

# Práctica 2.Soluciones.

Estadística CCAA 2020-21

Actualizado: 2024-10-24

## Introducción.

Los datos que aparecen en el fichero `p1-robles.csv` se refieren a un estudio realizado sobre un robledal cercano a una planta industrial, parte de los cuales se visualizan en la tabla. Se han seleccionado robles

- De dos variedades (A y B).
- Ubicados en cuatro zonas distintas.
- La mitad de ellos han sido sometidos a cierto tratamiento (codificados con 1), los no tratados (codificados con 0).

Sobre cada árbol se han medido las concentraciones (mg/kg) de ocho elementos químicos en sus hojas:

- Metales pesados: hierro, manganeso y zinc
- Metales alcalinotérreos: calcio y magnesio
- Metal alcalino: potasio
- No metales: fósforo y nitrógeno.

## Ejercicio 1

Calcula los cuartiles de la variable Hierro e interpretalos.

```
quantile(robles$Hierro)
```

```
##      0%      25%      50%      75%     100%  
## 0.0070 0.0110 0.0145 0.0590 0.0630
```

Quiere decir que

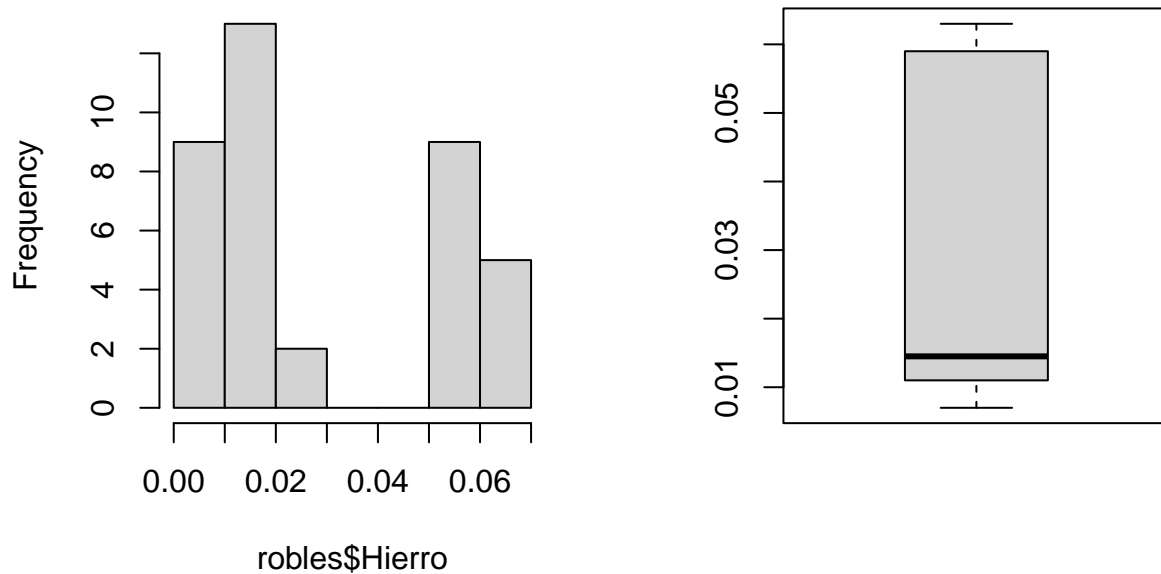
- El 25% de los robles tienen 0.011mg/kg de hierro o menos
- El 50% de los robles tienen 0.0145mg/kg de hierro o menos (y el otro 50% más de esa cantidad)
- El 25% de los robles tienen 0.063mg/kg de hierro o más

## Ejercicio 2

Considera la variable Hierro, calcula su histograma y su boxplot. Compara lo que te dice cada una de las figuras.

```
par(mfrow = c(1,2))  
hist(robles$Hierro)  
boxplot(robles$Hierro)
```

