

# Mercurio

Nicolas Cardenas

2022-09-05

## PROBLEMA

### Overview

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. Las variables que se midieron se encuentran en mercurio.csv. Descargar mercurio.csv y su descripción es la siguiente:

X1 = número de indentificación X2 = nombre del lago X3 = alcalinidad (mg/l de carbonato de calcio) 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)

Es importante utilizar los datos que tenemos para poder ver que factores afectan el agua ya que el mercurio es una de las sustancias más dañinas para el cuerpo humano. Nos basaremos en los factores dentro del agua ya que son cosas que podemos alterar científicamente con químicos y reducir el nivel de mercurio de los lagos ya que los peces que habitan ahí son el alimento de los habitantes de Florida. [1]

### Resumen

Para este análisis se busca contestar esta pregunta: ¿Cuáles son los factores principales que afectan los niveles de mercurio en los peces? Usaremos varios modelos y técnicas aprendidas durante este curso junto con la información que se nos dio en forma de un Excel para hacer un análisis más profundo utilizando estas herramientas. Primero hicimos una limpieza de datos y un análisis por encima de todas las variables. Procedimos a utilizar correlación para ir empezando a contestar la pregunta, sin embargo, lo vimos a detalle con dos modelos: ANOVA y Regresión. Durante la realización de estos modelos encontramos mucha información interesante al respecto y terminamos con una transformación de la Regresión Lineal para tener una Regresión Exponencial con las variables que sí tenían efecto. Por lo tanto, no solo averiguamos cuáles eran los factores principales para los niveles de mercurio en los peces si no también obtuvimos un modelo preciso para estimar los niveles utilizando estas dos variables. Al final igualmente respaldamos nuestra investigación estadística con investigación científica para respaldar lo que encontramos.

```
## [1] "Informacion General de los datos"
```

```
##   alcalinidad      ph      calcio      clorofila
## Min.      : 1.20   Min.    :3.600   Min.    : 1.1   Min.    : 0.70
```

```
## 1st Qu.: 6.60 1st Qu.:5.800 1st Qu.: 3.3 1st Qu.: 4.60
## Median : 19.60 Median :6.800 Median :12.6 Median : 12.80
## Mean : 37.53 Mean :6.591 Mean :22.2 Mean : 23.12
## 3rd Qu.: 66.50 3rd Qu.:7.400 3rd Qu.:35.6 3rd Qu.: 24.70
## Max. :128.00 Max. :9.100 Max. :90.7 Max. :152.40
## mediaMercurio numPeces minMercurio maxMercurio
## Min. :0.0400 Min. : 4.00 Min. :0.0400 Min. :0.0600
## 1st Qu.:0.2700 1st Qu.:10.00 1st Qu.:0.0900 1st Qu.:0.4800
## Median :0.4800 Median :12.00 Median :0.2500 Median :0.8400
## Mean :0.5272 Mean :13.06 Mean :0.2798 Mean :0.8745
## 3rd Qu.:0.7700 3rd Qu.:12.00 3rd Qu.:0.3300 3rd Qu.:1.3300
## Max. :1.3300 Max. :44.00 Max. :0.9200 Max. :2.0400
## estimacion madurez
## Min. :0.0400 Min. :0.0000
## 1st Qu.:0.2500 1st Qu.:1.0000
## Median :0.4500 Median :1.0000
## Mean :0.5132 Mean :0.8113
## 3rd Qu.:0.7000 3rd Qu.:1.0000
## Max. :1.5300 Max. :1.0000

## [1] "Desviacion Estandar"

## alcalinidad ph calcio clorofila mediaMercurio
## 38.2035267 1.2884493 24.9325744 30.8163214 0.3410356
## numPeces minMercurio maxMercurio estimacion madurez
## 8.5606773 0.2264058 0.5220469 0.3387294 0.3949977
```

## LIMPIEZA

A primer vistazo, no vemos ninguna discrepancia en los datos, todos los minimos y maximos se observan en orden, no hay datos duplicados, no hay N/A, la alcalinidad esta dentro de los rangos validos de la escala pH, asi que podemos proceder con un analisis general de los datos para empezar.

Lo unico que vamos a hacer es cambiar el dummy de joven y maduro (0 y 1) a una variable categorica para poder correr el anova test en el futuro.

Conteo de valores nulos por columna, vemos que todos son 0 por lo cual no contamos con discrepancias.

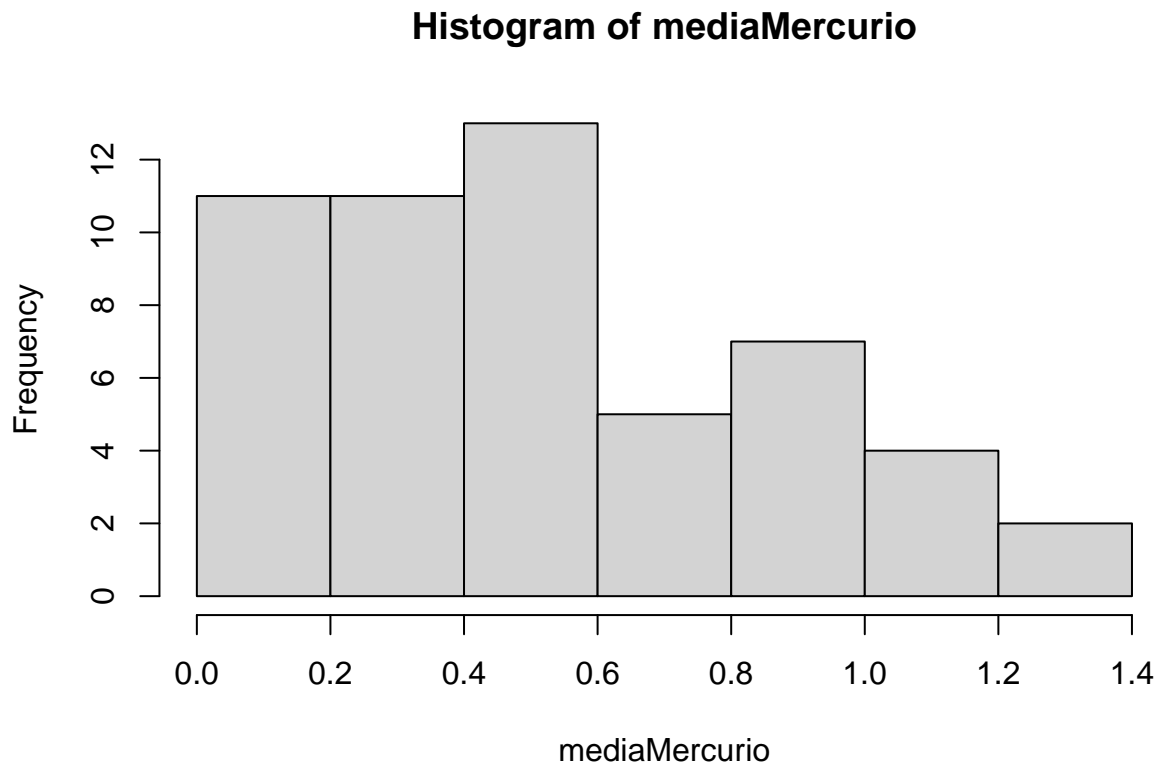
```
## alcalinidad ph calcio clorofila mediaMercurio
## 0 0 0 0 0
## numPeces minMercurio maxMercurio estimacion madurez
## 0 0 0 0 0
```

