

# Entregable2\_A01283727

Russel Rosique

2022-09-07

## #El Problema

La contaminación por mercurio de peces en el agua dulce comestibles es una amenaza directa contra nuestra salud. Se llevó a cabo un estudio reciente en 53 lagos de Florida con el fin de examinar los factores que influían en el nivel de contaminación por mercurio.

```
df=read.csv("mercurio.csv")
```

```
#DESCRIPCION DE LOS DATOS | Columna | Descripción | |-----|-----|
| X1 | Número de identificación | X2 | Nombre del lago | X3 | Alcalinidad | X4 | PH |
X5 | Calcio (mg/l) | X6 | Clorofila (mg/l) | X7 | concentración media de mercurio (parte por millón) en el
tejido muscular del grupo de peces estudiados en cada lago
X8 | Número de peces estudiados en el lago | X9 | mínimo de la concentración de mercurio en cada
grupo de peces | X10 | máximo de la concentración de mercurio en cada grupo de peces | X11 | estimación
(mediante regresión) de la concentración de mercurio en el pez de 3 años (o promedio de mercurio cuando la
edad no está disponible) | X12 | indicador de la edad de los peces (0: jóvenes; 1: maduros) |
```

#Exploración Variables Dado que tenemos un total de 12 columnas, es importante que se realice esta exploración ya que podremos tener un vistazo preliminar sobre la relevancia de cada columna, hacia las respuestas que queremos obtener.

```
edad_peces = df$X12
table(edad_peces)
```

```
## edad_peces
## 0 1
## 10 43
```

Se puede observar que la mayoría de los peces estudiados en los lagos son de edad avanzada.

```
cantidad_peces = df$X8
table(cantidad_peces)
```

```
## cantidad_peces
## 4 5 6 7 8 10 11 12 13 14 24 36 40 43 44
## 1 1 4 2 1 11 3 20 3 2 1 1 1 1 1
```

Por otro lado, se puede observar que hay lagos en los que la muestra de los peces es demasiado pequeña, en varios únicamente se estudió un pez, mientras que en otros lagos se estudiaron cantidades más adecuadas.

```
media_mercurio = df$X7
table(media_mercurio)
```

```
## media_mercurio
## 0.04 0.05 0.1 0.15 0.16 0.17 0.18 0.19 0.21 0.25 0.27 0.28 0.34 0.4 0.41 0.43
## 2 1 1 1 1 1 1 3 1 1 3 1 4 1 1 1
```

```
## 0.44 0.48 0.49 0.5 0.52 0.56 0.59 0.63 0.65 0.71 0.73 0.77 0.81 0.83 0.84 0.86
## 1 2 2 2 1 2 1 1 1 1 1 1 1 1 1 1
## 0.87 0.94 0.98 1.08 1.1 1.16 1.2 1.23 1.33
## 1 1 1 1 1 1 1 1 1
```

Con este se logra obtener una primera observación del rango de la concentración de mercurio en los peces.

Por último, se ejecuta la siguiente instrucción para verificar que no existan columnas vacías, y podamos proceder con la exploración.