## Histogram of robles\$Hierro



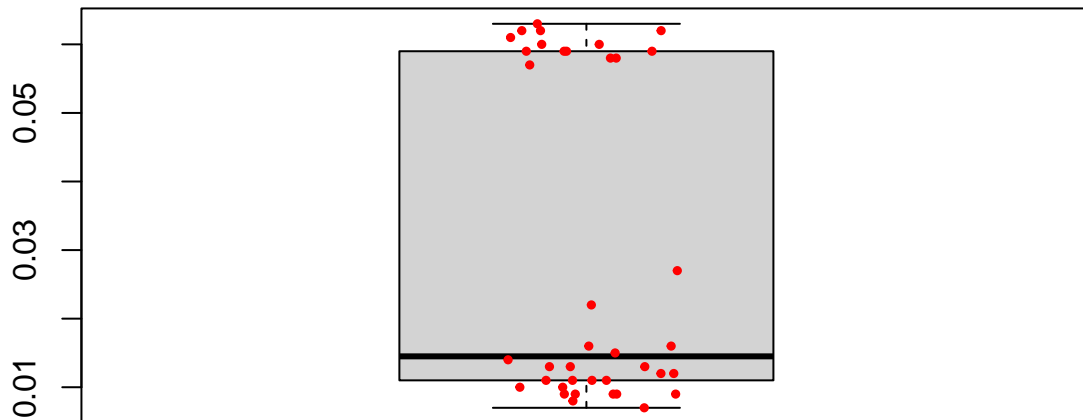
```
par(mfrow = c(1,1))
```

En el histograma se aprecia de forma evidente que los datos están agrupados en dos bloques. No hay ningún roble con una concentración de Hierro entre 0.03 y 0.05.

En el boxplot esto no es tan evidente (aunque algo se intuye). Se observa una distribución no uniforme de los datos: están muy concentrados por debajo de la mediana (bigote inferior próximo a la caja y mitad inferior de la caja muy estrecha) y por encima del tercer cuartil (bigote superior próximo a la caja). Están muy dispersos entre la mediana y el tercer cuartil (mitad superior de la caja).

Si sobreimpresionamos los valores en el boxplot

```
boxplot(robles$Hierro)
stripchart(robles$Hierro,
  method = "jitter", add = TRUE,
  pch=19, col= "red", cex=0.5, vertical = TRUE)
```



se evidencia la no homogeneidad en la distribución de los datos y el hecho de que ningún roble presenta una concentración de Hierro entre 0.03 y 0.05.

**Cosas de R:** fijate en que en `stripchart` el argumento `vertical` vale `TRUE`. Por defecto, el boxplot se representa verticalmente, pero los puntos horizontalmente (no sólo se usan con el boxplot, claro). Por eso hay que añadir, o bien en `boxplot` el argumento `horizontal = TRUE`, o bien en `stripchart` el argumento `vertical = TRUE`. Si haces las dos cosas simultaneamente, seguirás sin ver los puntos sobre el boxplot.

### Ejercicio 3

¿Entre qué valores se mueve el 80% central de la muestra respecto de la variable **Magnesio**? Se trata de los valores que están por encima del percentil 10 y por debajo del percentil 90

```
quantile(robles$Magnesio, probs = c(0.1, 0.9))
```

```
##      10%      90%
## 0.2949 0.4881
```

### Ejercicio 4

¿Qué porcentaje de los robles tienen una concentración de **Magnesio** mayor que 0.3?

```
(length(robles$Magnesio[robles$Magnesio>0.3])/length(robles$Magnesio))*100
```

```
## [1] 86.84211
```

### Ejercicio 5

¿Cuál es el valor mínimo que alcanza la variable **Magnesio**? ¿En qué individuo se observa?

```
min(robles$Magnesio)
```

```
## [1] 0.267
```

```
which(robles$Magnesio == min(robles$Magnesio))
```

```
## [1] 34 35
```

### Ejercicio 6

- **Calcula la media y la cuasidesviación típica muestral de la variable Nitrogeno.** Usaremos las funciones mean y sd

```
mean(robles$Nitrogeno)
```

```
## [1] 3.293737
```

```
sd(robles$Nitrogeno)
```

```
## [1] 0.6415097
```

- **Repite el cálculo para los primeros 14 individuos de la tabla.** Recuerda que 1:10 genera la sucesión 1, 2, 3, ..., 10, y que los corchetes [] sirven para referirse a los elementos de un vector/tabla

```
mean(robles$Nitrogeno[1:14])
```

```
## [1] 3.506857
```

```
sd(robles$Nitrogeno[1:14])
```

```
## [1] 0.6564765
```

