

La distrofia muscular de Duchenne (DMD) es una enfermedad transmitida genéticamente por la madre a los hijos. Las mujeres afectadas normalmente no sufren síntomas visibles y pueden ser portadoras de la enfermedad sin saberlo. La DMD se manifiesta primariamente en los varones debido a que el gen de la enfermedad se encuentra en el cromosoma X. El conjunto de datos `dmd.dat` proviene de Percy *et al.* (1981) y contiene observaciones de 194 mujeres, familiares de niños con DMD. La muestra estaba constituida por 67 portadoras de la DMD y 127 no portadoras. Esta muestra formó parte de un programa canadiense cuyo objetivo era informar a las mujeres de su probabilidad de ser portadoras, basándose en marcadores séricos, así como en su historial familiar. Los dos primeros marcadores, creatina quinasa y hemopexina, son baratos de medir, mientras que la piruvato quinasa o la lactato deshidrogenasa son costosas. Las variables del estudio son

Columna	Variable
1	Edad de la mujer
2	Nivel de creatina quinasa
3	Nivel de hemopexina
4	Nivel de piruvato quinasa
5	Nivel de lactato deshidrogenasa
6	Indicador de si una mujer es portadora (1) de DMD o no (0)

El siguiente código permite cargar los datos, extraer variables de la tabla y separar sus valores en los grupos de portadoras y no portadoras. Ejecuta el código paso a paso para entender qué es lo que hace, mirando la ventana de comandos (console) y las variables creadas en el environment.

```
Datos = read.table("dmd.dat")
C = Datos$V6 # Indicador de si es portadora (1) o no (0)
portadora = (C==1) # TRUE si son portadoras, FALSE si no lo son
CQ = Datos$V2 # Nivel de creatina quinasa
CQportadoras = CQ[portadora] # Creatina quinasa en portadoras
CQnoportadoras = CQ[!portadora] # Creatina quinasa en no portadoras
PQ = Datos$V4 # Nivel de piruvato quinasa
PQportadoras = PQ[portadora] # Piruvato quinasa en portadoras
PQnoportadoras = PQ[!portadora] # Piruvato quinasa en no portadoras
```

a) Examinar y ejecutar el siguiente código y explicar qué realiza. Interpretar los resultados.

```
summary(CQportadoras)
summary(CQnoportadoras)
boxplot(CQportadoras,CQnoportadoras,names=c("portadoras","no portadoras"))
boxplot(log(CQportadoras),log(CQnoportadoras),
        names=c("portadoras","no portadoras"))
```

b) Examinar y ejecutar el siguiente código y explicar qué realiza. Interpretar los resultados.

```
plot(log(CQportadoras),log(PQportadoras),type="p",
     xlab="log(CQportadoras)",ylab="log(PQportadoras)")
cor(log(CQportadoras),log(PQportadoras))
```

Solución:

a) El siguiente código calcula algunos estadísticos descriptivos básicos (máximo, mínimo, cuartiles y media) del nivel de creatina quinasa en portadoras y no portadoras.

```
summary(CQportadoras)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
19.0	55.0	101.0	185.8	235.0	1288.0

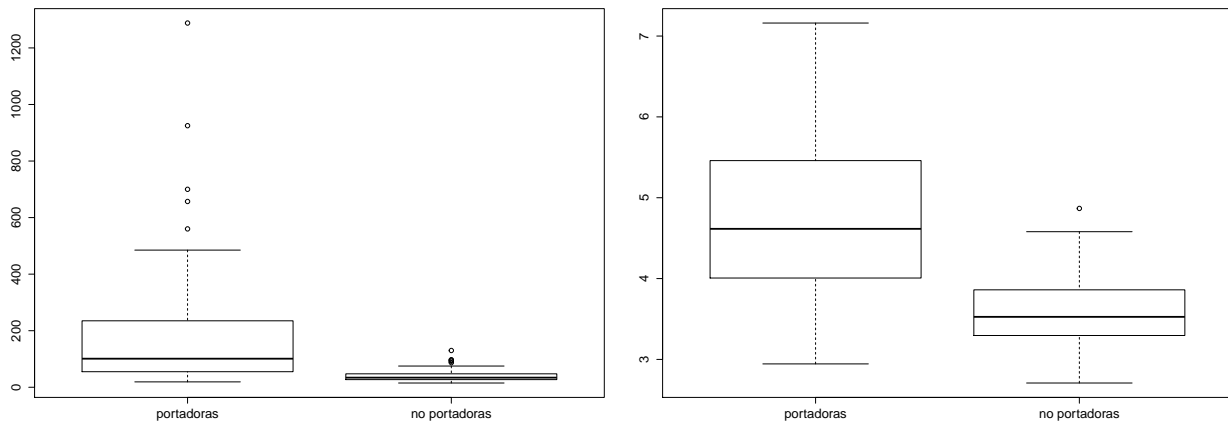
```
summary(CQnoportadoras)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.00	27.00	34.00	39.61	47.50	130.00

