Ejercicio 2

```r
datos <- scan("lamparas.txt")
datos
```

```
##  [1]  26.43  33.58  65.86  29.18   5.92  13.29  13.54  64.78  56.11  23.60
## [11]  33.39 100.32  28.04  29.63   2.41   3.17  11.99   6.47  23.59  17.96
## [21]  32.27   2.09  57.43  15.31  42.85   1.68  49.61
```

a) Estimar la probabilidad de que una lámpara producida por esta fábrica **dure más de 30 horas**.

b) Implementar y graficar la **función de distribución empírica** de este conjunto de datos.

c) Completar: *Estos datos permiten estimar que el 90 % de las lámparas producidas por esta fábrica dura más de . . . . . . horas y el 10 % dura menos de . . . . . . horas.*

Primero, para el item a nos piden la probabilidad de que una lampara dure mas de 30hs Para eso, como nuestros datos son diversos todos entre si podemos estimar la probabilidad de dicho suceso, para ello nos basamos en aplicar la funcion de distribucion empirica que nos dice:

$$\widehat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I(x \leq t)$$

¿Porque tomamos la funcion empirica? Pues porque tenemos una cantidad de datos los cuales van dando "saltos" a razon de 1/N por cada uno. La funcion empieiza nos permite calcular esa cantidad de saltoss en relacion a la cantidad total de elementos que tengamos

Donde n es la cantidad de datos totales y la sumatoria es que cada elemento cumpla, en nuestro caso que la duracion sea mayor a 30

```r
# Cantidad de datos > 30
length(datos[datos > 30])
```

```
## [1] 10
```

```r
# Cantidad total de datos
length(datos)
```

```
## [1] 27
```

```r
# Probabilidad estimada (empírica) de que dure más de 30hs
length(datos[datos > 30])/length(datos)
```

```
## [1] 0.3703704
```

Rta. a)

Probabilidad Empírica
$$P_{emp}(X_i > 30) = \frac{10}{27} \approx 0.370$$

Item B

Histograma $\equiv$ Estimador de Densidad

Construcción del Histograma:

- Sea la ***muestra*** $Y_n = X_1, X_2, \ldots, X_n$ de **variables aleatorias**.
- Sea la ***realización*** $y_n = x_1, x_2, \ldots, x_n$
  Donde ***realización*** son los **datos** (x_i **en** $\mathbb{R}^d$) **obtenidos en un experimento**
- Sea la ***realización ordenada*** $X_n = x_{(1)}, x_{(2)}, \ldots, x_{(n)}$
- Sea $I_1, I_2, \ldots, I_k$ intervalos
  típicamente $k = \lfloor\sqrt{n}\rfloor$

Sobre cada intervalo I_j elejimos una $altura_j$ tal que el área $I_j \times altura_j$ sea igual a la **frecuencia relativa** en ese intervalo I_j

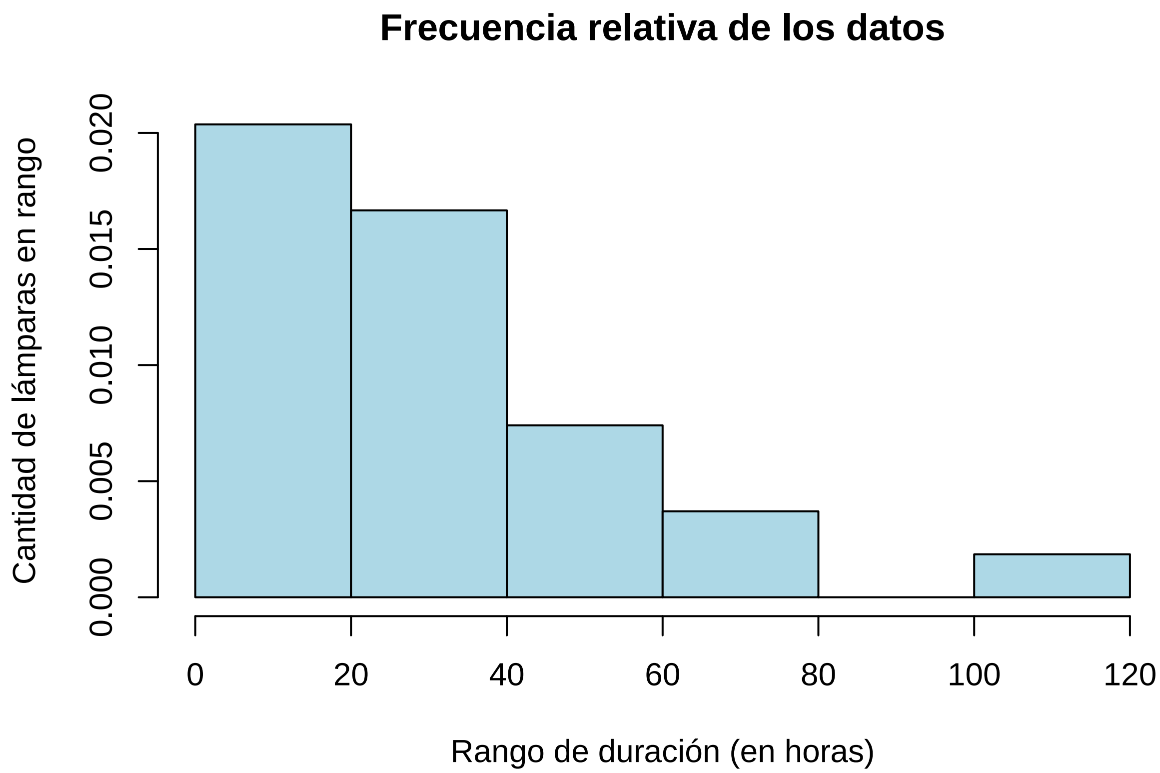
$$\acute{A}rea\ en\ I_j \equiv Frecuencia\ relativa\ en\ I_j$$

$$|I_j| \times altura_j = \frac{\sum_{i=1}^{n} \mathbf{1}\{x_i \in I_j\}}{n} = \frac{\text{cant. de datos} \in I_j}{\text{cant. total de datos}}$$

De esta forma obtenemos una representación o modelo de frecuencias relativas de los datos.

