

Lección 2: Estadística descriptiva

Bioestadística con R

Ejercicios

Ejercicio 1: Un botánico hizo crecer 15 pimientos en el mismo invernadero y manteniendo las condiciones de cada pimiento constantes durante 21 días. Una vez transcurridos los 21 días, ella midió la longitud total del tallo (cm) de cada planta, y obtuvo los siguientes resultados:

12.4	10.9	11.8	14.1	12.6
12.2	12.2	13.5	12.7	11.9
13.4	12.1	12.0	13.2	13.1

A partir de estos valores calcular:

- Calcular la media y mediana.
- Calcular el rango, la desviación estándar y la varianza.
- Calcular los tres cuartiles.
- Calcular el límite superior e inferior de los valores atípicos.
- *Actividad extra:* Realizar un gráfico de cajas utilizando `ggplot2`.

Ejercicio 2: Un genetista contó el número de cerdas en una región del abdomen de la mosca *Drosophila melanogaster*. El resultado de 119 individuos fue el siguiente:

Cerdas	Individuos	Cerdas	Individuos
29	1	38	18
30	0	39	13
31	1	40	10
32	2	41	15
33	2	42	10
34	6	43	2
35	9	44	2
36	11	45	3
37	12	46	2

Con base en estos resultados calcular:

- Media del número de cerdas.
- La desviación estándar de la muestra.
- ¿Cuál es el coeficiente de variación?
- *Actividad extra:* Realizar un gráfico de cajas utilizando `ggplot2`.

Ejercicio 3: Científicos midieron la concentración de calcio (nM) en la sangre de muestras de 38 personas. Los datos son los siguientes:

95	110	135	120	88	125
112	100	130	107	86	130
122	122	127	107	107	107
126	125	112	78	115	78
102	103	93	88	110	104
122	112	80	121	126	90
96	88				

Con base en estos datos:

- Calcular medidas de tendencia central y dispersión.
- Describir la forma de la distribución.
- *Actividad extra:* Realizar un gráfico de cajas utilizando `ggplot2`.

Soluciones

Ejercicio 1: Primero creamos un vector con nuestros datos:

```
ej1 <- c(12.4, 10.9, 11.8, 14.1, 12.6,  
         12.2, 12.2, 13.5, 12.7, 11.9,  
         13.4, 12.1, 12.0, 13.2, 13.1)
```

Una vez que tenemos nuestro vector, podemos calcular la media y la mediana con las funciones `mean()` y `median()`:

```
mean(ej1)
```

```
## [1] 12.54
```

```
median(ej1)
```

```
## [1] 12.4
```

Para el rango, podemos utilizar las funciones `max()` y `min()`, así como la función `range()` para conocer el valor mínimo y máximo.

```
range(ej1)
```

```
## [1] 10.9 14.1
```

```
max(ej1) - min(ej1)
```

```
## [1] 3.2
```

Los cuartiles podemos calcularlos usando la función `summary()`:

```
summary(ej1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.90   12.05   12.40   12.54   13.15   14.10
```

Para el rango intercuartil, utilizamos el valor del 1er y 3er cuartil:

```
13.15 + 1.5 * (max(ej1) - min(ej1))
```

```
## [1] 17.95
```

```
12.05 - 1.5 * (max(ej1) - min(ej1))
```

```
## [1] 7.25
```

Ejercicio 2: Creamos un vector con los valores de las cerdas. Nos apoyamos de la función `rep()` para el número de individuos:

```
ej2 <- c(29, 31, rep(32, 2), rep(33, 2), rep(34, 6), rep(35, 9), rep(36, 11),
        rep(37, 12), rep(38, 18), rep(39, 13), rep(40, 10), rep(41, 15), rep(42, 10),
        rep(43, 2), rep(44, 2), rep(45, 3), rep(46, 2))
```

Con este nuevo vector, podemos hacer el calculo de nuestra media y desviación estándar:

```
mean(ej2)
```

```
## [1] 38.45378
```

```
sd(ej2)
```

```
## [1] 3.196258
```

Para calcular el coeficiente de variación, sacamos la desviación estándar, la dividimos entre la media y multiplicamos por 100:

```
sd(ej2)/mean(ej2)*100
```

```
## [1] 8.311948
```

Ejercicio 3: Primero creamos nuestro vector con los datos:

```
ej3 <- c(95, 110, 135, 120, 88, 125, 112, 100, 130, 107, 86, 130, 122, 122,
        127, 107, 107, 107, 88, 126, 125, 112, 78, 115, 78, 102, 103, 93,
        88, 110, 104, 122, 112, 80, 121, 126, 90, 96)
```

Posteriormente calculamos las medidas de tendencia central (media y mediana):

```
mean(ej3)
```

```
## [1] 107.8684
```

```
median(ej3)
```

```
## [1] 108.5
```

Ahora calculamos las medidas de dispersión (rango, desviación estándar y varianza):

```
max(ej3) - min(ej3)
```

```
## [1] 57
```

```
sd(ej3)
```

