

## Resumen de los comandos básicos del programa R

**Introducción de datos:** Para introducir un conjunto de datos  $\{dato1, dato2, \dots\}$  con el nombre 'x' escribimos

```
x = c(dato1,dato2,dato3...)
```

**Media, mediana, desviación típica, varianza...**

```
mean(x)
median(x)
sd(x)
var(x)
```

**Comando de resumen (datos máximo y mínimo, cuartiles,...)**

```
summary(x)
```

**Histograma, diagrama de caja, diagrama de tallo y hojas,...**

```
hist(x)
boxplot(x)
stem(x)
```

**Comando de ayuda**

```
help(duda)
```

por ejemplo `help(hist)` nos explica cómo usar el comando "hist".

### NOTAS:

1.- El programa **R** define la desviación típica como  $dt(x) = \sigma_{n-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ , y análogamente  $var(x) = \sigma_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ . Para calcular  $\sigma_n$  hay que hacer la multiplicación correspondiente, es decir

$$\sigma_n = sd(x) * ((n - 1)/n)^{1/2}.$$

2.- El comando `hist(x)` dibuja por defecto un *histograma de frecuencias absolutas* con intervalos de la forma  $(a_{i-1}, a_i]$ , y puntos de separación  $a_i$  elegidos según la "regla de Sturges". Si se desea un histograma de densidades, intervalos abiertos por la derecha  $[a_{i-1}, a_i)$ , o elegir otros puntos de separación se pueden usar respectivamente los comandos

```
hist(x, freq = F)
hist(x, right = F)
hist(x, breaks = c(a0, a1, ..., ak))
```

3.- Si se desea poner nombres en la cabecera del histograma, o en los ejes  $x$  e  $y$  se pueden usar los comandos `main`, `xlab`, `ylab`, por ejemplo

```
hist(x, main='ejercicio 1', xlab='alturas', ylab='numero individuos')
```

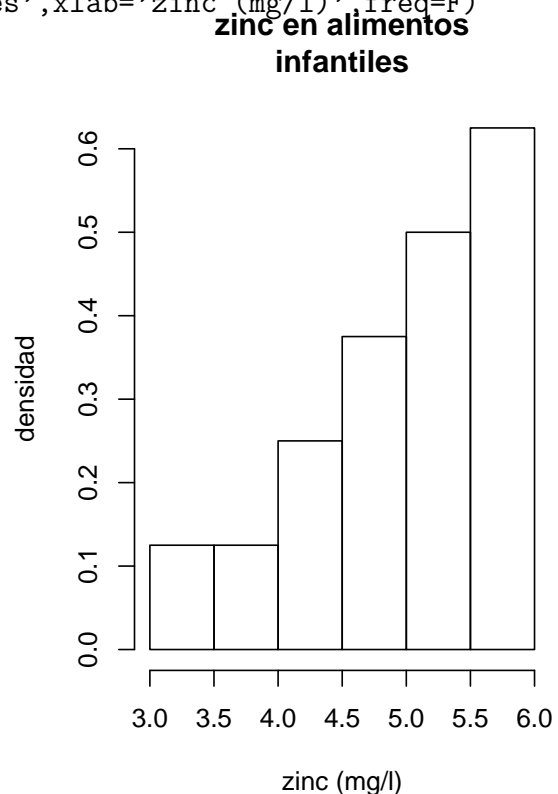
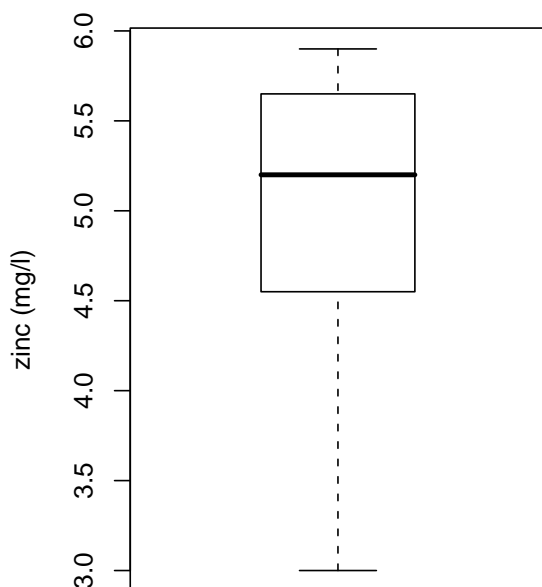
**Ejemplo:** La cantidad de zinc (en mg/l) en 16 muestras de alimentos infantiles viene dada por  
 3.0 5.8 5.6 4.8 5.1 3.6 5.5 4.7 5.7 5.0 5.9 5.7 4.4 5.4 4.2 5.3  
 Hallar media, desviación típica, mediana, cuartiles, y dibujar diagrama de tallos y hojas, box-plot e histograma.

