

Guía 3

Probabilidad y Estadística

Ingeniería Industrial - Facultad de Ciencias Agrarias

Ejercicio 1

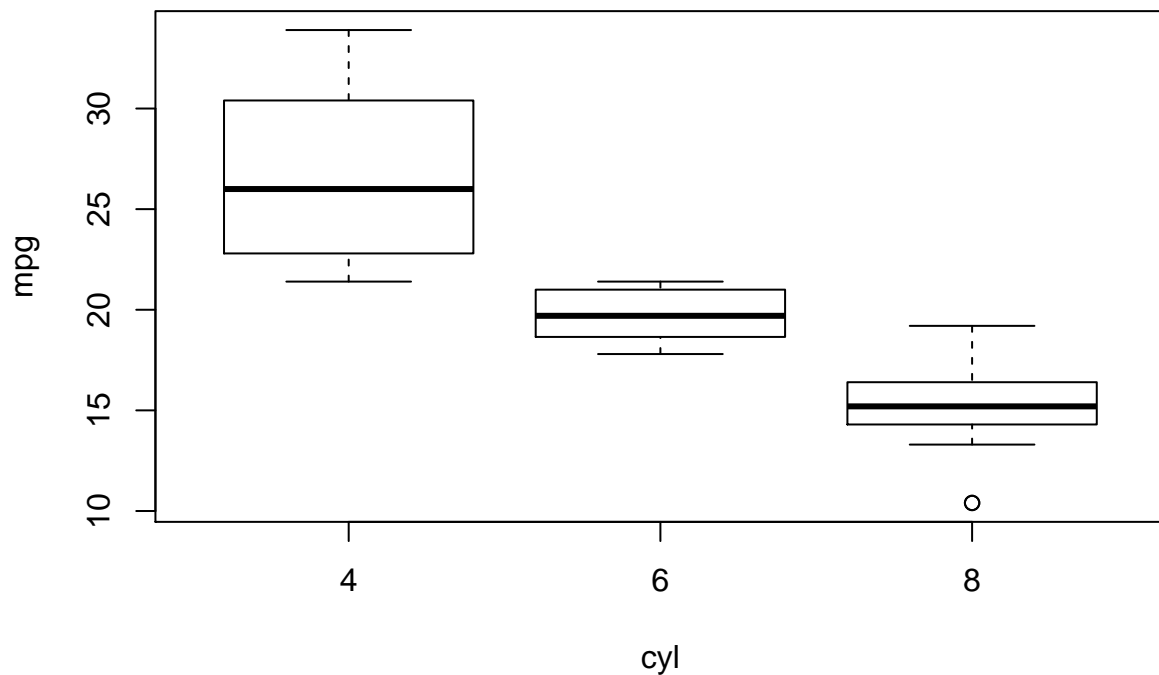
El gráfico siguiente representa la variable millas por galón del dataset mtcars que se encuentra guardado en R.

1. Para la variable millas por galón correspondientes a 4 cilindradas indicar los valores aproximados de la media y los cuantiles 5 y 95.
2. ¿Qué cilindrada presentó mayor variabilidad?
3. Describa la variable millas por galón, en el grupo de 8 cilindradas.

En R

Podemos graficar mediante la función **boxplot()** eligiendo las variables 'mpg' y 'cyl'

```
boxplot(mpg~cyl, data=mtcars) #gráfico de boxplot
```



También podemos elegir otros dos pares de variables

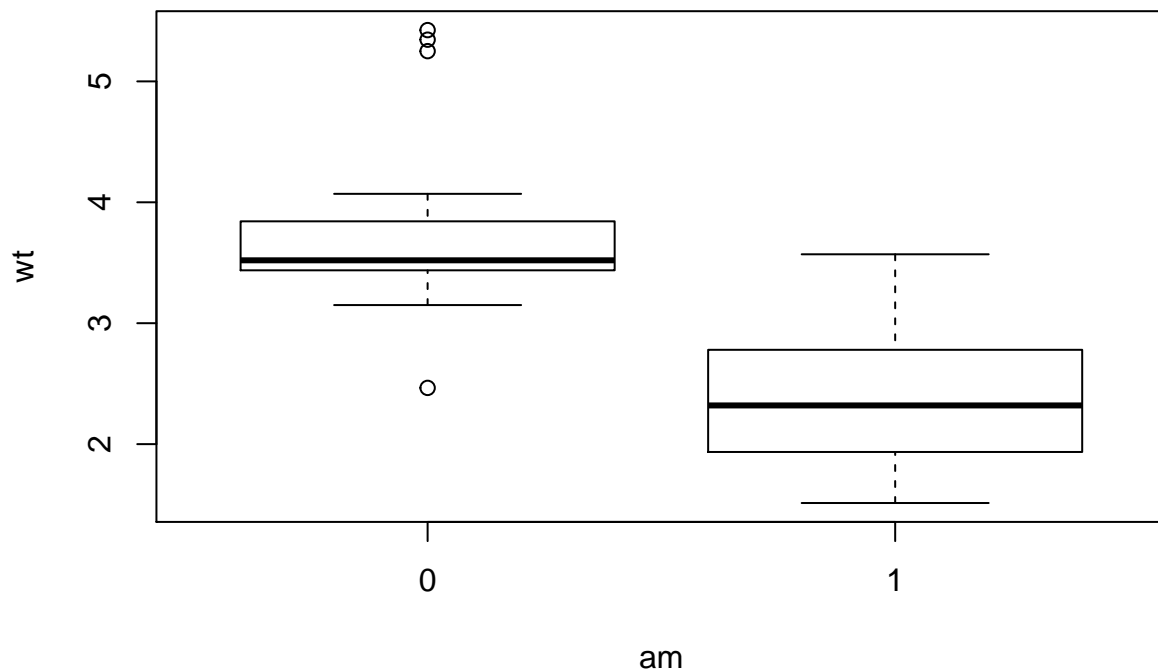
1. wt(peso)
2. am(transmisión)

En la transmisión automática se observan valores extremos ¿cómo interpreta? ¿En qué caso?