Como el área de histograma suma 1, podemos usarlo como **Función de Densidad Empírica**, donde ***Empírica*** hace referencia a que es definida a través de observaciones (en nuestro caso, los `datos`)

```r
# Histograma
hist(datos,
     main="Frecuencia relativa de los datos",
     xlab="Rango de duración (en horas)",
     ylab="Cantidad de lámparas en rango",
     col="lightblue",
     freq=FALSE,
     breaks=6)
```



Ahora vamos con el item c Observemos que podemos sacar deduciendo los datos

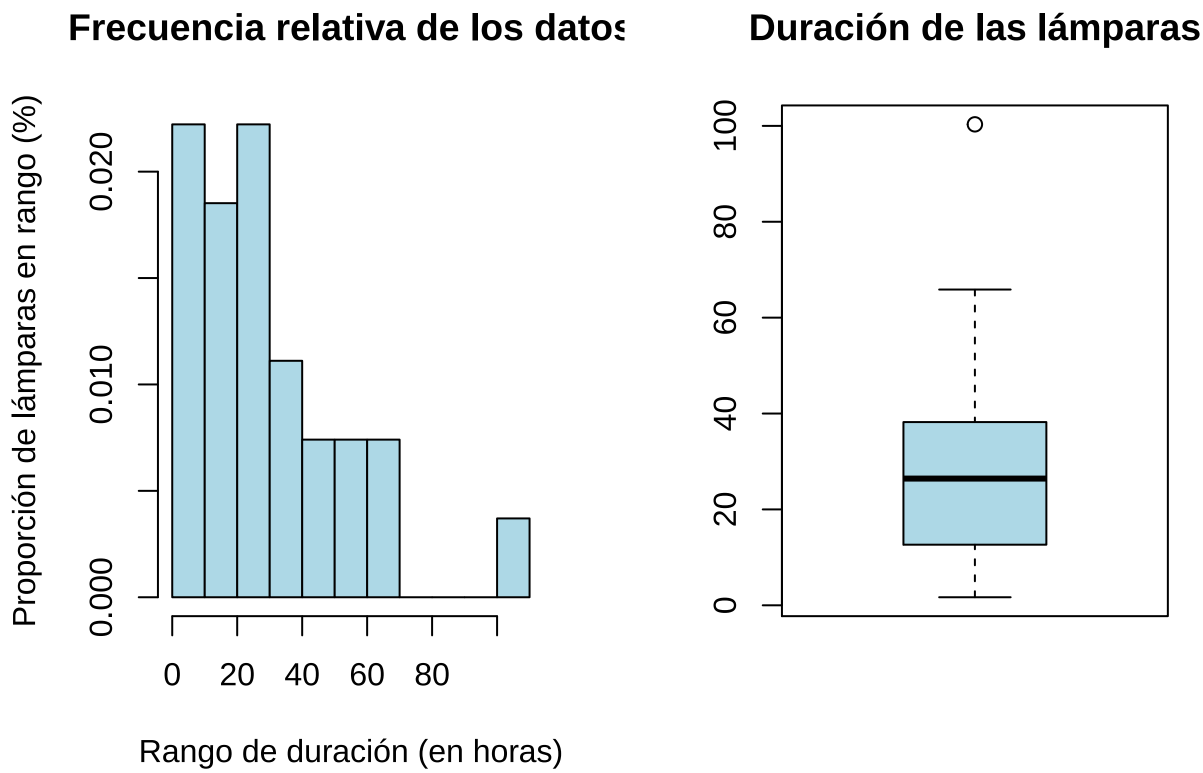
```
sort(datos)
```

```
##  [1]   1.68   2.09   2.41   3.17   5.92   6.47  11.99  13.29  13.54  15.31
## [11]  17.96  23.59  23.60  26.43  28.04  29.18  29.63  32.27  33.39  33.58
## [21]  42.85  49.61  56.11  57.43  64.78  65.86 100.32
```

Podemos hacer un grafico asi tenemos una idea de que estamos haciendo

```
par(mfrow=c(1,2))
# Frecuencia de los datos
hist(datos,
     main="Frecuencia relativa de los datos",
     xlab="Rango de duración (en horas)",
     ylab="Proporción de lámparas en rango (%)",
     col="lightblue",
     freq=FALSE,
     breaks=11)


# Boxplot de datos
boxplot(datos,
        main="Duración de las lámparas",
        col="lightblue")
```



Ahora teniendo un mejor panorama, podemos decir que el 90% de las lamparas producidas en dicha fabrica dura mas de… Buscamos entonces, como tenemos al 90%, el percentil 10 a la derecha

```r
quantile(datos, seq(0,1,0.1))
```

```
##      0%     10%     20%     30%     40%     50%     60%     70%     80%     90%
##   1.680   2.866   7.574  13.490  20.212  26.430  29.450  33.428  48.258  60.370
##    100%
## 100.320
```

Teniendo estos datos en cuenta, podemos decir que el 90% de las lamparitas dura mas de 2.866 horas, por ende # Rta

Respuesta
Estos datos permiten estimar que el ***90 % de las lámparas*** *producidas por esta fábrica* ***dura más de*** 2.866 *horas*

Equivalentemente, usando el ***90-percentil***:

Respuesta
y el ***10 % dura menos de*** 60.37 *horas*