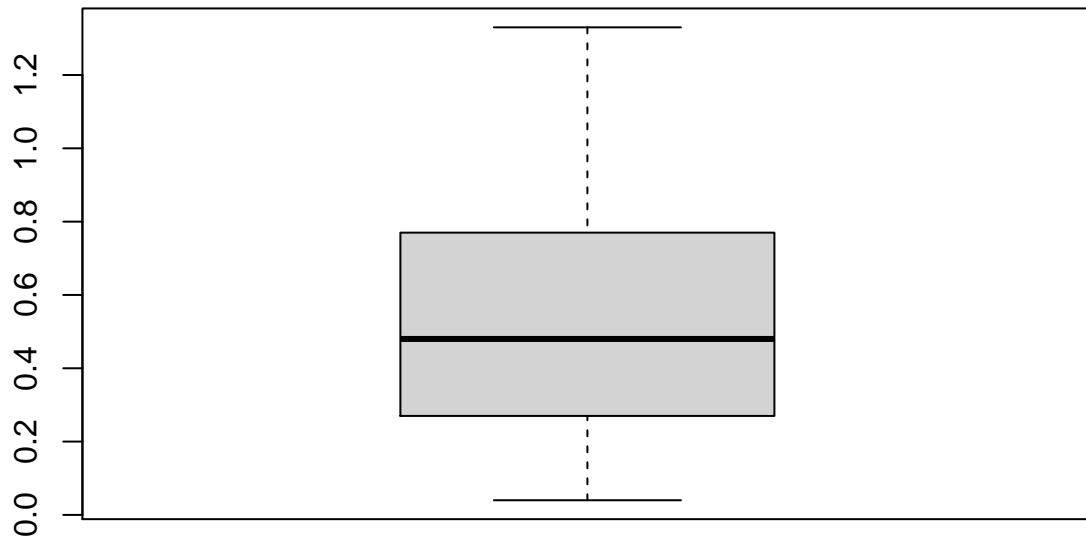
En cuanto a limpieza de datos atipicos/extremos los veremos en el analisis.

## ANALISIS

Para nuestra variable objetivo durante este analisis debemos primero determinar lo que nos pide el problema: medir los factores de de mercurio. Dentro de nuestra tabla tenemos varias variables para medir el nivel de mercurio: media, minimo, maximo, y la de estimacion. Primero que nada descartare la de estimacion ya que esa fue hecha con regresion lineal cuando no estaba la informacion en cuanto a la edad asi que esa ya contiene un grado de error, ademas utilizaron la media por lo tanto muestra alta correlacion con la media.

Y es por esto que creo que debemos enfocarnos en la media de mercurio en los peces ya que esta nos da una cifra pero tambien nos puede dar un intervalo de confianza donde podemos poner la media de la poblacion. Al principio estaba debatiendo si analizar el maximo del mercurio ya que este es el que importa, no queremos que ningun pez sobresalga de el maximo permitido para la salud. Sin embargo, como vemos que hay una correlacion alta entre la media y el maximo en este caso, creo que la media nos brindaria mas informacion y podemos hacer inferencias del maximo, asimismo, no queremos que datos atipicos nos alteren el maximo y tengamos una percepcion alterada.

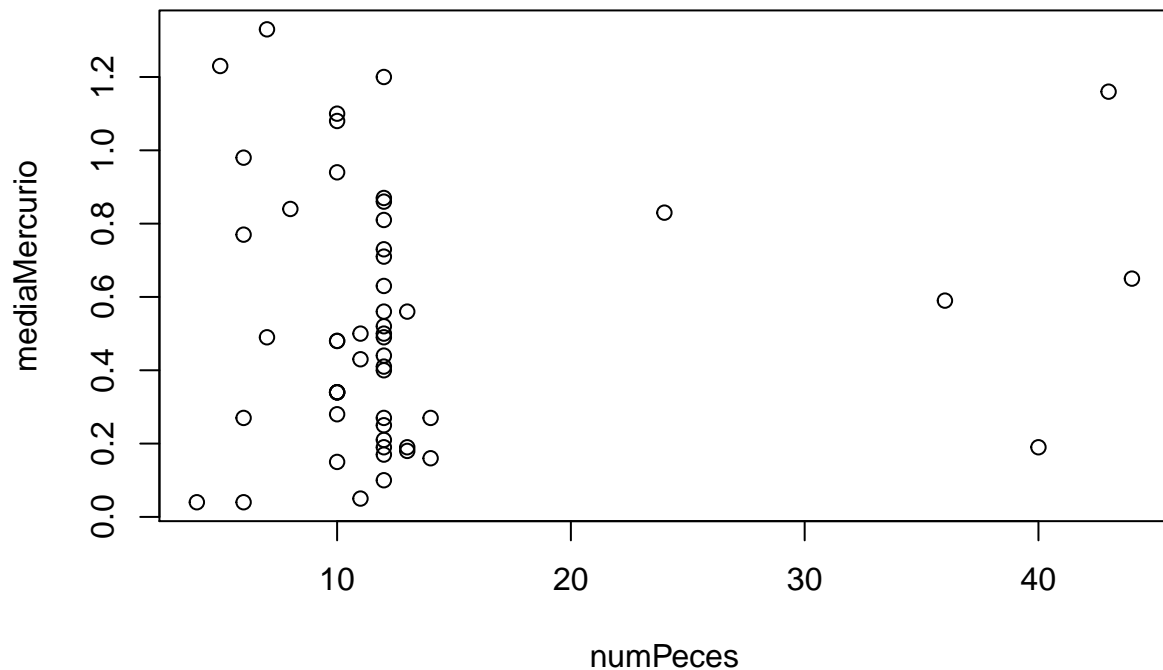




Vemos que en cuanto a la media de mercurio no hay datos atipicos, tenemos una distribucion bastante uniforme. No hay datos atipicos que tengamos que limpiar

## Medias

Ya que nuestra variable objetivo es la media del nivel del mercurio encontrado de los peces de ese lago, quisiera ver que tanto se relacion al numero de peces medidos. Lo primero que haremos es ver la grafica entre numPeces y la media de mercurio.



No vemos un ningun tipo de correlacion entre estas o algun patron que podamos analizar lo cual es bueno, significa que no porque haya mas peces y la media sea mas precisa o cambie drasticamente.

## Toxicidad

Antes de proceder a analizar cuales seran los factores principales en cuanto a toxicidad de mercurio en el agua, hay que observar la media en cuanto todos los lagos para ver si realmente es un problema que hay en Florida. Para esto utilizaremos los reglamentos internacionales CE 1881/2006 y Codex Standard 193-1995 que indican que la concentración promedio de mercurio en productos de la pesca no debe superar los 0.5 mg de Hg/kg.

$H_0$  = La media de la media de mercurio en los peces en lagos de Florida es  $< 0.5$   $H_1$  = La media de la media de mercurio en los peces en lagos de Florida es  $> 0.5$

Para esta prueba de hipotesis utilizaremos un alpha de 0.05

Para la distribucion opte por utilizar la t de student debido a que tenemos mas de 30 datos y no conocemos sigma.

Primero debemos encontrar  $t_0$ .

```
## [1] -1.675285
```

Esta prueba es solo cola hacia la derecha ya que solo se rechaza si es mayor.

Por lo tanto rechazaremos si: \* Valor p es menor a 0.05 \* Si  $|t \text{ estrella}| > |t_0$

```
## [1] 0.5799957
```

```
## [1] 1.435529
```

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

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

Vemos que el valor  $p$  no es menor que  $\alpha$  y que  $t_{estrella}$  no es menor que el valor absoluto de  $t$  estrella no es tampoco entonces no rechazamos la hipótesis nula y que nuestra media si es menos de 0.5, por lo cual no hay cantidades de mercurio tóxicas en promedio pero si se busca disminuir debemos encontrar que factores son los que aumentan esta toxicidad.

Otra cosa que podríamos analizar que también considero que podría ser extremadamente útil sería sacar el intervalo de la media para cada lago, para saber específicamente por lago que tan tóxico es y no solo en promedio de Florida.