### Ejercicio 7

¿Cuántas observaciones corresponden a robles con una concentración de Potasio mayor o igual que 2? Puedes recuperar los valores del Potasio de los robles que tienen más de 2 de Potasio

```
robles$Potasio[robles$Potasio >= 2]
```

y, a continuación, calcular la longitud de dicho vector

```
length(robles$Potasio[robles$Potasio >= 2])
```

```
## [1] 15
```

Otra alternativa es comparar cada elemento del vector robles\$Potasio con 2

```
robles$Potasio >= 2
```

```
## [1] TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE
```

```
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
## [25] FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE FALSE TRUE FALSE TRUE
```

```
## [37] FALSE TRUE
```

para obtener TRUEs y FALSEs según se cumpla o no esa condición. Recuerda que TRUE vale 1 y FALSE cero, con lo que sumas todos esos unos y ceros obtienes la cantidad deseada

```
sum(robles$Potasio >= 2)
```

```
## [1] 15
```

### Ejercicio 8

Ahora calcula la media de la variable Nitrogeno para aquellas observaciones en las que Potasio es mayor o igual que 2. Compárala con la media que has calculado en el anterior apartado. Haz lo mismo con la cuasidesviación típica. Basta con calcular ambas medias

```
mean(robles$Nitrogeno)
```

```
## [1] 3.293737
```

```
mean(robles$Nitrogeno[robles$Potasio >= 2])
```

```
## [1] 3.570467
```

en el segundo caso la media es ligeramente superior, y las cuasidesviaciones típicas

```
sd(robles$Nitrogeno)
```

```
## [1] 0.6415097
```

```
sd(robles$Nitrogeno[robles$Potasio >= 2])
```

```
## [1] 0.5133733
```

### Ejercicio 9

Calcula la media de la variable Nitrogeno para los individuos de la Variedad A (es decir, Variedad == "A") Recicla el código ya escrito para obtener

```
mean(robles$Nitrogeno[robles$Variedad == "A"])
```

```
## [1] 3.206143
```

### Ejercicio 10

Calcula la mediana de la variable Potasio para los individuos de cada Zona Recicla el código ya escrito para obtener

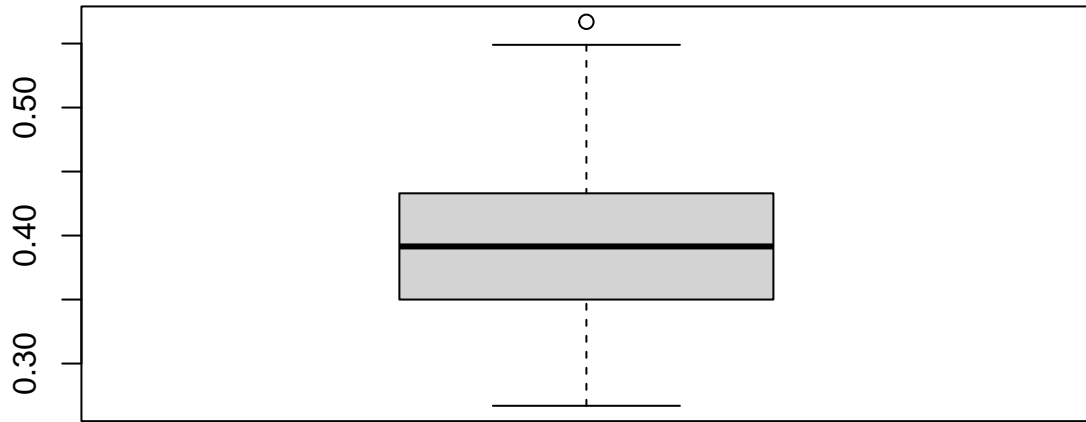
```
aggregate(robles$Nitrogeno ~ robles$Zona, FUN = median)
```

```
##   robles$Zona robles$Nitrogeno
## 1           1           3.4855
## 2           2           3.3270
## 3           3           2.5250
## 4           4           3.3930
```