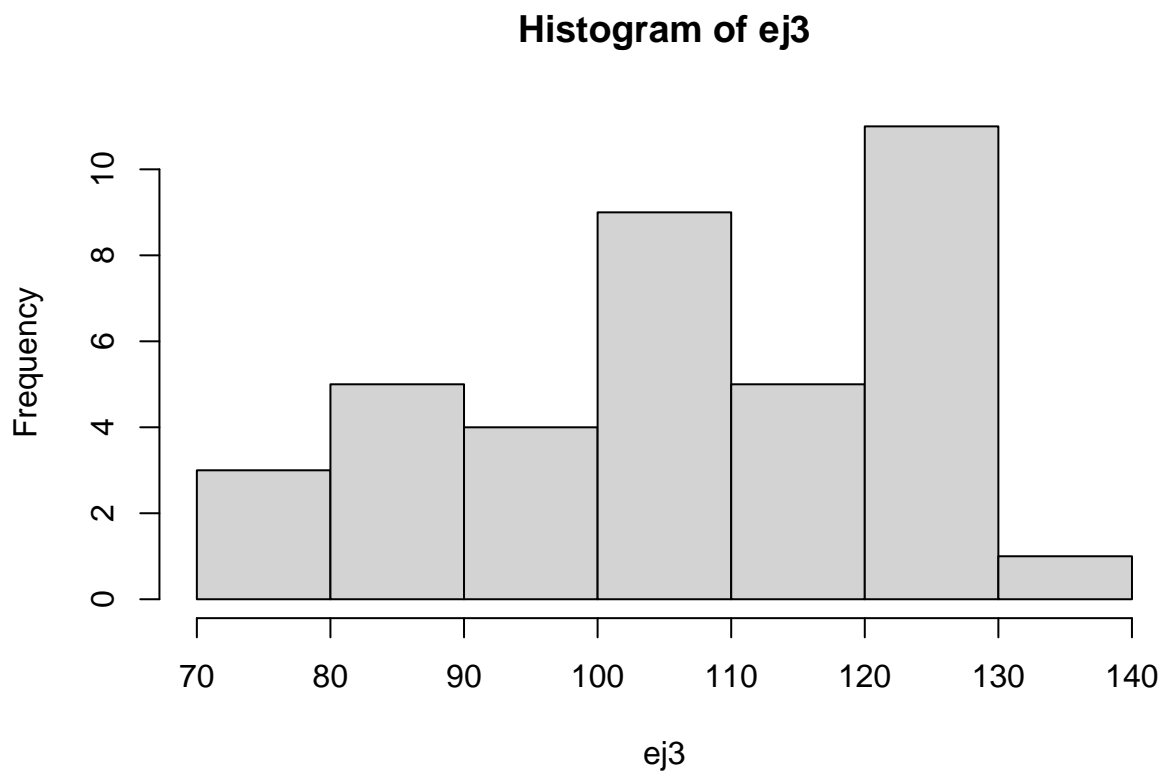
```
## [1] 16.0778
```

```
sd(ej3)^2
```

```
## [1] 258.4957
```

Para visualizar la distribución, podemos generar un histograma de manera rápida:

```
hist(ej3)
```

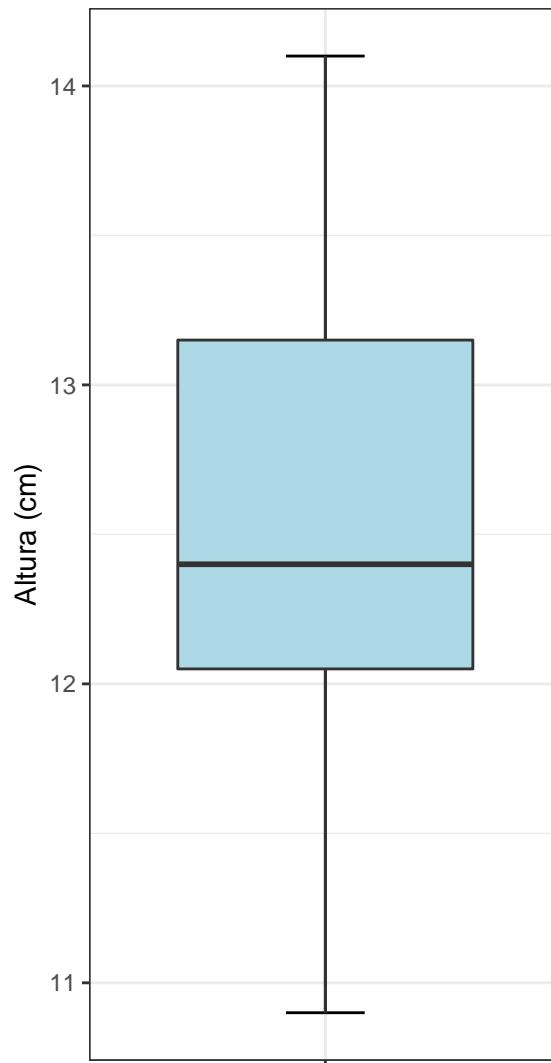


Actividades extra: Primero necesitamos cargar la librería `ggplot2`.

```
library(ggplot2)
```

Realizaremos el gráfico de caja del primer ejercicio:

```
ggplot(mapping = aes(y = ej1, x = "")) +  
  stat_boxplot(geom = "errorbar", width = 0.2) +  
  geom_boxplot(fill = "lightblue") +  
  theme_bw() +  
  labs(x = "", y = "Altura (cm)")
```



Para el resto de ejercicios podemos utilizar el mismo código, cambiando el nombre de la variable a utilizar.