
## Median :5.800    Median :3.000    Median :4.350    Median :1.300
## Mean   :5.843    Mean   :3.057    Mean   :3.758    Mean   :1.199
## 3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
## Max.   :7.900    Max.   :4.400    Max.   :6.900    Max.   :2.500
##      Species
## setosa      :50
## versicolor:50
## virginica   :50
##
##
##
```

Frecuencias

Frecuencia absoluta

Es el número de datos que hay.

Frecuencia relativa

Es el porcentaje de datos que hay ### Código en R Con sample obtener 12 elementos del 40 al 55 donde los elementos se pueden repetir. En Y estamos generando 12 elementos si y no, los cuales están medidos en un vector y se terminarán convirtiendo en un factor, debido a eso nos arroja que hay dos niveles. La función table nos ayuda a contar los elementos que existen.

```
x = sample(40:55, size = 12, replace=TRUE)
y = factor(sample(c("si", "no"), size=12, replace = TRUE))
x
```

```
## [1] 52 49 41 42 51 52 40 43 52 42 42 46
```

```
y
```

```
## [1] si no no si si si no si no no no no
## Levels: no si
```

Tambien podemos usar la funcion names que nos dara los niveles.

```
table(x)
```

```
## x
## 40 41 42 43 46 49 51 52
## 1 1 3 1 1 1 1 3
```

```
table(y)
```

```
## y
## no si
## 7 5
```

```
names(table(x))
```

```
## [1] "40" "41" "42" "43" "46" "49" "51" "52"
```

Table tiene una desventaja puesto que solo aparecen los valores mayores a cero. Si nuestro datos tienen un nivel el cual no tiene ningun elemento, no se nos sera mostrado. Por lo cual podemos convertirlos a factores. Creamos nx que recogera a x como factores con niveles del 40:55

```
nx=factor(x, levels=40:55)
table(nx)
```

```
## nx
## 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55
## 1 1 3 1 0 0 1 0 0 1 0 1 3 0 0 0
```

Si quisieramos encontrar un valor dentro de la tabla podemos indicarle la posicion o podemos hacer que busque. Solo bastara con usar:

```
table(nx)["54"]
```

```
## 54
## 0
```

Para conocer la frecuencia relativa deberemos usar prop.table. Nos da el porcentaje

```
prop.table(table(nx))
```

```
## nx
##      40      41      42      43      44      45      46
## 0.08333333 0.08333333 0.25000000 0.08333333 0.00000000 0.00000000 0.08333333
##      47      48      49      50      51      52      53
## 0.00000000 0.00000000 0.08333333 0.00000000 0.08333333 0.25000000 0.00000000
##      54      55
## 0.00000000 0.00000000
```

```
prop.table(table(y))
```

```
## y
##      no      si
## 0.5833333 0.4166667
```

```
#Si queremos obtener el porcentaje, debemos multiplicar por 100
prop.table(table(y))*100
```

```
## y
##      no      si
## 58.33333 41.66667
```

```
#Tambien podemos obtener la frecuencia relativa con la division
table(y)/length(y)
```

```
## y
##      no      si
## 0.5833333 0.4166667
```

```
table(x)==2
```

```
## x
##  40  41  42  43  46  49  51  52
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Frecuencias Bidimensionales

Construiremos un nuevo conjunto de datos. LO que haremos sera cruzar la informacion de Y. Con table haremos que cada “se” pueda vincularse con el valor en el orden correspondiente a “y”

```
se=sample(c("H","M"), replace = TRUE, size=length(y))
table(se,y)
```

```
##      y
## se no si
##  H  2  1
##  M  5  4
```

Podemos decir que 4 hombres dijeron que no, 4 mujeres dijeron que no.

Frecuencia relativa global

Se divide cada elemento entre el total. Ejemplo $4/12$ para el elemento (1,1) o $1/1$ para el elemento (2,2)

```
prop.table(table(se,y))
```

```
##      y
## se      no      si
##  H 0.1666667 0.08333333
##  M 0.4166667 0.3333333
```

El 33% del total son hombres que dijeron que no. el 25% son hombres que dijeron que si. ### Frecuencia relativa marginal

Con margin=1 obtenemos la FRM de las filas y con =2 obtenemo la FRM de las columnas

```
prop.table(table(se,y), margin=1)
```

```
##      y
## se      no      si
##  H 0.6666667 0.3333333
##  M 0.5555556 0.4444444
```

Analizaremos por filas, es decir del total de hombres el 0.66 respondió no, y el 0.33 respondió que si.

```
prop.table(table(se,y), margin=2)
```

```
##      y
## se      no      si
##  H 0.2857143 0.2000000
##  M 0.7142857 0.8000000
```

Analizaremos por columnas, es decir de los que dijeron si, el 75% fueron hombres y el 25% mujeres.