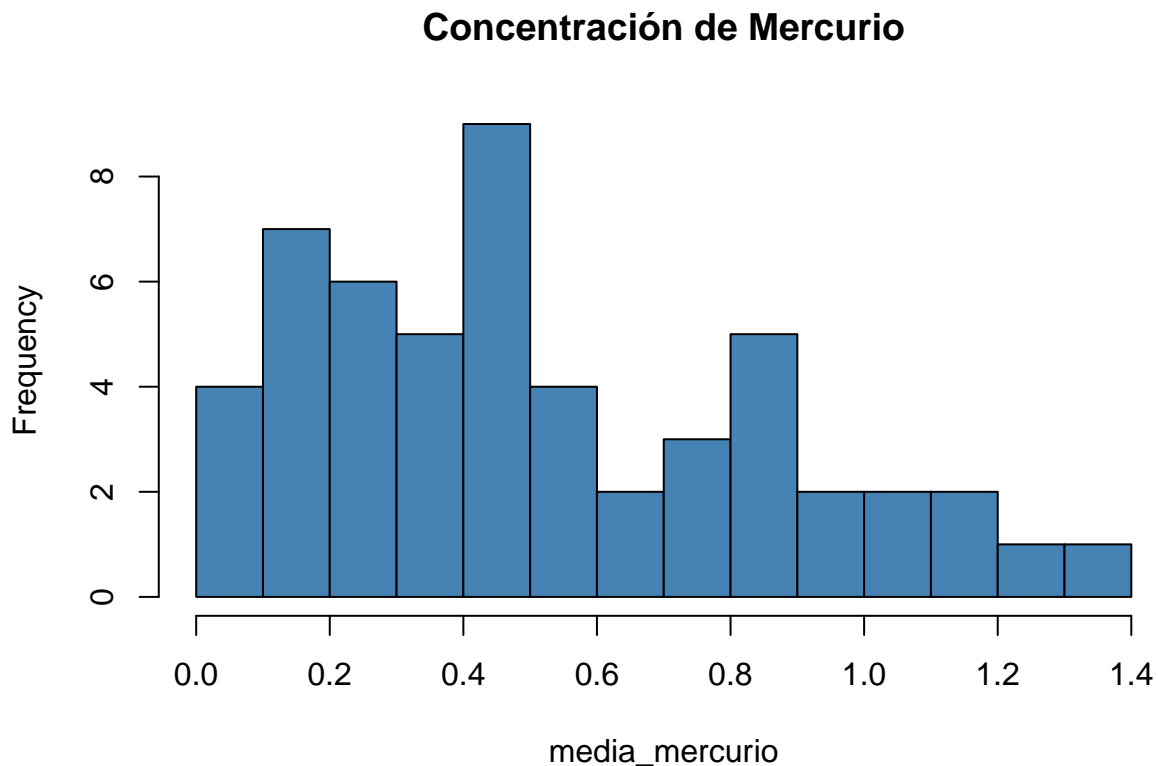
```
colSums(is.na(df))
```

```
## X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12
## 0 0 0 0 0 0 0 0 0 0 0 0
```

```
#EXPLORACION
```

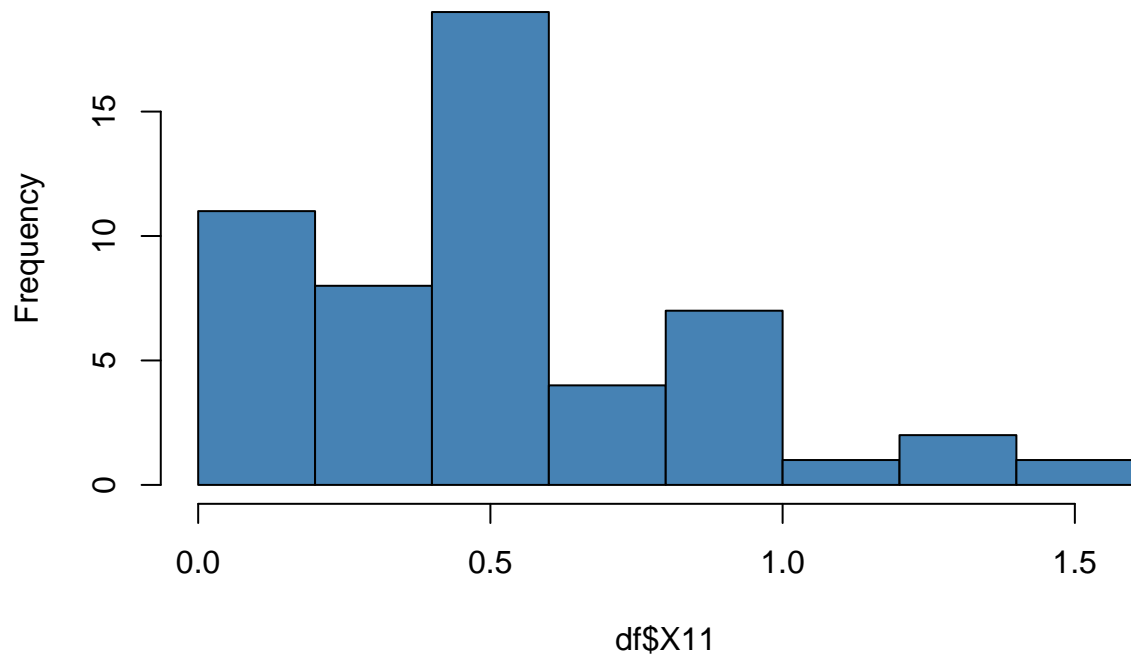
Primero que nada, buscaremos encontrar algunos datos generales sobre nuestra base de datos.

```
hist(media_mercurio, col='steelblue', main='Concentración de Mercurio', breaks = 12)
```



```
hist(df$X11, col='steelblue', main='Estimación de la concentración en peces de 3 años', breaks = 10)
```