## Correlacion

```
## Warning: package 'Hmisc' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Warning: package 'Formula' was built under R version 4.0.3
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      format.pval, units
```

```
##          alcalinidad    ph calcio clorofila mediaMercurio numPeces
## alcalinidad          1.00  0.72  0.83      0.48        -0.59    0.01
## ph                   0.72  1.00  0.58      0.61        -0.58   -0.02
## calcio                0.83  0.58  1.00      0.41        -0.40   -0.09
## clorofila             0.48  0.61  0.41      1.00        -0.49   -0.01
## mediaMercurio        -0.59 -0.58 -0.40     -0.49         1.00    0.08
## numPeces              0.01 -0.02 -0.09     -0.01         0.08    1.00
## minMercurio          -0.53 -0.54 -0.33     -0.40         0.93   -0.08
## maxMercurio          -0.60 -0.55 -0.41     -0.48         0.92    0.16
## estimacion           -0.63 -0.61 -0.46     -0.51         0.96    0.03
## madurez              -0.09  0.04  0.00     -0.28         0.11    0.21
##          minMercurio maxMercurio estimacion madurez
## alcalinidad        -0.53        -0.60      -0.63   -0.09
## ph                  -0.54        -0.55      -0.61    0.04
```

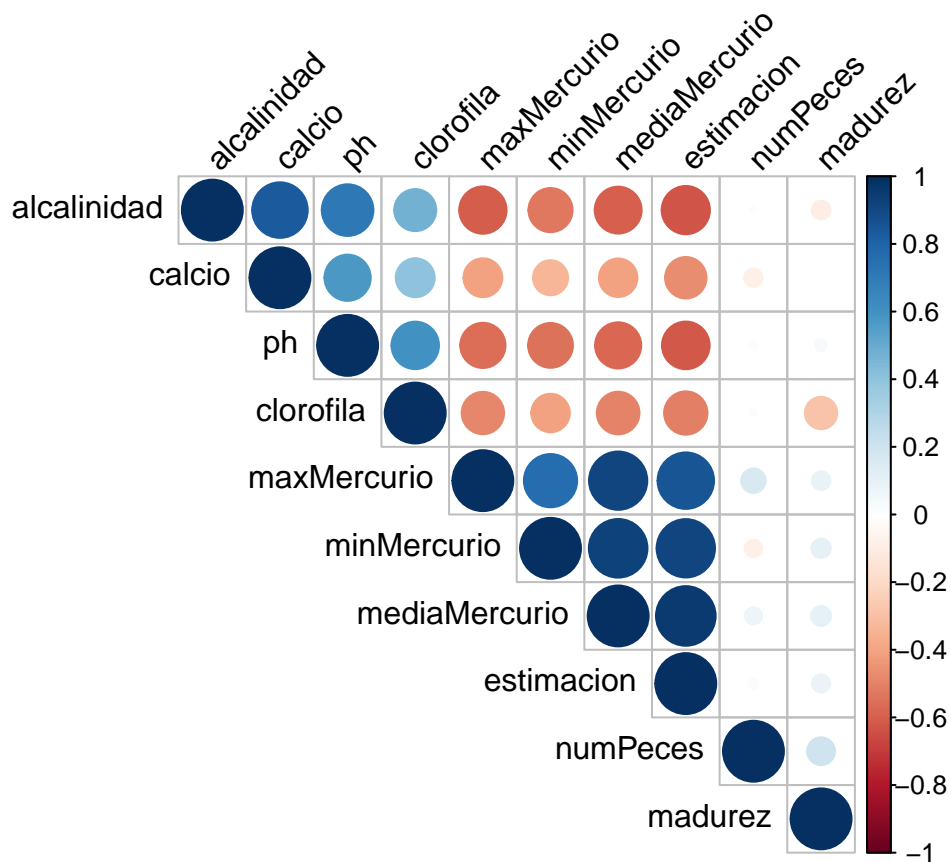
```

## calcio          -0.33      -0.41      -0.46      0.00
## clorofila       -0.40      -0.48      -0.51     -0.28
## mediaMercurio   0.93       0.92       0.96      0.11
## numPeces       -0.08       0.16       0.03      0.21
## minMercurio     1.00       0.77       0.92      0.10
## maxMercurio     0.77       1.00       0.86      0.09
## estimacion      0.92       0.86       1.00      0.09
## madurez         0.10       0.09       0.09      1.00
##
## n= 53
##
##
## P
##          alcalinidad ph      calcio clorofila mediaMercurio numPeces
## alcalinidad      0.0000 0.0000 0.0003 0.0000 0.0000 0.9417
## ph               0.0000      0.0000 0.0000 0.0000 0.0000 0.8948
## calcio           0.0000 0.0000      0.0023 0.0029 0.0029 0.5245
## clorofila        0.0003 0.0000 0.0023      0.0002 0.0002 0.9331
## mediaMercurio    0.0000 0.0000 0.0029 0.0002      0.0002 0.5737
## numPeces         0.9417 0.8948 0.5245 0.9331 0.5737      0.5611
## minMercurio      0.0000 0.0000 0.0150 0.0030 0.0000 0.5611
## maxMercurio      0.0000 0.0000 0.0024 0.0002 0.0000 0.2492
## estimacion       0.0000 0.0000 0.0005 0.0001 0.0000 0.8545
## madurez          0.4989 0.7870 0.9880 0.0400 0.4383 0.1351
##
##          minMercurio maxMercurio estimacion madurez
## alcalinidad 0.0000 0.0000 0.0000 0.4989
## ph          0.0000 0.0000 0.0000 0.7870
## calcio      0.0150 0.0024 0.0005 0.9880
## clorofila   0.0030 0.0002 0.0001 0.0400
## mediaMercurio 0.0000 0.0000 0.0000 0.4383
## numPeces     0.5611 0.2492 0.8545 0.1351
## minMercurio      0.0000 0.0000 0.4733
## maxMercurio      0.0000 0.0000 0.5043
## estimacion       0.0000 0.0000 0.5243
## madurez          0.4733 0.5043 0.5243

##          a p clc clr mdM n mnM mxM e md
## alcalinidad 1
## ph          , 1
## calcio      + . 1
## clorofila   . , . 1
## mediaMercurio . . . . 1
## numPeces      . . . . 1
## minMercurio . . . . * 1
## maxMercurio , . . . * , 1
## estimacion , , . . B * + 1
## madurez      . . . . . 1
## attr("legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

## corrplot 0.92 loaded

```



```
## Warning: package 'PerformanceAnalytics' was built under R version 4.0.5
```

```
## Loading required package: xts
```

```
## Warning: package 'xts' was built under R version 4.0.5
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
##
```

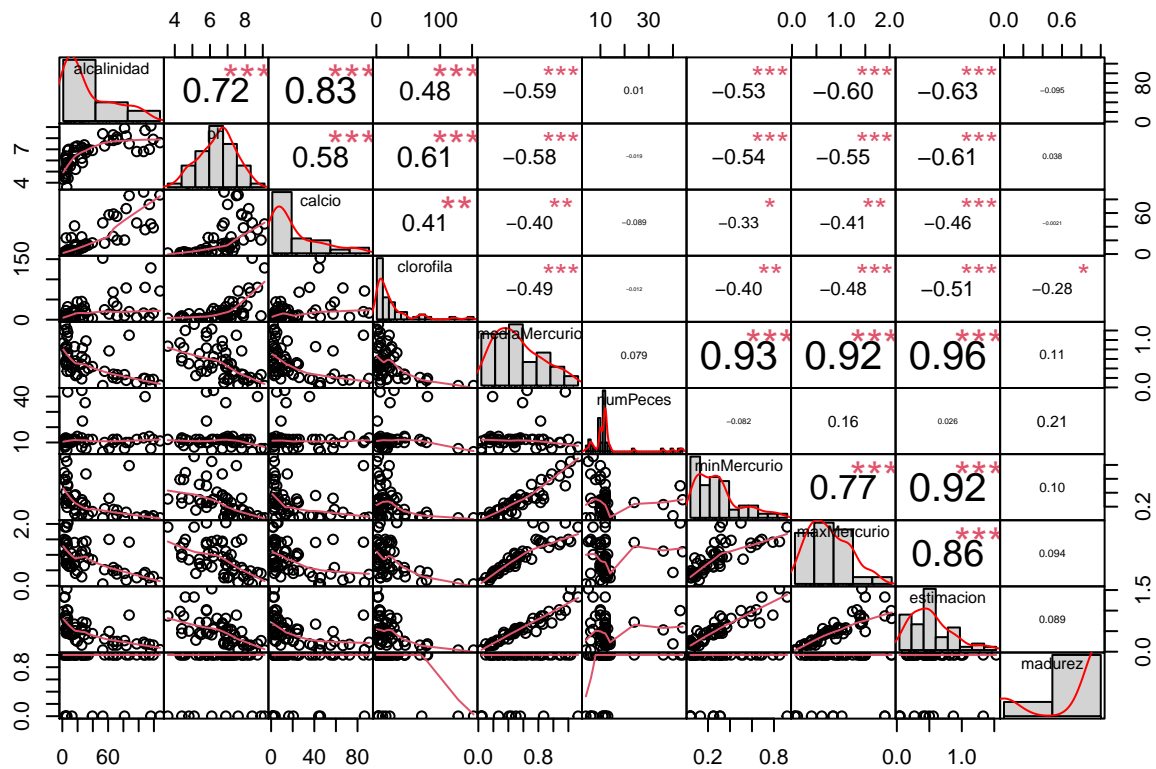
```
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
```

```
##
```

```
## legend
```





Como podemos ver hay alta correlacion entre maxMercurio,minMercurio, mediaMercurio y tambien la estimacion lo cual no es sorprendente debido a que estan van de la mano. Lo que si es interesante que el numero de peces tambien esta altamente correlacionado con esta cifra. Otra de las correlaciones medias-altas que observamos es el calcio con el ph y la clorofila con el ph.

