

Ejemplo 5: Estadística Descriptiva(Cangrejo *Xanthidae*)

Este ejemplo es el mismo que el ejemplo 4. La diferencia es que incluiremos en este fichero el código R. De forma que luego de procesarlo con la función

```
odfWeave('ejemplo5odfWeave.odt','ejemplo5odfWeave_OUT.odt')
```

Obtendremos como salida el resultado del ejemplo. Notemos que en la salida que está en la carpeta de ejemplos del lab2 se han modificado los anclajes de los dibujos después. Para modificarlos seleccionad el dibujo con el botón derecho del ratón y en el menú desplegable que aparecerá haced Anclaje → Como carácter.

Dado el gran número de huevos que produce en cada puesta el cangrejo *Xanthidae* esta variable resulta ser de interés. Cometeremos el abuso de considerar que la variable es continua. Las siguientes son las observaciones obtenidas para las puesta de 45 una hembra del cangrejo:

2802	2462	4000	3378	7343	4189	8973	4327
2412	7624	1548	4801	737	5321	849	5749
6837	8639	7417	6082	10241	962	3894	1801
5099	6627	4484	5633	4148	6588	5847	4632
6472	8372	8225	6142	12130			

Responder a las siguientes preguntas realizando un *script* de R y siguiendo las recomendaciones de la lección 2 de Estadística Descriptiva (página 3) del tema transversal. Poned comentarios a cada bloque de instrucciones que expliquen al lector qué es lo que hacen.

- Determinar la observación mayor y la más pequeña. Determinar el rango de las observaciones.
- Según la fórmula de *Sturges* ¿Cuántas clases, k , que se necesitan para subdividir estos datos?
- ¿Cuál la amplitud mínima necesaria para cubrir el intervalo, si se utilizan el número de clases del apartado anterior? No se os olvide corregir la amplitud a las unidades de medida.
- Determinar el límite inferior para la primera clase restando al valor mínimo la precisión. A partir de éste construir un vector numérico de nombre L_j que contenga los límites de las k clases de la forma $L_1, L_2, \dots, L_k, L_{k+1}$.
- Utilizar la instrucción `cut` para recodificar las observaciones por intervalos, guardar el resultado en la variable `conteo`. Calcular las frecuencias con la instrucción `cut` y guardarlas en la variable `frec.absolutas`. ¿Qué clase de objeto es `frec.absolutas`?, interpretarlo.
- Dibujad `hist(x,breaks=Lj,freq=TRUE)`, comentad el resultado.
- Dibujad el histograma de los datos con `hist(x,breaks=Lj,freq=FALSE)`, comentad el resultado.

h) Utilizando la función `par(mfrow=c(1,2))` R permite dibujar en un sólo gráfico una “fila de dos dibujos”. Ejecutad el código y comentar las diferencias entre los dos gráficos:

```
par(mfrow=c(1,2))
hist(x,breaks=Lj,freq=TRUE)
hist(x,breaks=Lj,freq=FALSE)
#volvemos a dejar 1 dibujo
par(mfrow=c(1,1))
```

Solución:

Introducimos los datos de la variable `numero` de huevos en la variable `num`

```
> num <- c(2802, 2462, 4000, 3378, 7343, 4189, 8973, 4327, 2412,
+ 7624, 1548, 4801, 737, 5321, 849, 5749, 6837, 8639, 7417,
+ 6082, 10241, 962, 3894, 1801, 5099, 6627, 4484, 5633, 4148,
+ 6588, 5847, 4632, 6472, 8372, 8225, 6142, 12130)
```

Apartado a) determinar le mínimo (`m`) y el Máximo (`M`) de `num`. Determinar el rango =`M-m`

```
> m = min(num)
> m
[1] 737
> M = max(num)
> M
[1] 12130
> rango = M - m
> rango
[1] 11393
```

Apartado b) Según la fórmula de Sturges ¿Cuántas clases, que se necesitan para subdividir estos datos?

```
> n = length(num)
> k = ceiling(1 + log(n, 2))
> k = nclass.Sturges(num)
> k
[1] 7
```

Apartado c) ¿Cuál la amplitud mínima necesaria para cubrir el intervalo, si se utilizan el número de clases del apartado anterior? No se os olvide corregir la amplitud a las unidades de medida.

```
> A = rango/k
> A
[1] 1627.571
> A = ceiling(A)
> A
[1] 1628
```

Apartado d) Determinar el límite inferior para la primera clase restando al valor mínimo la precisión. A partir de éste construir un array con el nombre `Lj` que contenga los límites de las clases `L1,L2,...,Lk,Lk+1`.