### Ejercicio 11

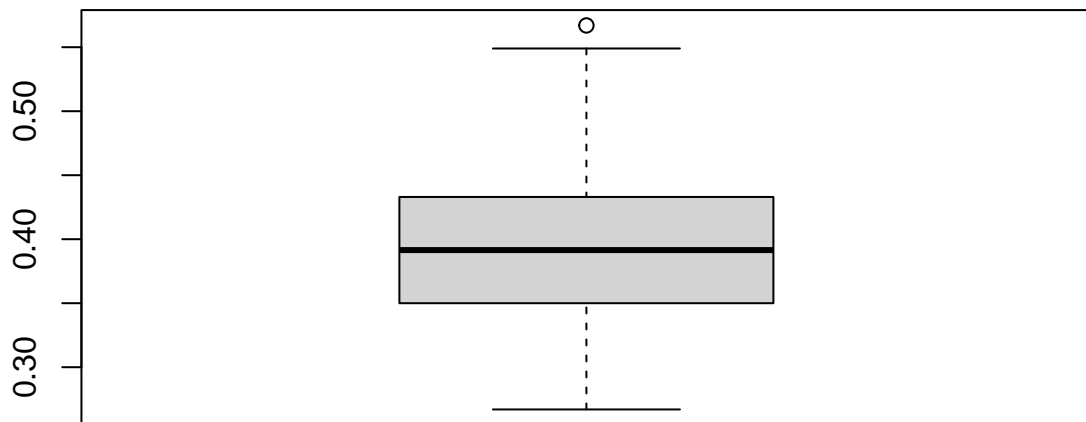
A partir del boxplot de la variable Magnesio, ¿observas algún dato atípico? ¿Qué posición ocupa en la tabla?

```
boxplot(robles$Magnesio)
```



Sí, hay uno. Para calcular la posición que ocupa

```
pb = boxplot(robles$Magnesio)
```



```
which(robles$Magnesio == pb$out)
```

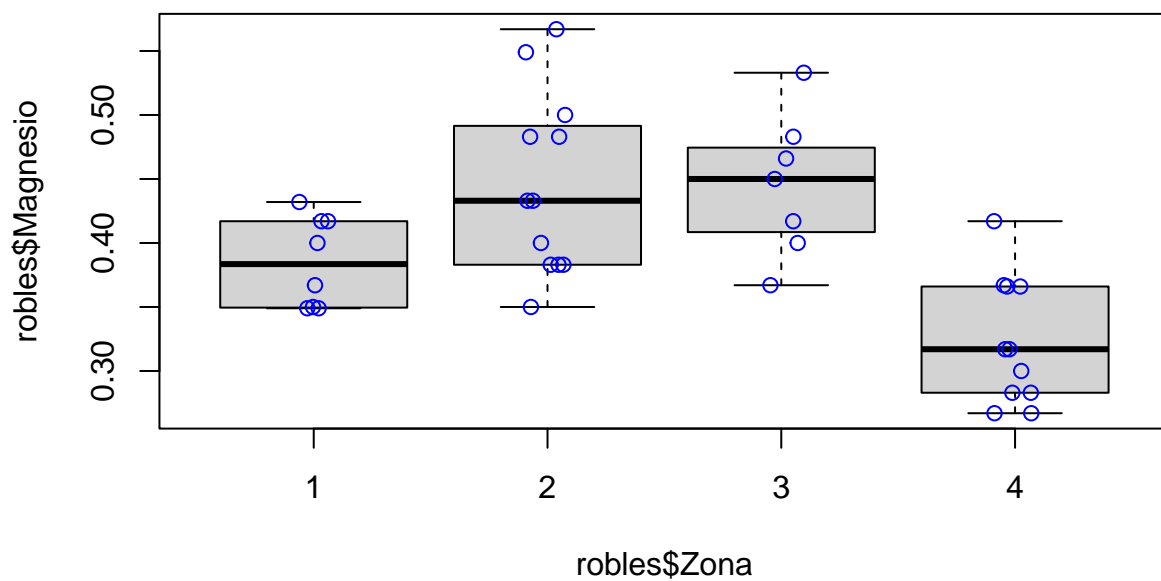
```
## [1] 17
```

## Ejercicio 12

Representa los boxplots de la variable Magnesio para cada una de las zonas, ¿observas algún patrón?

```
boxplot(robles$Magnesio ~ robles$Zona)
```

```
stripchart(robles$Magnesio ~ robles$Zona, method = "jitter", col="blue", add=TRUE, pch = 21, vertical =
```



No olvides representar los puntos sobre los boxplots. Y sí, se observan ciertas tendencias:

- En la zona 4 parece haber menos Magnesio.
- En la zona 1 parece estar menos variabilidad.
- Los valores más altos se encuentran en las zonas 2 y 3