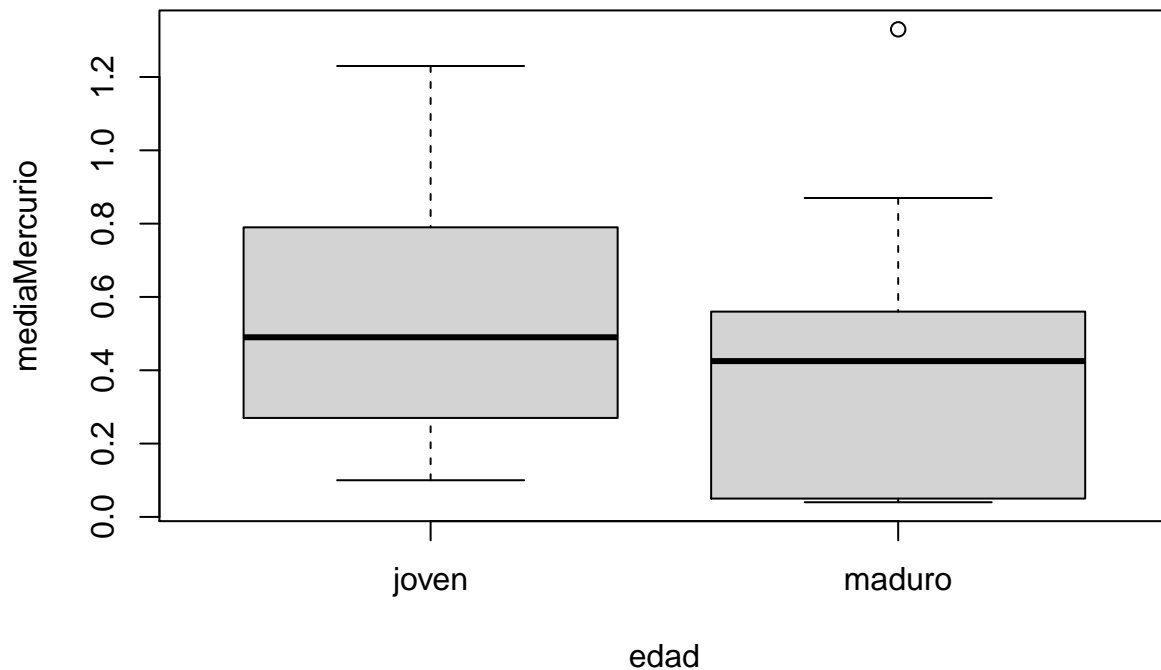
## Anova

Primero que nada haremos una prueba de anova con la madurez de los peces, y como ya lo mencionamos antes usaremos el promedio de mercurio en los peces como variable dependiente. Haremos esto primero ya que es nuestra unica variable categorica.

H0: No hay efecto debido a la edad en la media de los mercurio en los peces H1: Si hay efecto debido a la edad en la media de los mercurio en los peces

Criterio de Rechazo  $P(F^* > F_0) < 0.05$

Primero veremos un boxplot representando la edad de los peces (joven,adulto)



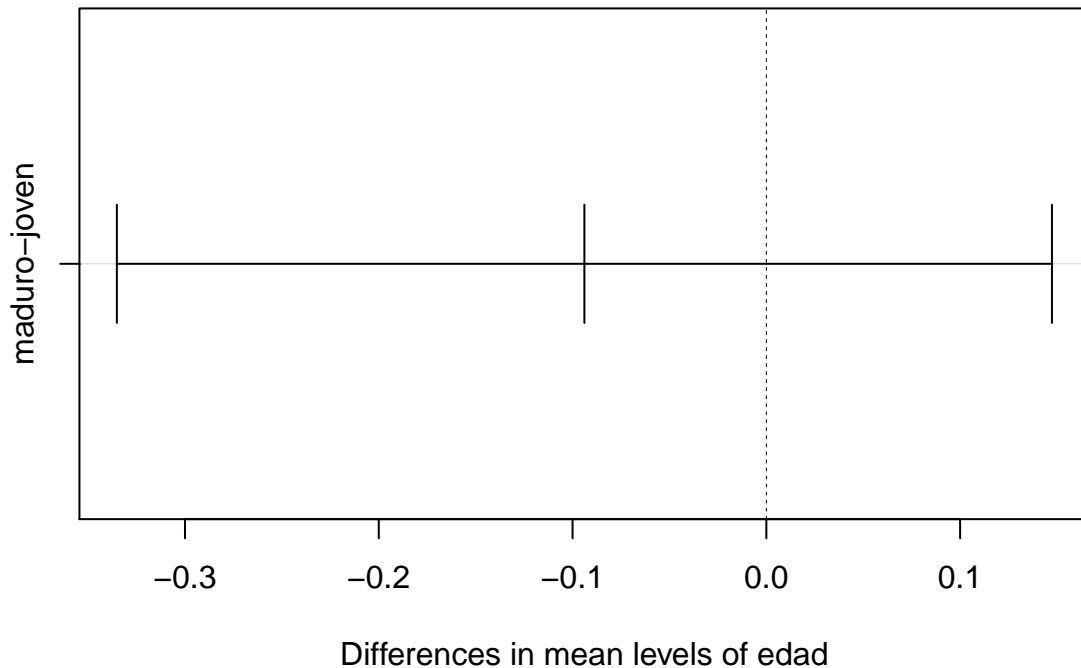
Como podemos ver la media es similar en ambos, sin embargo se ven diferentes los diagramas en terminos de rangos intercuartiles, asi que procederemos con el ANOVA test.