La unidad mínima de media es 1 la precisión es $\frac{1}{2}$

```
> precision = 1/2
> L1 = m - precision
> L1
[1] 736.5
> aux = c(0:k) * A
> aux
```

```
[1] 0 1628 3256 4884 6512 8140 9768 11396
```

```
> Lj = L1 + aux
```

```
> Lj
```

```
[1] 736.5 2364.5 3992.5 5620.5 7248.5 8876.5 10504.5 12132.5
```

Comprobad que el mínimo m está en el primer intervalo y el máximo M está en el último intervalo.

Efectivamente el mínimo y el máximo están incluidos en el primer y último intervalos respectivamente.

Apartado e) Utilizar la instrucción `cut` para contar las frecuencias absolutas de los datos en las clases obtenidas en el apartado anterior.

```
> conteo = cut(x, breaks = Lj)
```

```
> class(conteo)
```

```
[1] "factor"
```

```
> str(conteo)
```

```
Factor w/ 7 levels "(736,2.36e+03]",...: 2 2 3 2 5 3 6 3 2 5 ...
```

```
> frec.abs = table(conteo)
```

```
> frec.abs
```

```
conteo
```

```
(736,2.36e+03] (2.36e+03,3.99e+03] (3.99e+03,5.62e+03] (5.62e+03,7.25e+03]
```

```
5 5 9 9
```

```
(7.25e+03,8.88e+03] (8.88e+03,1.05e+04] (1.05e+04,1.21e+04]
```

```
6 2 1
```

```
> class(frec.abs)
```

```
[1] "table"
```

```
> names(frec.abs)
```

```
[1] "(736,2.36e+03]" "(2.36e+03,3.99e+03]" "(3.99e+03,5.62e+03]"
```

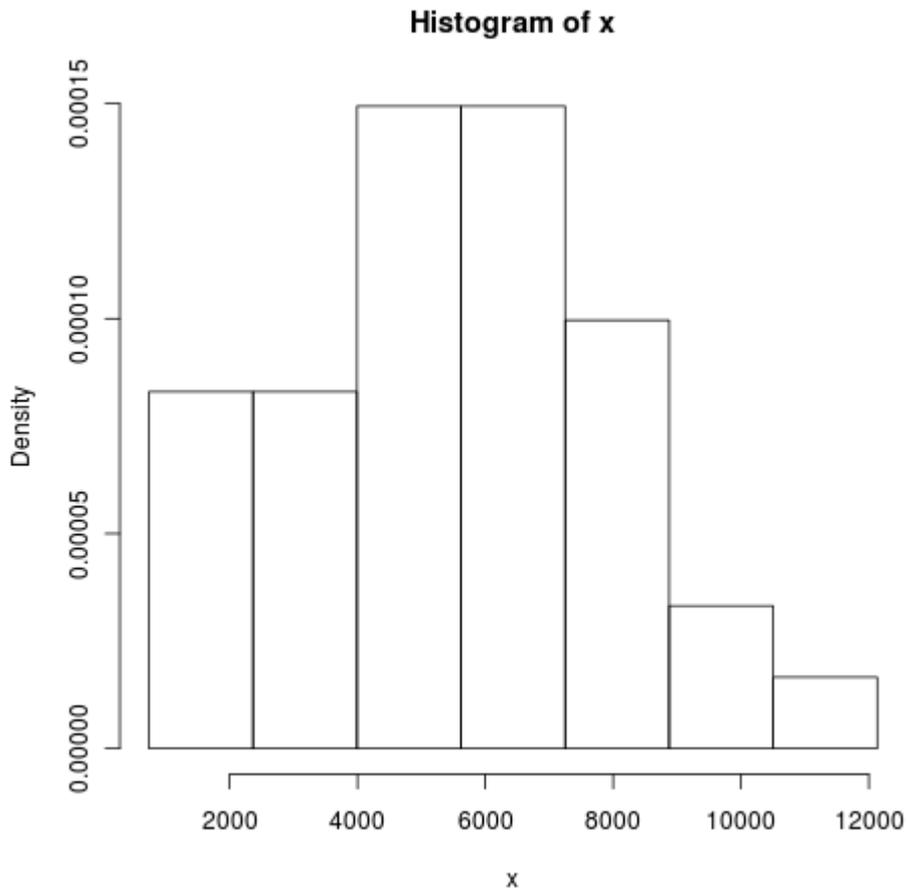
```
[4] "(5.62e+03,7.25e+03]" "(7.25e+03,8.88e+03]" "(8.88e+03,1.05e+04]"
```

```
[7] "(1.05e+04,1.21e+04]"
```

