

Datos cuantitativos agrupados

Ramon Ceballos

4/2/2021

Ejemplo de Histogramas con dataset de datacrab

1. Datos de partida

Vamos a seguir trabajando con nuestra variable **cw** y, esta vez, lo que haremos será calcular histogramas de todas las formas explicadas anteriormente.

Recordemos todo lo que habíamos obtenido sobre nuestra variable **cw**:

```
#Variable cuantitativa de trabajo  
crabs = read.table("../../../data/datacrab.txt", header = TRUE)  
cw = crabs$width
```

```
#Amplitud de cada intervalo  
A = 1.3
```

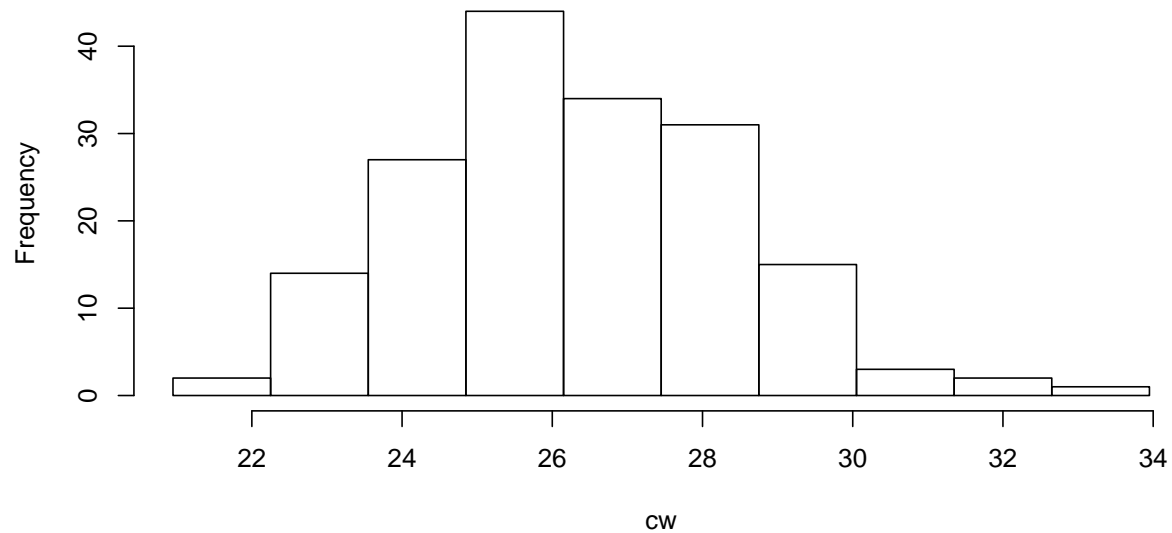
```
#Extremo 1 de los intervalos  
L1 = min(cw) - 1/2 * 0.1  
  
#Vector con todos los extremos de los intervalos  
L = L1 + A * (0:10)
```

2. Empleo de función hist() para dibujar el histograma

Dibujamos el histograma con **hist** y luego observamos su información interna.

```
hist(cw, breaks = L,  
     right = FALSE,  
     main = "Histograma de las anchuras de los cangrejos")
```

Histograma de las anchuras de los cangrejos



#Estructura interna del histograma

```
hist(cw, breaks = L,  
     right = FALSE, plot = FALSE)
```

```
## $breaks  
## [1] 20.95 22.25 23.55 24.85 26.15 27.45 28.75 30.05 31.35 32.65 33.95  
##  
## $counts  
## [1]  2 14 27 44 34 31 15  3  2  1  
##  
## $density  
## [1] 0.008892841 0.062249889 0.120053357 0.195642508 0.151178301 0.137839040  
## [7] 0.066696309 0.013339262 0.008892841 0.004446421  
##  
## $mids  
## [1] 21.6 22.9 24.2 25.5 26.8 28.1 29.4 30.7 32.0 33.3  
##  
## $xname  
## [1] "cw"  
##  
## $equidist  
## [1] TRUE  
##  
## attr("class")  
## [1] "histogram"
```

2.1 Histograma de frecuencias absolutas

Definimos la función **histAbs** en el presente informe de RMarkdown.