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## edad      1  0.072  0.07151    0.61  0.438
## Residuals 51  5.976  0.11718
```

Como podemos ver, vemos que no hay efecto.  $Pr > F$  es demasiado alta por lo tanto no rechazamos la hipotesis nula y concluimos que NO hay efecto. La edad es la unica variable categorica que podemos analizar con ANOVA. Por lo tanto utilizaremos otros metodos para analizar las otras variables.

Ya sabemos que hay efecto pero igual para confirmar procederemos con un Turkey Test:

## 95% family-wise confidence level



El intervalo pasa por cero, lo cual me indica que son iguales y como lo mencionamos anteriormente no hay efecto.

## Regresion Lineal

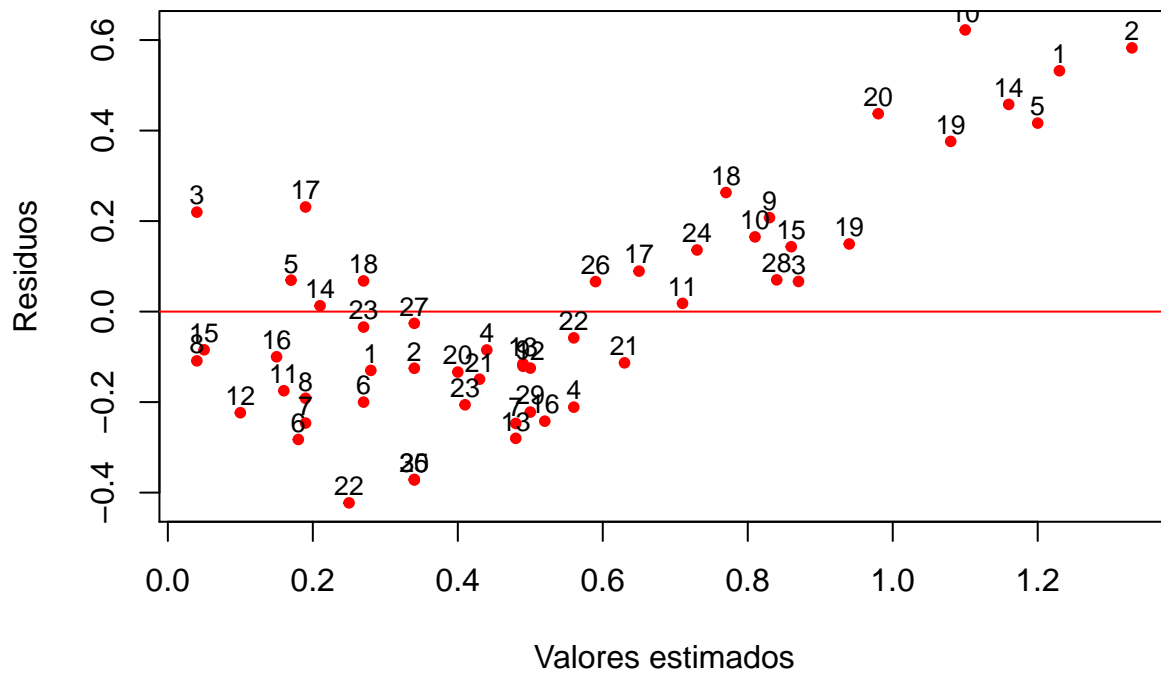
Utilizando nuestra matriz anterior, haremos una regresion lineal para ver cuales variables tienen mayor efecto en nuestra variable dependiente (media del mercurio)

Como vimos en el diagrama de correlacion, observamos que el numPeces, maxMercurio, minMercurio, mediaMercurio y tambien la son los que tienen correlacion mas alta. Sin embargo, el maximo y minimo de las medidas del mercurio no pueden ser realizadas para observar comportamiento diferente de la media ya que esta utilizaria datos que fueran parte del maximo y del minimo. Lo que se busca con esta investigacion es poder predecir y disminuir la media del mercurio con los factores que tenemos en el agua, determinar cuales son los factores y apartir de ahi tomar accion, por lo tanto no agregare las variables que forman parte de las