

# Aprendizaje Estadístico - (84:04)

Docente: Jemina García

Alumnos: Alejandro Verri Kozlowski, José Manuel Pereiras

## Contents

## Estimación no-paramétrica

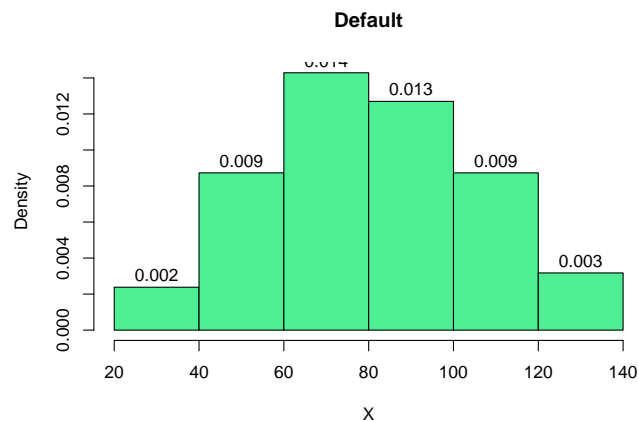
### Ejercicio 1

En el archivo `buffalo.txt` se encuentran los datos que corresponden a la mediciones de cantidad de nieve caída (en pulgadas) en Buffalo en los inviernos de 1910/1911 a 1972/1973. Realizar un histograma para estos datos utilizando los parámetros por defecto.

```
DT <- fread("data/buffalo.txt", sep=" ") |> transpose()
X <- DT$V1 |> sort()
summary(X)
```

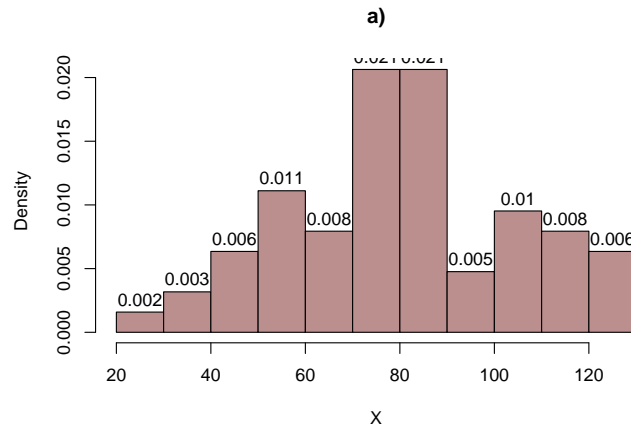
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  25.00   64.50   79.60   80.30   97.65  126.40
```