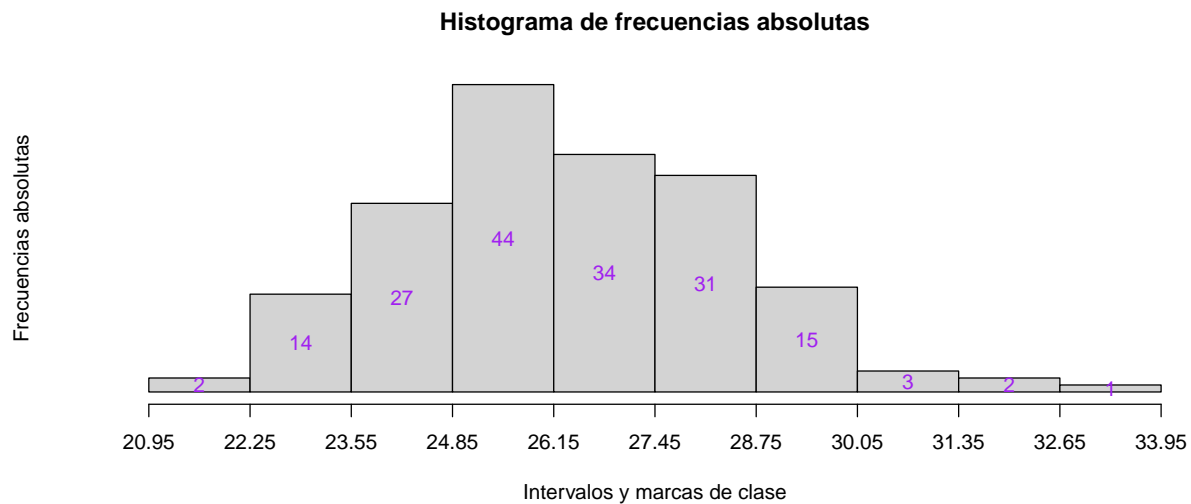
```
histAbs = function(x,L) {
  h = hist(x, breaks = L,
    right = FALSE,
    freq = FALSE,
    xaxt = "n",
    yaxt = "n",
    col = "lightgray",
    main = "Histograma de frecuencias absolutas",
    xlab = "Intervalos y marcas de clase",
    ylab = "Frecuencias absolutas")

  axis(1, at=L)

  text(h$mids, h$density/2, labels=h$counts, col="purple")
}
```

Ahora, dibujamos el histograma con **histAbs**.

```
histAbs(cw,L)
```



2.2 Histograma de frecuencias absolutas acumuladas

Definimos la función **histAbsCum** en el presente informe de RMarkdown.

```
histAbsCum = function(x,L) {

  h = hist(x, breaks = L,
    right = FALSE ,
    plot = FALSE)

  h$density = cumsum(h$density)
```

```

plot(h, freq = FALSE,
     xaxt = "n", yaxt = "n",
     col = "lightgray",
     main = "Histograma de frecuencias\nabsolutas acumuladas",
     xlab = "Intervalos",
     ylab = "Frec. absolutas acumuladas")

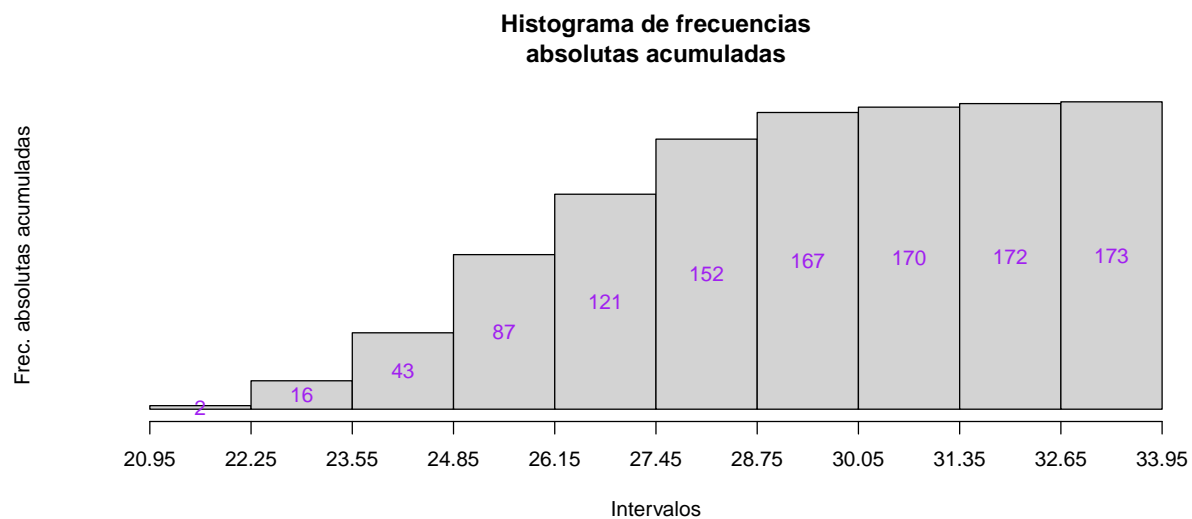
axis(1, at=L)

text(h$mids, h$density/2,
     labels = cumsum(h$counts),
     col = "purple")
}