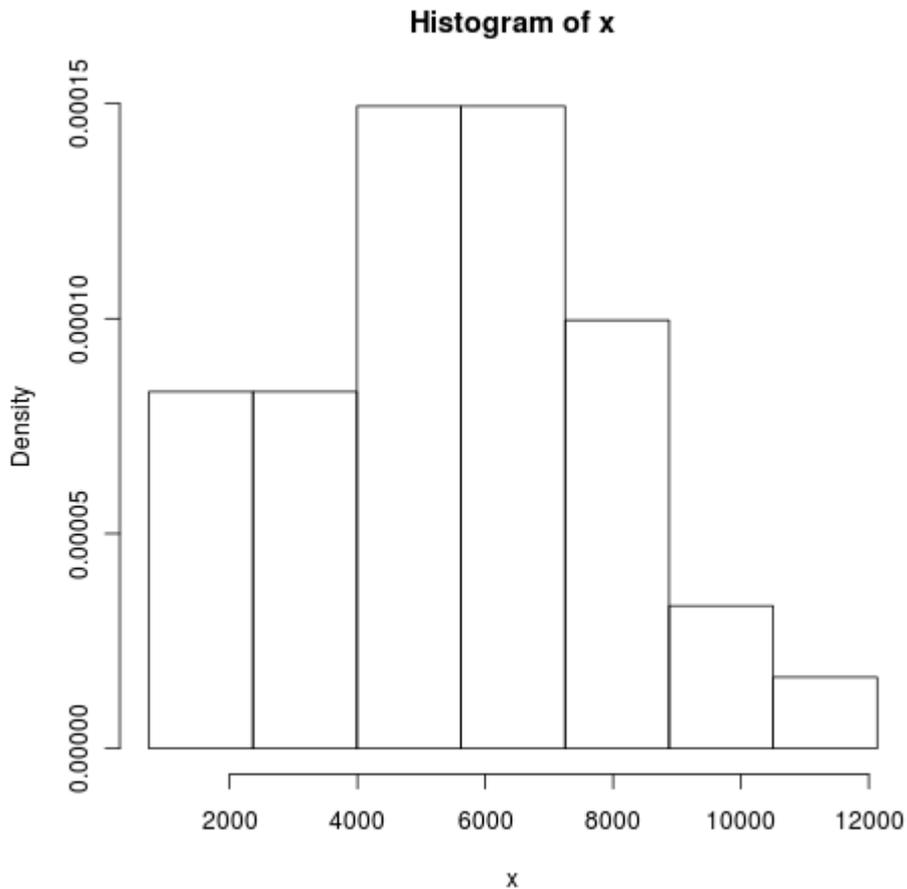
Como se ve con el objeto `conteo` es un `factor`, mientras que el objeto `frec.abs` es de la clase `table`.