En la pantalla de **R**, tras introducir los comandos correspondientes (después del símbolo de línea ">") nos aparecen como respuestas:

```
>x=c(3.0,5.8,5.6,4.8,5.1,3.6,5.5,4.7,5.7,5.0,5.9,5.7,4.4,5.4,4.2,5.3)
>mean(x)
  4.98125
>sd(x)
  0.8320407
>summary(x)
  Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
  3.000   4.625   5.200   4.981   5.625   5.900
>stem(x)
  The decimal point is at the |
  3 | 06
  4 | 2478
  5 | 0134567789
>boxplot(x, ylab='zinc (mg/l)')
>hist(x,main='zinc en alimentos infantiles', xlab='zinc (mg/l)', ylab='%',freq=F)
```

**NOTA:** Para que varias figuras aparezcan simultaneamente en la misma pantalla, se puede usar el comando `split.screen(c(n,m))`, que subdivide la pantalla en una matriz de `n` filas y `m` columnas, e incorpora una figura en cada casilla de la matriz. En el ejemplo anterior

```
>split.screen(c(1,2))
>boxplot(x, ylab='zinc (mg/l)')
>screen(2)
>hist(x,main='zinc en alimentos infantiles',xlab='zinc (mg/l)',freq=F)
```



## REGRESIÓN LINEAL

### Introducir datos

```
x = c(x1, x2, ...)  
y = c(y1, y2, ...)
```

### Covarianza y correlación

```
cov(x, y)  
cor(x, y)
```

### Coeficientes de la recta de regresión: $Y = A + BX$

```
lm(y ~ x)
```

### Gráficas de la nube de puntos y de la recta de regresión

```
plot(x, y)  
abline(lm(y ~ x))
```

### NOTAS:

- 1.- El programa **R** define la covarianza como  $\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ .
- 2.- Es importante no confundir los comandos `lm(y ~ x)` y `lm(x ~ y)`: el primero da la recta de regresión  $Y = A + BX$  y el segundo la recta  $X = A + BY$  (que en general no coinciden).
- 3.- Se pueden añadir nombres en la cabecera de las gráficas o etiquetas en los ejes, usando los comandos `main`, `xlab`, `ylab`, por ejemplo

```
plot(x, y, main='ejercicio 2', xlab='altura padres', ylab='altura hijos')
```

**Ejemplo:** *En el experimento de Galton se estudia la relación entre el tamaño de una semilla y el tamaño de sus descendientes, obteniéndose los datos (diámetro de la semilla en mm)*

tamaño padres	15	16	17	18	19	20	21
tamaño hijos	15,4	15,7	16	16,3	16,6	17	17,3

*Calcular la correlación entre las variables y la recta de regresión, esbozando la gráfica.*

```
>x=c(15, 16, 17, 18, 19, 20, 21)
```

```
>y=c(15.4, 15.7, 16, 16.3, 16.6, 17, 17.3)
```

```
>cor(x, y)
```

```
0.9990545
```

```
>lm(y~x)
```

```
Call: lm(formula = y ~ x)
```

```
Coefficients:      (Intercept)          x  
10.6071      0.3179
```

*Solución:* la recta de regresión es  $y = 10,6071 + 0,3179x$ .

```
>plot(x, y, main='experimento de Galton', xlab='diametro padres', ylab='diametro hijos')
```

```
>abline(lm(y~x))
```

### experimento de Galton