```

Dibujamos el histograma con `histAbsCum`.

```
histAbsCum(cw,L)
```

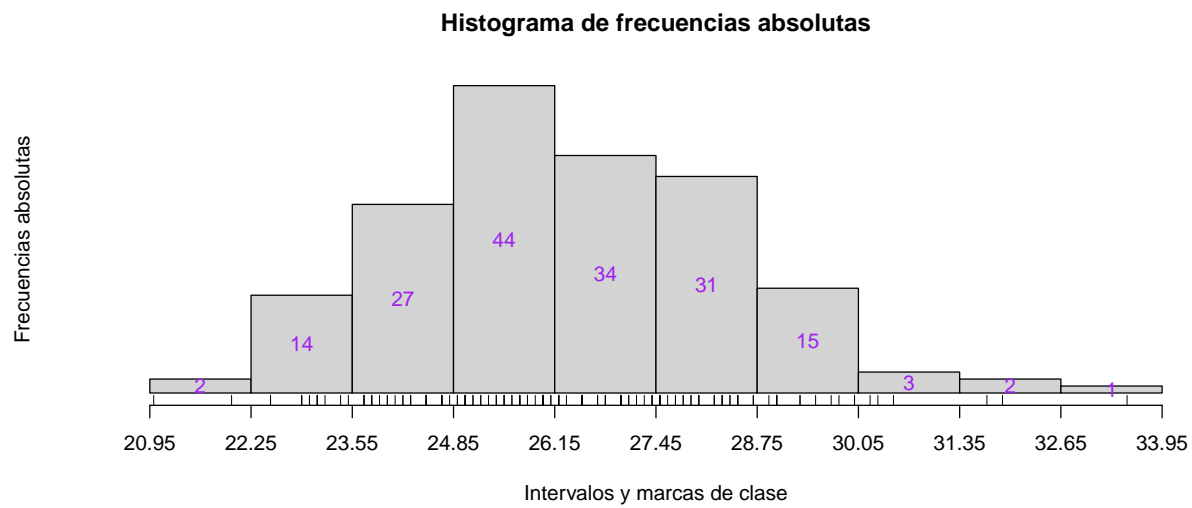


Hacemos uso de las funciones `rug` y `jitter`.

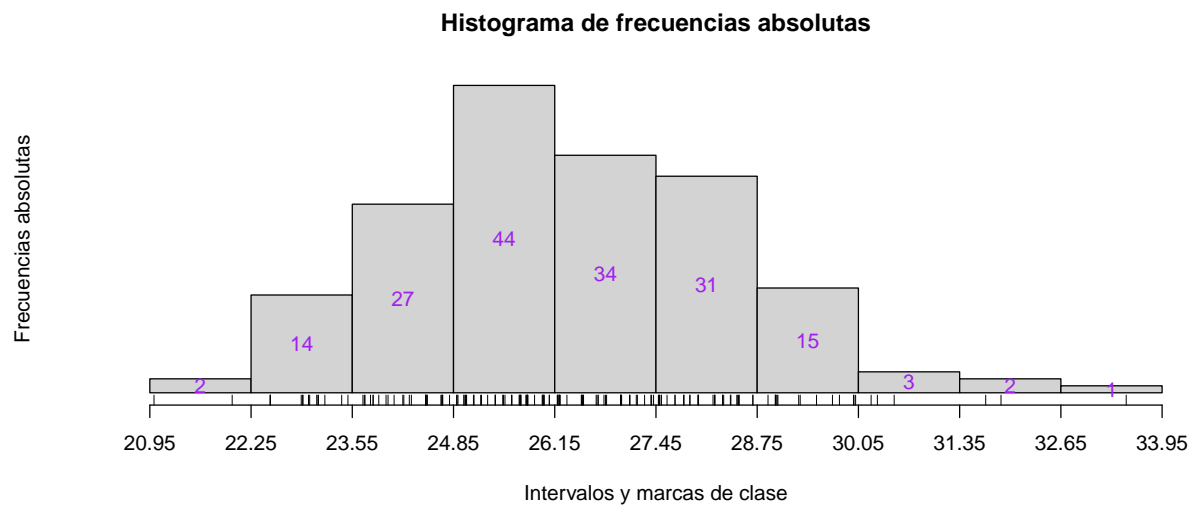
```

histAbs(cw,L)
rug(cw)