En R

Graficamos ahora con las variables 'am' y 'wt'.

```
boxplot(wt~am, data=mtcars) #gráfico de boxplot
```



Ejercicio 2

Los datos de la tabla adjunta corresponden a valores de mg/kg de diferentes fracciones de fósforo en 60 muestras de suelo de la provincia de Corrientes.

11	17	14	20	22	30	20	16	9	15	18	17	41	22	17	18	22	19	23	15
13	20	23	29	21	14	14	16	18	18	43	23	24	19	34	16	24	16	35	24
11	15	12	13	9	41	33	17	12	15	25	18	22	35	20	12	13	11	13	40

1. Ingresar los datos a R como vector de valores enteros con la función c().
2. Calcular la media, la mediana, mínimo, máximo, el primer y tercer cuartil mediante la función summary() e interpretar los resultados.
3. Calcular los percentiles de 5 y 95
4. Realizar un diagrama de cajas mediante la función boxplot(). Interpretar

En R

Ingresamos los datos como *vector de valores enteros*

```
suelo<-c(11, 17, 14, 20, 22, 30, 20, 16, 9, 15, 18, 17, 41, 22, 17, 18, 22, 19, 23, 15,
        13, 20, 23, 29, 21, 14, 14, 16, 18, 18, 43, 23, 24, 19, 34, 16, 24, 16, 35, 24,
        11, 15, 12, 13, 9, 41, 33, 17, 12, 15, 25, 18, 22, 35, 20, 12, 13, 11, 13, 40)
```

En R podemos usar la función **summary()** para conocer la media, la mediana, valores mínimos, máximos, el primer y tercer cuartil.

```
summary(suelo)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      9.00  14.75   18.00   20.28   23.00   43.00
```

En R la función **quantile()** nos permite conocer los cuartiles

```
quantile(suelo)
```

```
##      0%   25%   50%   75%  100%
##      9.00 14.75 18.00 23.00 43.00
```

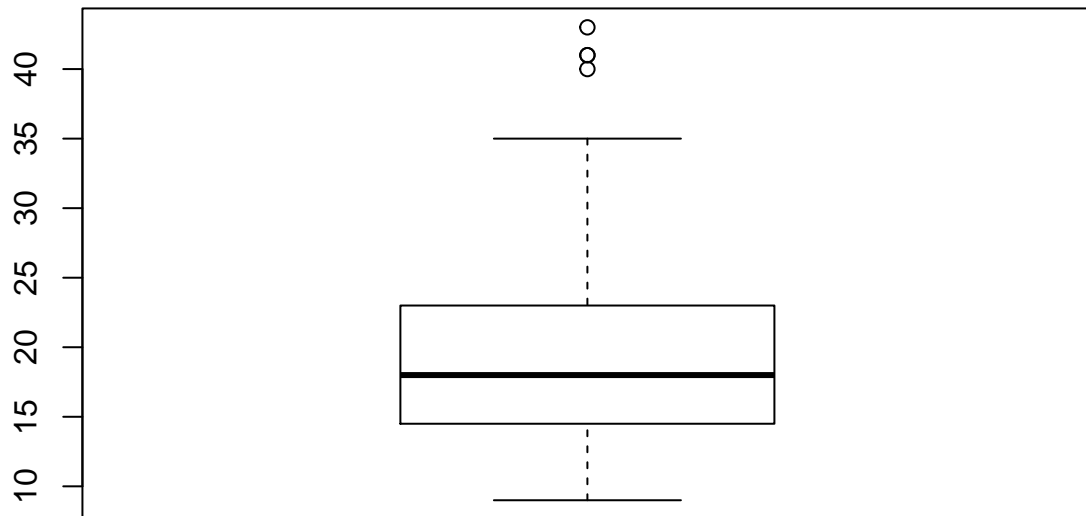
En R, podemos conocer el percentil 5% y 95% mediante el siguiente comando

```
quantile(suelo, probs=c(0.05, 0.95))
```

```
##      5%   95%
##     11.00 40.05
```

Un diagrama de cajas lo podemos realizar mediante la función **boxplot()**

```
boxplot(suelo)
```



Ejercicio 3

En la siguiente tabla se presentan estimadores de medidas de posición y dispersión de las variables altura de planta, peso de granos, humedad de granos, y rendimiento de dos parcelas de trigo.

Parcela	Variable	n	Media	D.E	Var(n-1)	CV	Mediana
1	Altura	16	56.41	10.27	105.57	18.21	59.20
1	Peso Grano	16	49.19	16.69	278.54	33.93	47.05
1	Humedad	16	10.35	3.75	14.09	36.26	9.40
1	Rendimiento	16	1304.19	356.41	127029.76	27.33	1380.00

1. ¿Qué variables presenta mayor variabilidad? Fundamente.
2. ¿A qué se debe la diferencia entre la Media y la Mediana en la variable Rendimiento?