Apartado f) Dibujad `hist(x, breaks=Lj, freq=TRUE)`, comentad el resultado.

```
> hist(x, breaks = Lj, freq = TRUE)
```



Es un falso histograma de frecuencias absolutas pues las alturas y no las áreas de los rectángulos son las frecuencias absolutas. Notemos que R denomina al eje vertical frecuencias



Apartado g) Lo mismo que el anterior apartado pero para frecuencias relativas.

```
> hist(x, breaks = Lj, freq = FALSE)
```

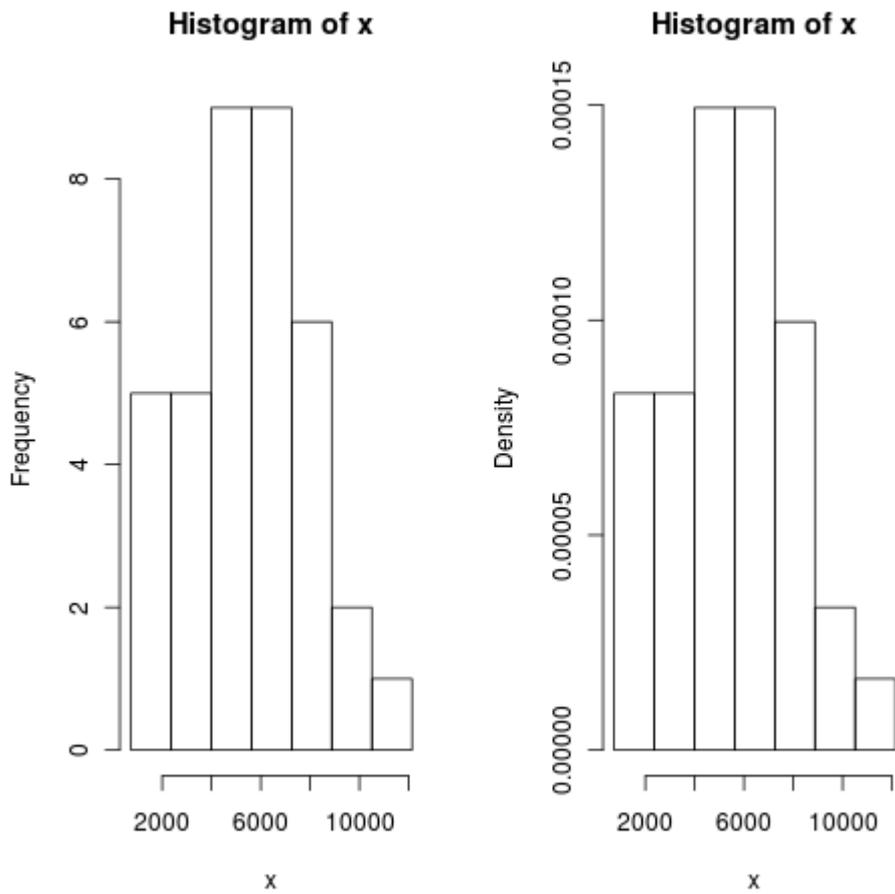
Es un histograma de frecuencias absolutas auténtico pues las áreas de los rectángulos son las frecuencias absolutas.

Notemos que R denomina al eje vertical densidades y no frecuencias. Las densidades son las alturas que hacen que el área de cada rectángulo sea la frecuencia absoluta.

Apartado h) Utilizando la función `par(mfrow=c(1,2))` de R permite dibujar en un sólo gráfico una “fila” de dos dibujos. Ejecutad el código y comentar las diferencias entre los dos gráficos:

```
par(mfrow=c(1,2))
hist(x,breaks=Lj,freq=TRUE)
hist(x,breaks=Lj,freq=FALSE)
#volvemos a dejar los parámetros igual que estaban.
par(mfrow=c(1,1))
```

```
> par(mfrow = c(1, 2))
> hist(x, breaks = Lj, freq = TRUE)
> hist(x, breaks = Lj, freq = FALSE)
> par(mfrow = c(1, 1))
```



Al poner uno al lado del otro los dos "histogramas", podemos ver las diferencias que se explicaron en los apartados anteriores. En este caso al ser los intervalos de igual amplitud no hay diferencia en el perfil de los rectángulos.