```



```
histAbs(cw,L)
rug(jitter(cw))
```



2.3 Histograma de frecuencias relativas

Definimos la funci3n **histRel** en el presente informe de RMarkdown.

```
histRel = function(x,L) {
  h = hist(x, breaks=L, right=FALSE , plot=FALSE)

  t = round(1.1*max(max(density(x)[[2]]),h$density),2)

  plot(h, freq = FALSE,
       col = "lightgray",
```

```

main = "Histograma de frec. relativas\ny curva de densidad estimada",
xaxt="n",
ylim=c(0,t),
xlab="Intervalos",
ylab="Densidades")

axis(1, at = L)

text(h$mids,
     h$density/2,
     labels = round(h$counts/length(x),2),
     col = "blue")

lines(density(x), col = "purple", lwd = 2)
}

```

A continuación, calculamos la **densidad de cw** y la representamos con **histRel**.

```

#densidad de "cw" (512 observaciones)
#Se obtienen 512 datos en el eje "x" y en el eje "y" que representan 512 ptos
#Estos ptos unidos generan la curva de densidad
str(density(cw))

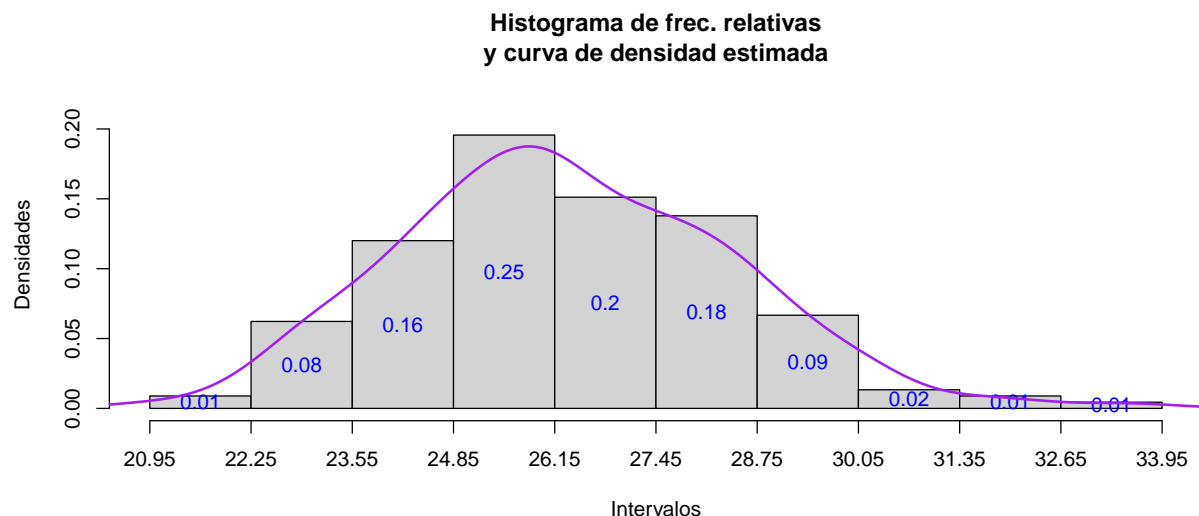
```

```

## List of 7
## $ x      : num [1:512] 19 19 19.1 19.1 19.1 ...
## $ y      : num [1:512] 3.90e-05 4.50e-05 5.17e-05 5.94e-05 6.82e-05 ...
## $ bw     : num 0.671
## $ n      : int 173
## $ call   : language density.default(x = cw)
## $ data.name: chr "cw"
## $ has.na  : logi FALSE
## - attr(*, "class")= chr "density"