medidas del mercurio ya que ya tenemos la media del mercurio como variable dependiente. Agregare ph, clorofila y alcalinidad ya que no fueron tan fuertemente correlacionados pero mostraron correlacion en todas las medidas del mercurio decente (0.5-0.6). Calcio no mostro mucha correlacion y aunque el numero de peces mostro mucha correlacion, no es factor en si afecta incrementa los niveles del mercurio ya que no es un factor ambiental, pero creo que si esta afectando nuestra media, analizaremos eso despues. Por ahora, meteremos todos los factores relevantes en cuanto al agua en el modelo.

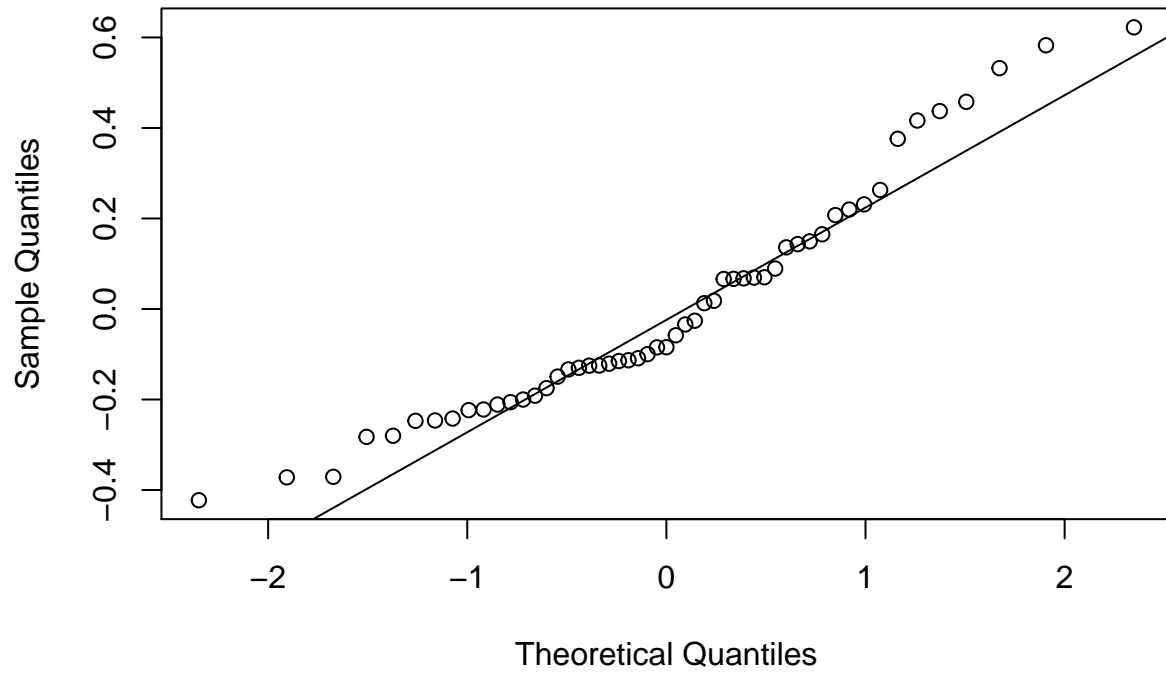
Antes de empezar a realizar los modelos me gustaria tener una funcion para tambien analizar los residuos:

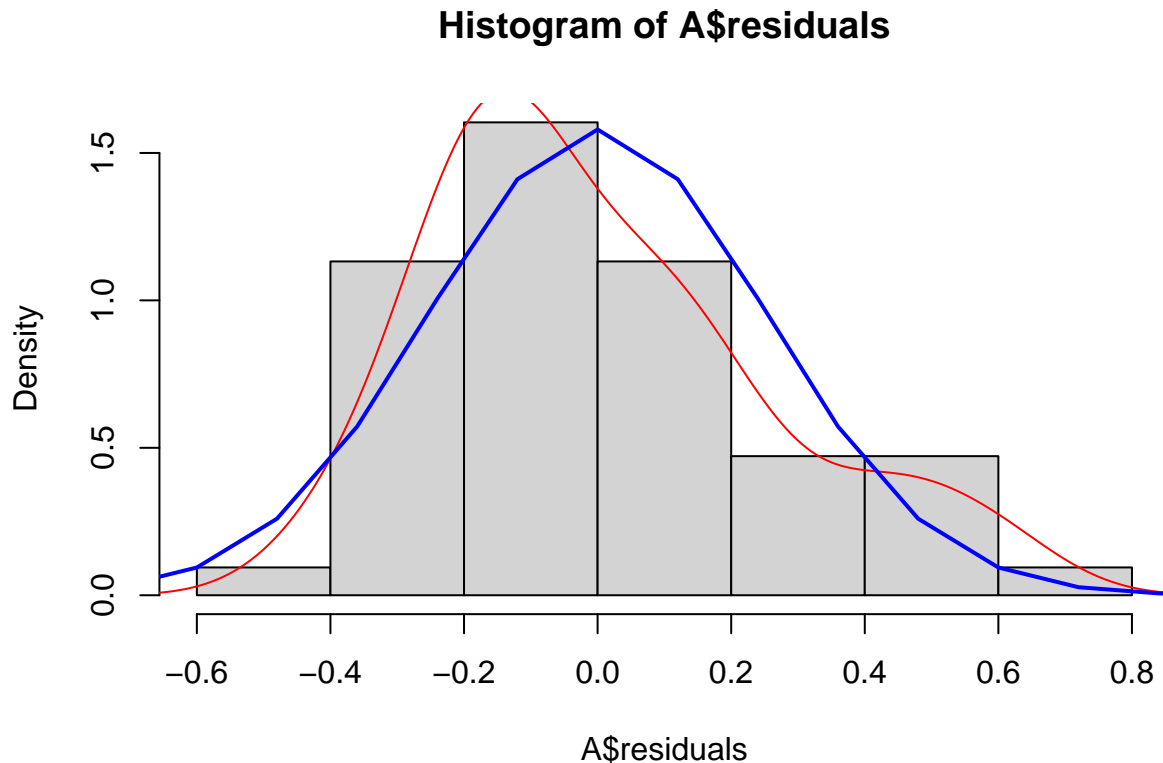
```
##  
## Call:  
## lm(formula = mediaMercurio ~ ph + alcalinidad + clorofila + calcio)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42260 -0.19155 -0.08438  0.14334  0.62234
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.004440   0.257561   3.900 0.000299 ***
## ph          -0.046709   0.045329  -1.030 0.307968
## alcalinidad -0.005503   0.002028  -2.713 0.009224 **
## clorofila   -0.002361   0.001497  -1.577 0.121257
## calcio       0.004129   0.002648   1.559 0.125484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2629 on 48 degrees of freedom
## Multiple R-squared:  0.4515, Adjusted R-squared:  0.4058
## F-statistic: 9.879 on 4 and 48 DF,  p-value: 6.499e-06
```



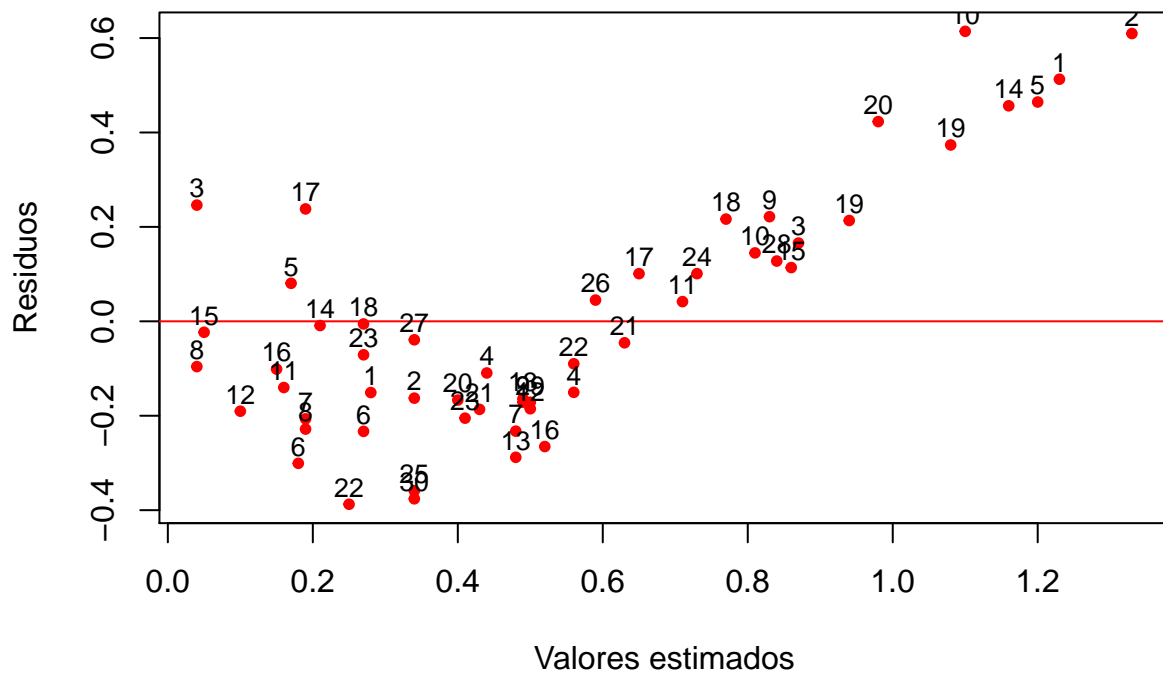
Normal Q-Q Plot