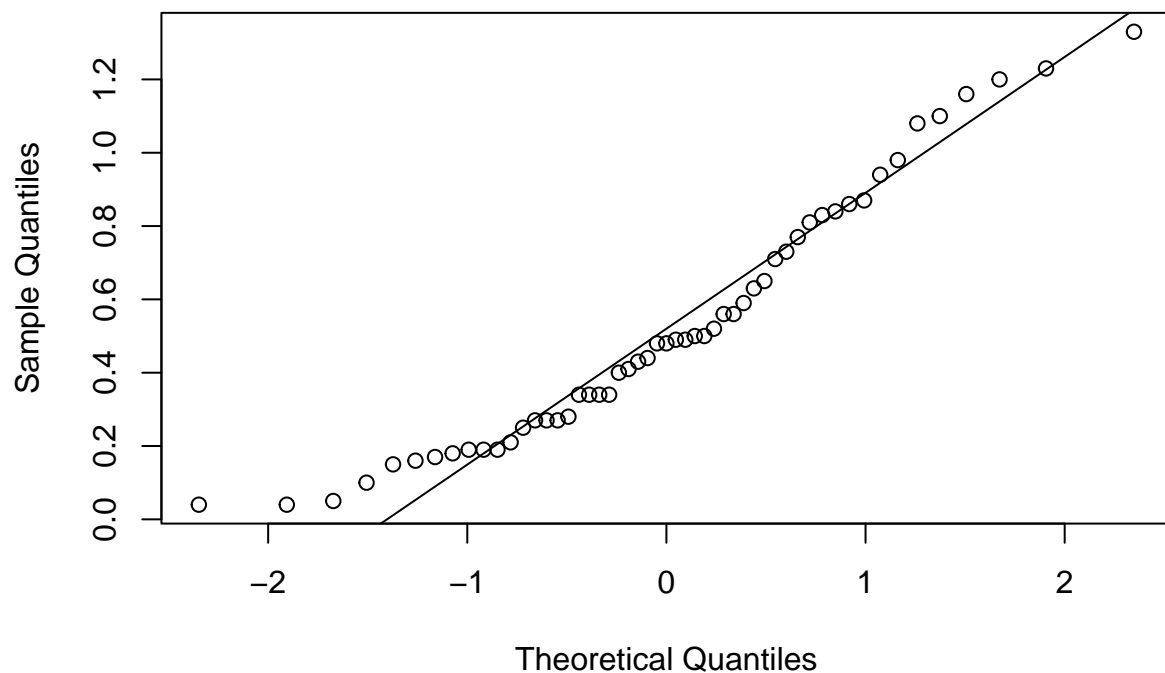
## Estimación de la concentración en peces de 3 años



Se puede observar a partir del histograma que los datos no se encuentran distribuidos de forma normal. Esto se podrá observar con mayor claridad en el siguiente diagrama.

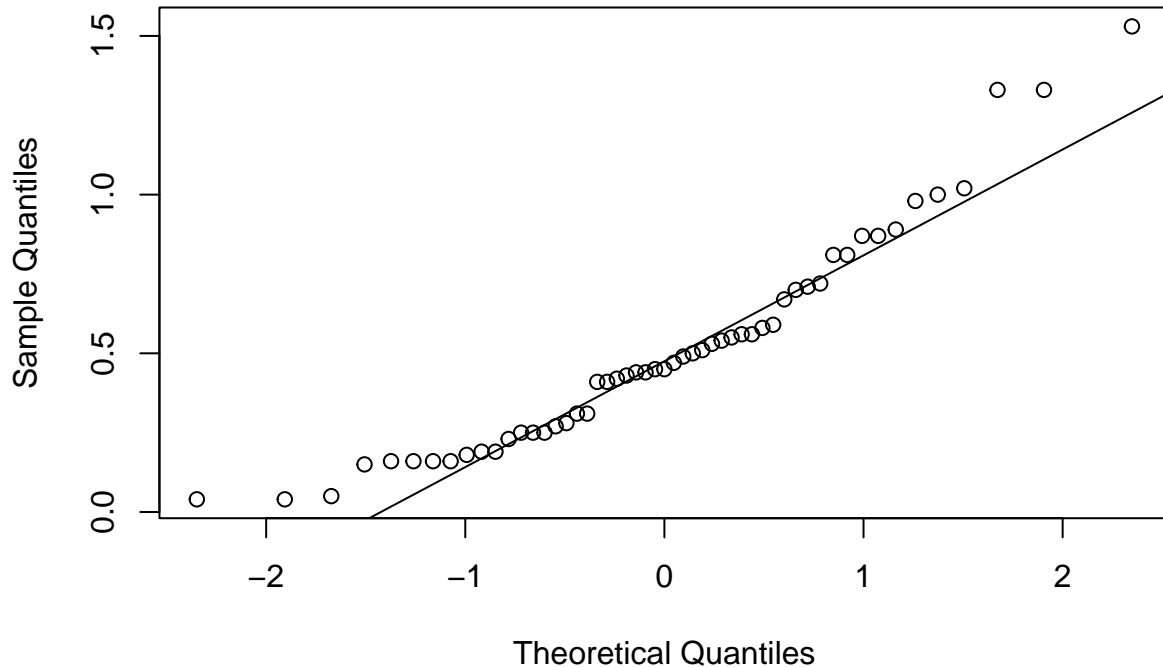
```
qqnorm(media_mercurio, main='QQ Plot Media de Concentración de Mercurio')  
qqline(media_mercurio)
```