```

```
histRel(cw,L)
```



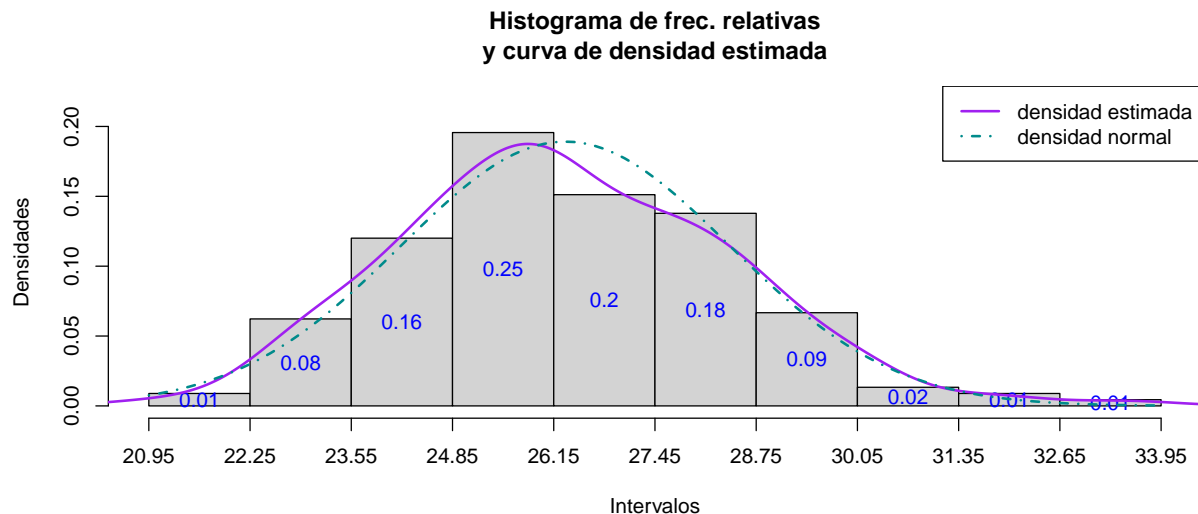
La curva de densidad que hemos obtenido en este gráfico tiene una forma de campana que nos recuerda la campana de Gauss. Para explorar este parecido, vamos a añadir al histograma la gráfica de la función densidad de una distribución normal (campana de Gauss) de media y desviación típica las del conjunto de datos original.

Así, aplicando las instrucciones siguientes, acabamos obteniendo:

```
histRel(cw,L)

# Curva añadida al histograma
# dnorm("data","mean","sd") es la función de distribución normal
curve(dnorm(x, mean(cw), sd(cw)),
      col="cyan4",
      lty=4,
      lwd=2,
      add=TRUE) #add=TRUE (para añadirlo a la figura)

legend("topright",
      lwd=c(2,2), #anchura de linea
      lty=c(1,4), #tipo de linea
      col=c("purple","cyan4"),
      legend=c("densidad estimada","densidad normal"))
```



2.4 Histograma de frecuencias relativas acumuladas

Definimos la función `histRelCum` en el presente informe de RMarkdown.

```
histRelCum = function(x,L){
  h = hist(x, breaks = L,
           right = FALSE,
           plot = FALSE)

  h$density = cumsum(h$counts)/length(x)
```

```

plot(h, freq = FALSE,
     main = "Histograma de frec. rel. acumuladas\n y curva de distribución estimada",
     xaxt = "n",
     col = "lightgray",
     xlab = "Intervalos",
     ylab = "Frec. relativas acumuladas")

axis(1, at = L)

text(h$mids,
     h$density/2,
     labels = round(h$density ,2),
     col = "blue")

dens.x = density(x)

dens.x$y = cumsum(dens.x$y)*(dens.x$x[2]-dens.x$x[1])

lines(dens.x,col = "purple",lwd = 2)
}

```

Dibujamos el histograma con `histRelCum`.

```
histRelCum(cw,L)
```