El primer diagrama de caja, de los dibujados a continuación, representa los niveles de creatina quinasa por separado en los grupos de portadoras y no portadoras. Observamos una gran asimetría hacia la derecha en ambas poblaciones, que además dificulta la comparación entre ambas. Al tomar logaritmos de las observaciones y representar nuevamente los datos (que es lo que hace el segundo diagrama de caja), se corrige esta asimetría y se facilita la comparación entre portadoras y no portadoras.

```
boxplot(CQportadoras,CQnoportadoras,names=c("portadoras","no portadoras"))
```

```
boxplot(log(CQportadoras),log(CQnoportadoras),  
        names=c("portadoras","no portadoras"))
```

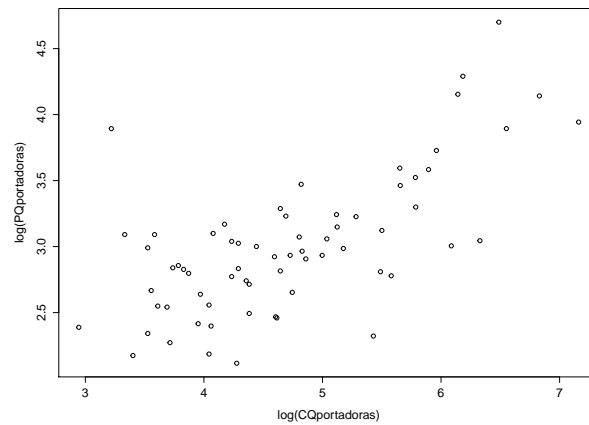


Los resultados muestran que los niveles de creatina quinasa en portadoras de la DMD son mucho más elevados que en las no portadoras. De hecho, en www.duchenne-spain.org aparece la siguiente información:

“Las siguientes pruebas se utilizan normalmente para confirmar un diagnóstico preliminar de Duchenne:

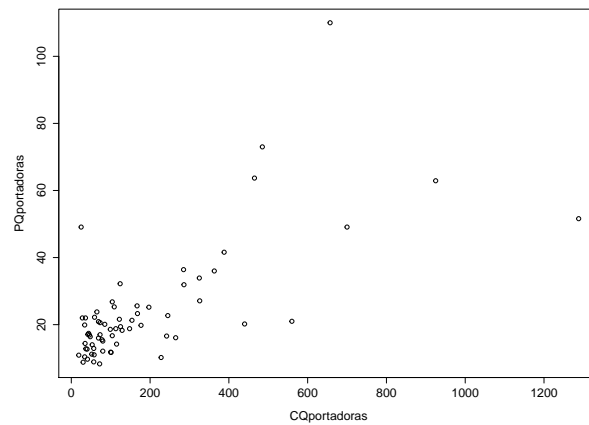
- *La creatina quinasa: fosfato de creatina quinasa (también conocida como CPK o CK) es una enzima que normalmente vive dentro de los músculos. Los chicos con Duchenne a menudo tienen niveles de CK de 10 a 100 veces el rango normal. Los niveles elevados de CK indican daño del músculo, a pesar de una alta CK no permite confirmar un diagnóstico de Duchenne.”*

b) Obtenemos el siguiente diagrama de dispersión del logaritmo de la piruvato quinasa en términos del logaritmo de la creatina quinasa:



y también el coeficiente de correlación lineal, 0.6692976. El diagrama y la correlación muestran una evidente relación lineal entre ambas variables. Quizá otro tipo de transformación daría mejor ajuste lineal. En cualquier caso, el gráfico es más explicativo que el de las variables sin transformar:

```
plot(CQportadoras,PQportadoras,type="p",
      xlab="CQportadoras",ylab="PQportadoras")
```



He probado tomando raíces cuadradas en lugar de logaritmos y sí que parece que mejora el ajuste y aumenta ligeramente el coeficiente de correlación hasta 0.7208222.