```
hist(X, main="Default", col = "seagreen2", labels = TRUE, freq=FALSE)
```

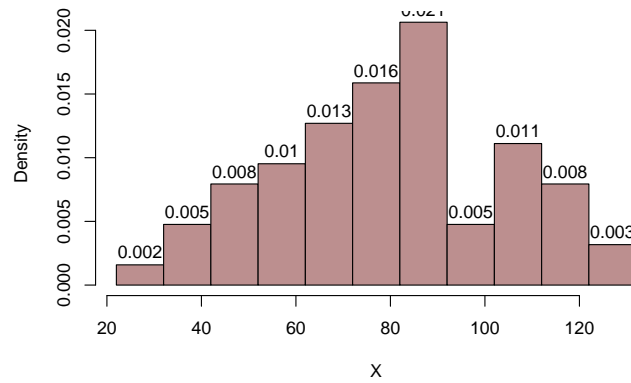


Repetir eligiendo como puntos de corte las siguientes secuencias

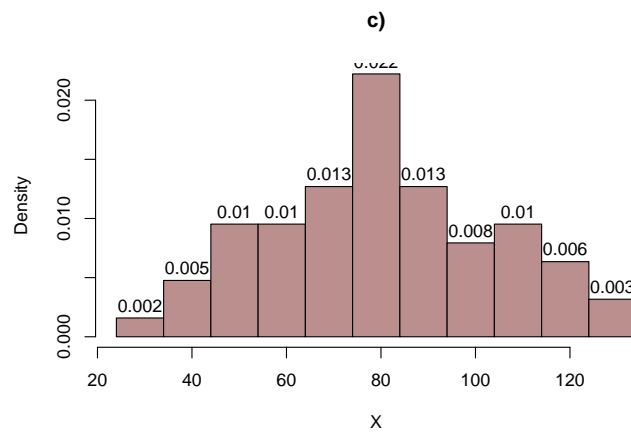
- De 20 a 130 con un paso de 10



b. De 22 a 132 con un paso de 10



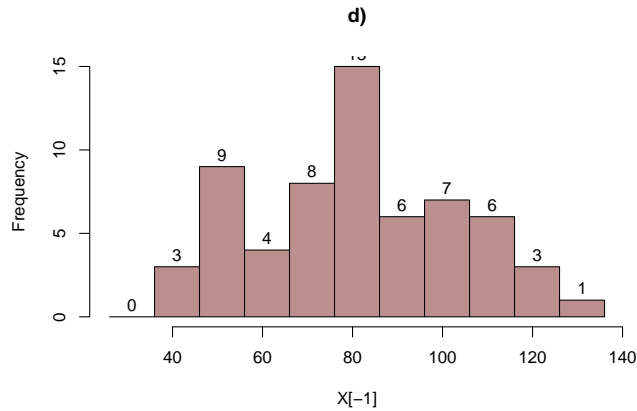
c. De 24 a 134 con un paso de 10



d. De 26 a 136 con un paso de 10 <sup>1</sup>

---

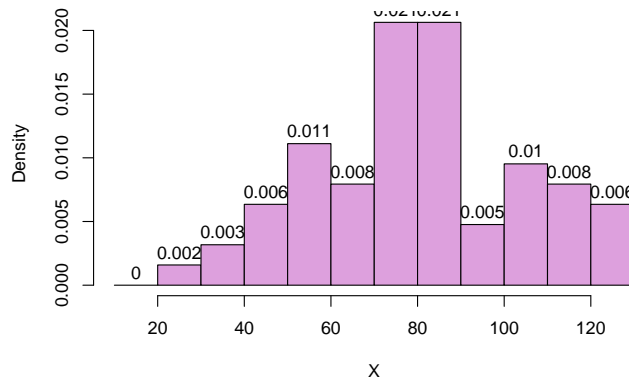
<sup>1</sup>Este histograma no puede visualizarse con las propiedades por defecto de la función `hist()`. El valor mínimo de `X` 25 queda fuera del límite inferior de los puntos de corte. Para poder visualizar este histograma debe excluirse el primer punto del vector aleatorio `X`



## Ejercicio 2

Realizar un histograma para estas observaciones utilizando puntos de corte (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130). Comparar todos los histogramas obtenidos. ¿Tiene algún efecto la elección del punto inicial para estos datos?

```
BRK <- seq(from=10,to=130,by=10)
hist(X,breaks=BRK,main="", col = "plum", labels = TRUE,freq=FALSE)
```



Los histogramas con el mismo ancho de contenedor  $h$  (binwidth) son diferentes y dependen fuertemente de la elección del origen  $x_0$ . Para un mismo  $h$ , la elección del punto inicial  $x_0$  cambia completamente la forma (distribución de frecuencias) del histograma.

## Ejercicio 3

Sea  $X$  la cantidad de nieve caída en un invierno en Buffalo. Implementar una función que dados los valores  $x$ ,  $h$  y un conjunto de datos  $X$ , permita estimar a  $P[X \in (x - h, x + h)] = P[x - h \leq X \leq x + h]$  para cada valor  $x$ .

*Solución:* Si  $X$  es una V.A.C, la probabilidad de  $X$  esté comprendida en el intervalo abierto  $(x - h, x + h)$  puede estimarse mediante

$$P[X \in (x - h, x + h)] \approx n^{-1} \sum_{i=1}^n \mathbf{I}[X_i \in (x - h, x + h)]$$