

Introducción al análisis de datos biológicos con R

Daniela Lamattina

Contenidos

1. Frecuencias
2. Histogramas

1. Frecuencias

- La frecuencia de un evento es el número de veces que dicho evento se repite en una muestra o experimento.
- Se pueden generar tablas de frecuencias para ayudar a organizar la información de una base de datos
- Generalmente, la distribución de las frecuencias de un conjunto de datos se visualiza mediante histogramas.

Tomemos un ejemplo

- En el marco de un estudio sobre patógenos transmitidos por garrapatas, realizamos la captura de coatíes (*Nasua nasua*) en el Parque Nacional Iguazú y los revisamos en procura de garrapatas. Aquí los datos en crudo:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30
<i>N° de garrapatas</i>	17	0	0	6	6	10	6	1	1	5	3	4	6	8	10	2	14	4	3	2	2	4	3	9	2	4	3	2	14	0



Tipos de frecuencias

- **Frecuencia absoluta:** es el número de veces que se repite una observación.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30
<i>Nº de garrapatas</i>	17	0	0	6	6	10	6	1	1	5	3	4	6	8	10	2	14	4	3	2	2	4	3	9	2	4	3	2	14	0

- R nos permite construir tablas de frecuencia de una vía usando la función `table`

Xi	fi
0	3
1	2
2	5
3	4
4	4
5	1
6	4
8	1
9	1
10	2
14	2
17	1

```
fa<-table(garrapas$ngarrapatas)
fa

##  0  1  2  3  4  5  6  8  9  10  14  17
##  3  2  5  4  4  1  4  1  1  2  2  1
```

Tipos de frecuencias

- **Frecuencia relativa:** es la proporción de la frecuencia absoluta de cada valor en relación al total. Es decir, es el número de veces que aparece un valor dividido por el total de observaciones.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30
N° de garrapatas	17	0	0	6	6	10	6	1	1	5	3	4	6	8	10	2	14	4	3	2	2	4	3	9	2	4	3	2	14	0

```
prop.table(fa)
```

```
# 0          1          2          3          4
# 0.10000000 0.06666667 0.16666667 0.13333333 0.13333333
# 5          6          8          9         10
# 0.03333333 0.13333333 0.03333333 0.03333333 0.06666667
# 14         17
# 0.06666667 0.03333333
```

→ si quiero expresarla en porcentaje uso
`prop.table()*100`

Xi	fa	fr
0	3	0.1
1	2	0.07
2	5	0.17
3	4	0.13
4	4	0.13
5	1	0.03
6	4	0.13
8	1	0.03
9	1	0.03
10	2	0.07
14	2	0.07
17	1	0.03

Tipos de frecuencias

- **Frecuencia absoluta acumulada:** es el total de frecuencias absolutas que corresponden a valores iguales o menores a cierto valor

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30
<i>N° de garrapatas</i>	17	0	0	6	6	10	6	1	1	5	3	4	6	8	10	2	14	4	3	2	2	4	3	9	2	4	3	2	14	0

Xi	fa	fr	Fa
0	3	0.1	3
1	2	0.07	5
2	5	0.17	10
3	4	0.13	14
4	4	0.13	18
5	1	0.03	19
6	4	0.13	23
8	1	0.03	24
9	1	0.03	25
10	2	0.07	27
14	2	0.07	29
17	1	0.03	30

Tipos de frecuencias

- **Frecuencia relativa acumulada:** es el total de frecuencias relativas que corresponden a valores iguales o menores a cierto valor.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30
<i>N° de garrapatas</i>	17	0	0	6	6	10	6	1	1	5	3	4	6	8	10	2	14	4	3	2	2	4	3	9	2	4	3	2	14	0

Xi	fa	fr	Fa	Fr
0	3	0.1	3	0.1
1	2	0.07	5	0.17
2	5	0.17	10	0.33
3	4	0.13	14	0.47
4	4	0.13	18	0.6
5	1	0.03	19	0.63
6	4	0.13	23	0.77
8	1	0.03	24	0.8
9	1	0.03	25	0.83
10	2	0.07	27	0.9
14	2	0.07	29	0.97
17	1	0.03	30	1

Distribución de frecuencias

- Es la especificación de las frecuencias correspondientes a cada uno de sus valores.

```
install.packages("fdth")  
library(fdth)  
freq.dist,table <- fdt(garrapas$ngarrapatas,start=0, end=18, h=1)  
freq.dist,table
```

Xi	fa	fr	Fa	Fr
0	3	0.1	3	0.1
1	2	0.07	5	0.17
2	5	0.17	10	0.33
3	4	0.13	14	0.47
4	4	0.13	18	0.6
5	1	0.03	19	0.63
6	4	0.13	23	0.77
8	1	0.03	24	0.8
9	1	0.03	25	0.83
10	2	0.07	27	0.9
14	2	0.07	29	0.97
17	1	0.03	30	1

```
Class limits f rf rf(%) cf cf(%)  
[0,1) 3 0.10 10.00 3 10.00  
[1,2) 2 0.07 6.67 5 16.67  
[2,3) 5 0.17 16.67 10 33.33  
[3,4) 4 0.13 13.33 14 46.67  
[4,5) 4 0.13 13.33 18 60.00  
[5,6) 1 0.03 3.33 19 63.33  
[6,7) 4 0.13 13.33 23 76.67  
[7,8) 0 0.00 0.00 23 76.67  
[8,9) 1 0.03 3.33 24 80.00  
[9,10) 1 0.03 3.33 25 83.33  
[10,11) 2 0.07 6.67 27 90.00  
[11,12) 0 0.00 0.00 27 90.00  
[12,13) 0 0.00 0.00 27 90.00  
[13,14) 0 0.00 0.00 27 90.00  
[14,15) 2 0.07 6.67 29 96.67  
[15,16) 0 0.00 0.00 29 96.67  
[16,17) 0 0.00 0.00 29 96.67  
[17,18) 1 0.03 3.33 30 100.00
```

2. Histogramas

- Los histogramas se usan para visualizar la distribución de frecuencias

```
hist(garrapas$ngarrapatas,xlab="Número de  
garrapatas",ylab="Frecuencia",col="dark violet",breaks=20,  
main="Conteo de garrapatas en Nasua nasua",border = "green")
```