## QQ Plot Media de Concentración de Mercurio



```
qqnorm(df$X11, main='QQ Plot Estimación de Concentración de Mercurio en el pez de 3 años')
qqline(df$X11)
```

## QQ Plot Estimación de Concentración de Mercurio en el pez de 3 años



Dado lo anterior, se puede observar que no existe normalidad aunque se eliminaran los datos atípicos.

#SESGO A continuación se encontrará la cantidad de sesgo que existe en el conjunto de datos.

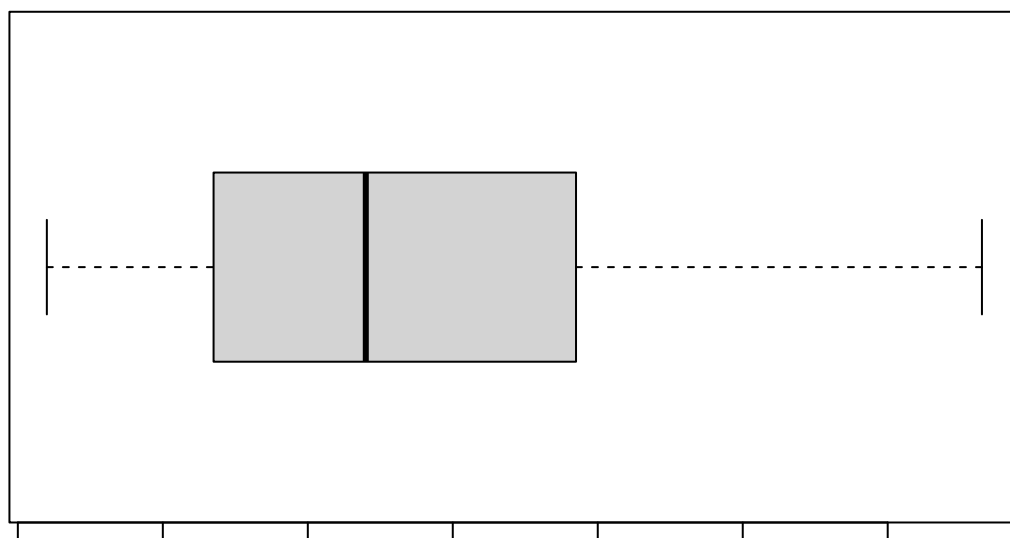
```
library(e1071)
skewness(media_mercurio)
```

```
## [1] 0.5986343
```

Dado que tenemos un sesgo positivo, se puede analizar que se tiene un sesgo hacia la izquierda.

#DATOS ATÍPICOS Al realizar la exploración de las variables, se pudo observar la posibilidad de datos atípicos en el conjunto de datos. Para verificar esto, se utilizará la herramienta de boxplot.

```
boxplot(media_mercurio, horizontal = TRUE)
```



0.0 0.2 0.4 0.6 0.8 1.0 1.2 Tras la visualización del boxplot, se verifica que no existe ningún dato atípico entre los datos.

```
iqr <- IQR(media_mercurio)
print(iqr)
```

```
## [1] 0.5
```