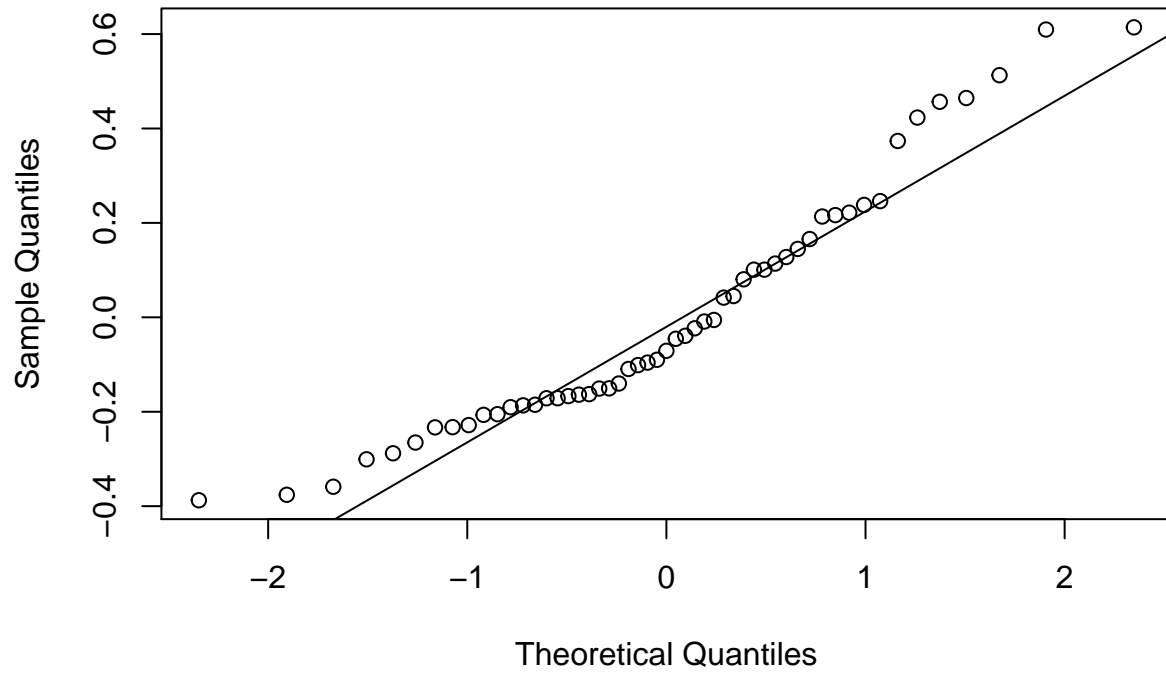


Como podemos ver, los únicos que sí tienen efecto en la media es la alcalinidad, sin embargo tenemos 2 variables que también tienen un valor p bajo, no abajo de nuestra alfa, pero considero que son importantes mantener. Por ahora quitaremos solo el que tuvo el valor p más alto que fue el pH y cambiaremos el modelo para ver cómo interactúa ahora. En cuanto a los residuos podemos ver que no se distribuyen normalmente, lo cual indica un mal modelo. Trataremos ahora de modelar con solo las variables que fueron lo suficientemente significativas:

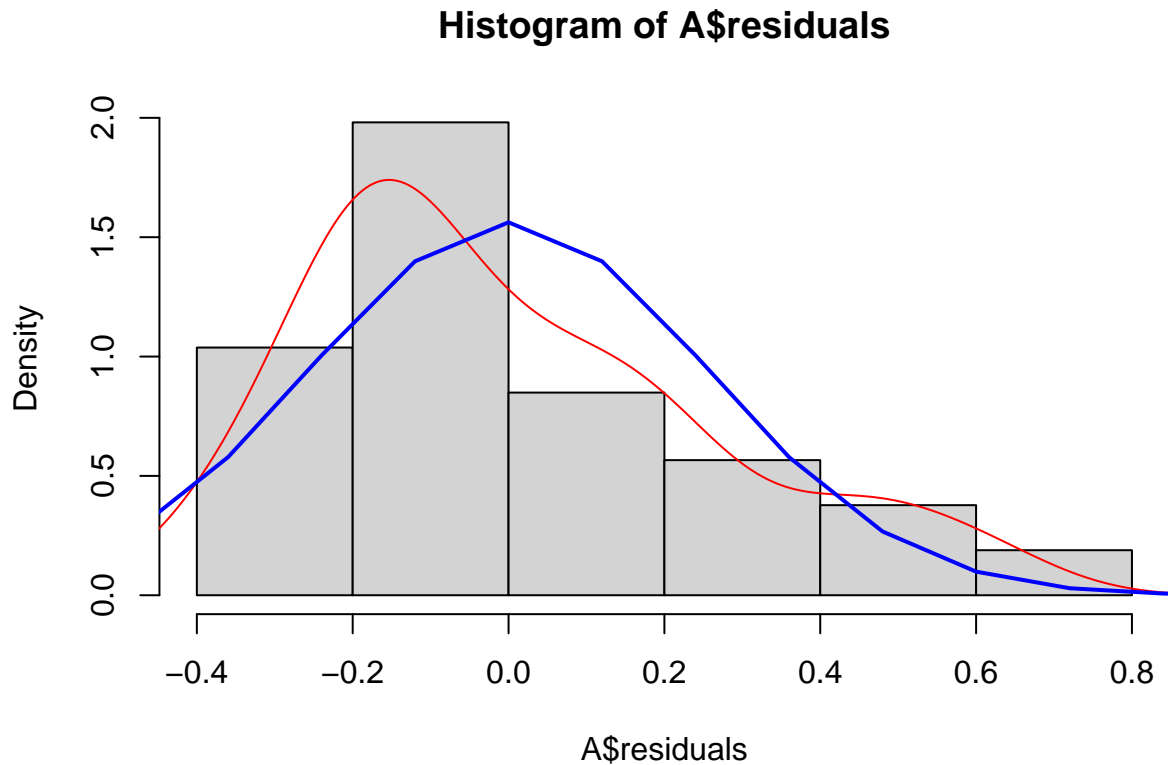
```
##
## Call:
## lm(formula = mediaMercurio ~ clorofila + alcalinidad + calcio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38746 -0.18520 -0.07092  0.14490  0.61422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.744583   0.052401  14.209  < 2e-16 ***
## clorofila    -0.003035   0.001348  -2.252  0.028862 *
## alcalinidad  -0.006487   0.001790  -3.624  0.000689 ***
## calcio       0.004333   0.002642   1.640  0.107445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.263 on 49 degrees of freedom
## Multiple R-squared:  0.4394, Adjusted R-squared:  0.4051
## F-statistic: 12.8 on 3 and 49 DF, p-value: 2.676e-06
```



Normal Q-Q Plot

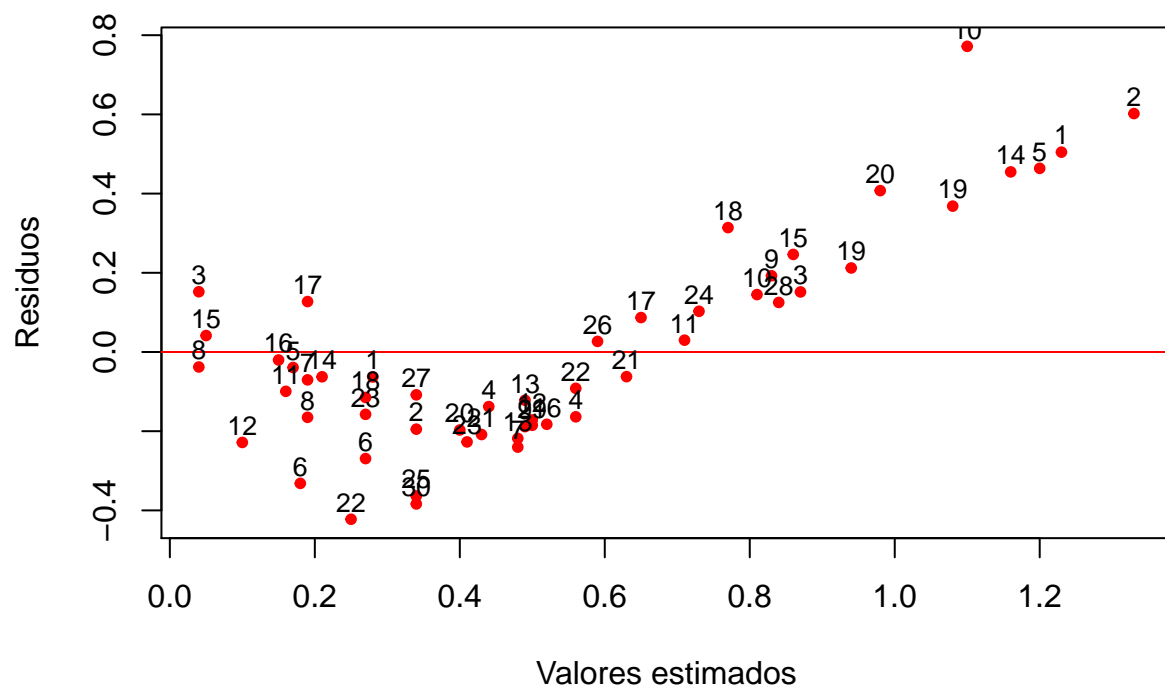




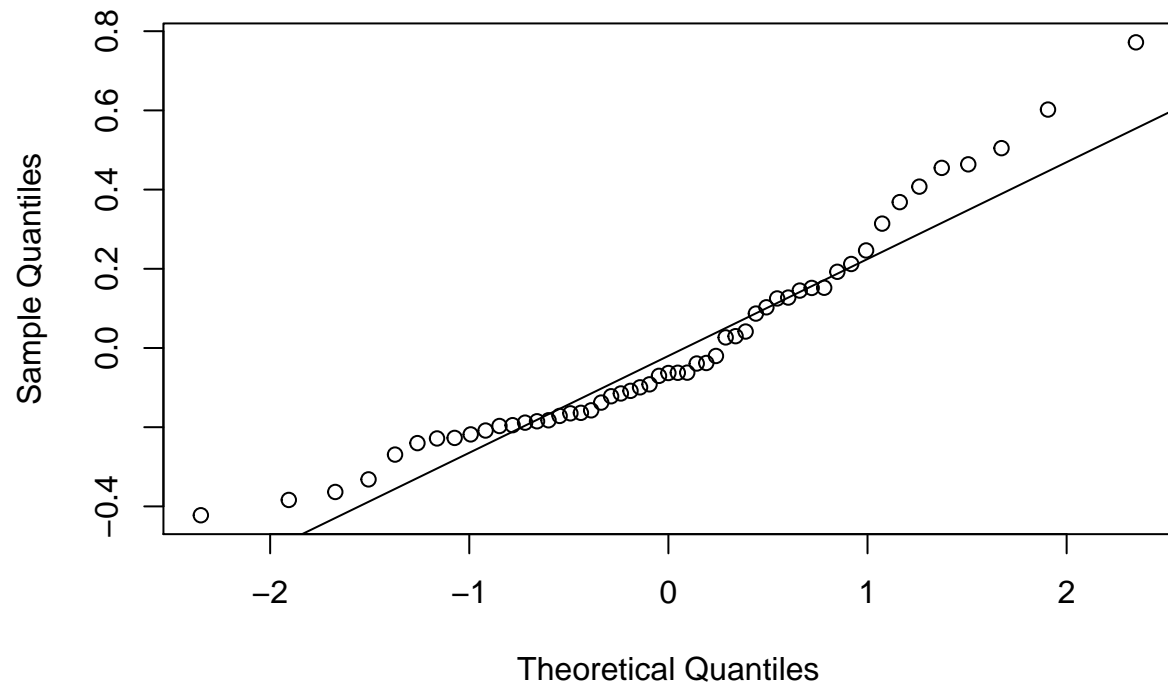


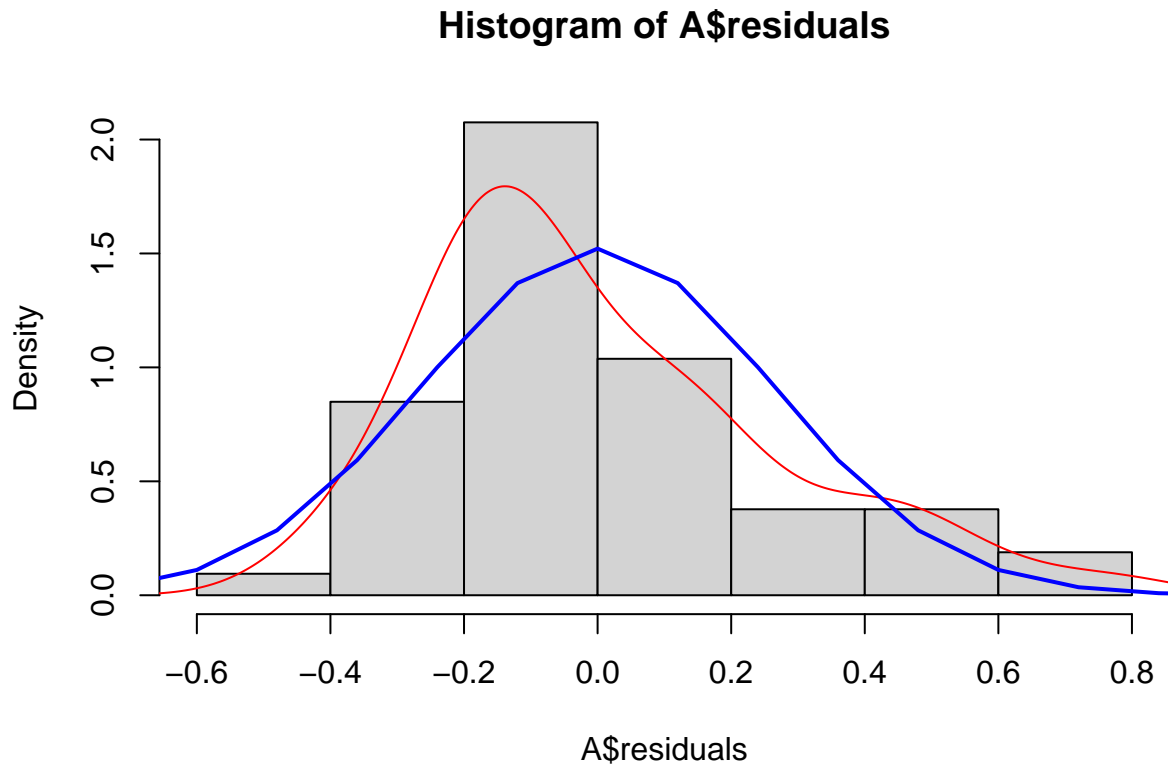
Finalmente modelamos con solo nuestras variables que fueron significantes para obtener un nuevo modelo:

```
##
## Call:
## lm(formula = mediaMercurio ~ clorofila + alcalinidad)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42232 -0.18500 -0.06314  0.14502  0.77185
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.751936   0.053084  14.165 < 2e-16 ***
## clorofila    -0.002979   0.001370  -2.174 0.034433 *
## alcalinidad  -0.004154   0.001105  -3.760 0.000446 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2675 on 50 degrees of freedom
## Multiple R-squared:  0.4086, Adjusted R-squared:  0.385
## F-statistic: 17.27 on 2 and 50 DF,  p-value: 1.979e-06
```



Normal Q-Q Plot

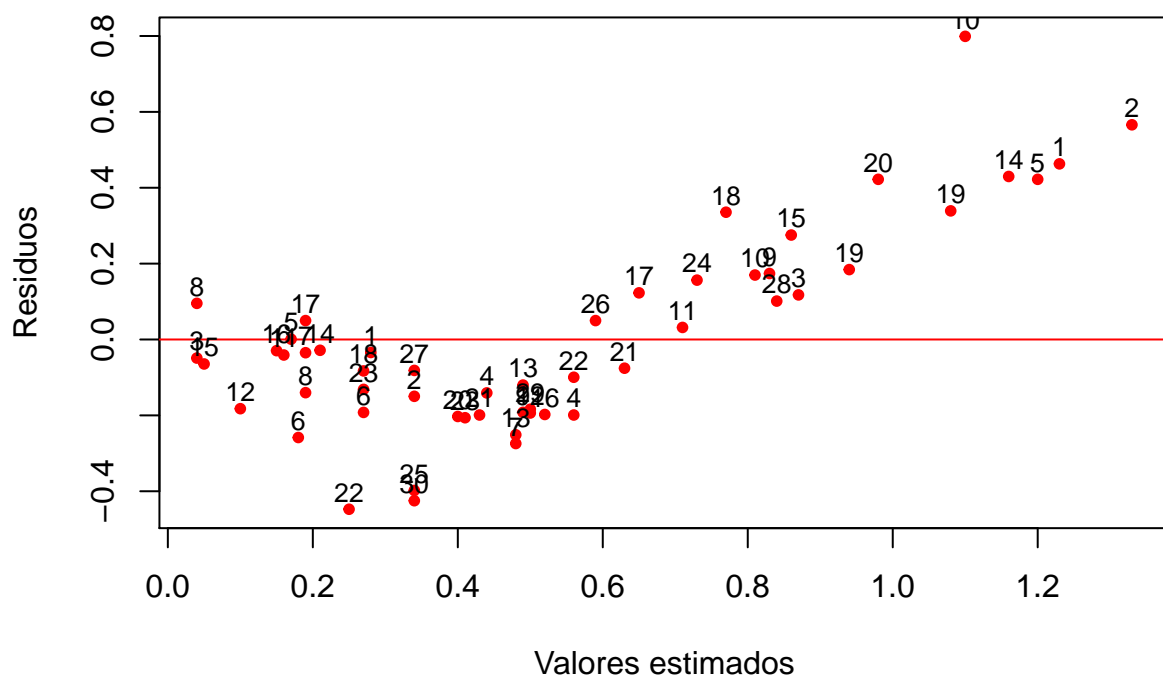




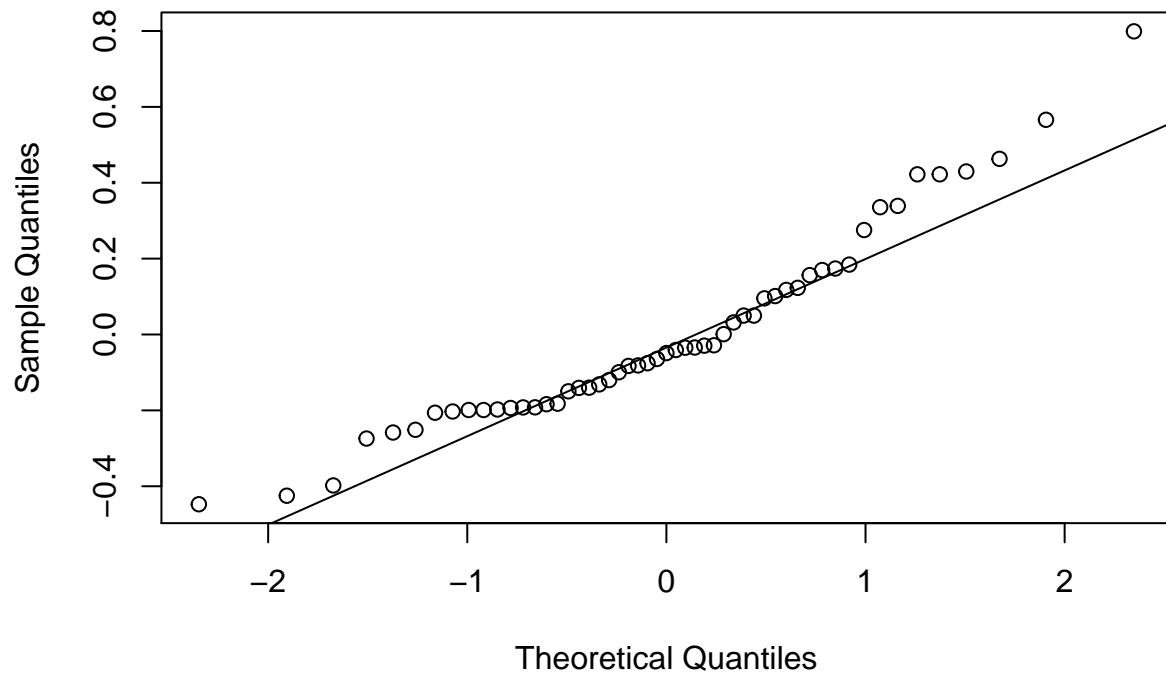
Como podemos ver el valor p para clorofila y alcalinidad son menores que nuestra alfa, indicando que si hay efecto, sin embargo, este modelo no lo esta modelando apropiadamente ya que tenemos una  $R^2$  mas baja ahora.

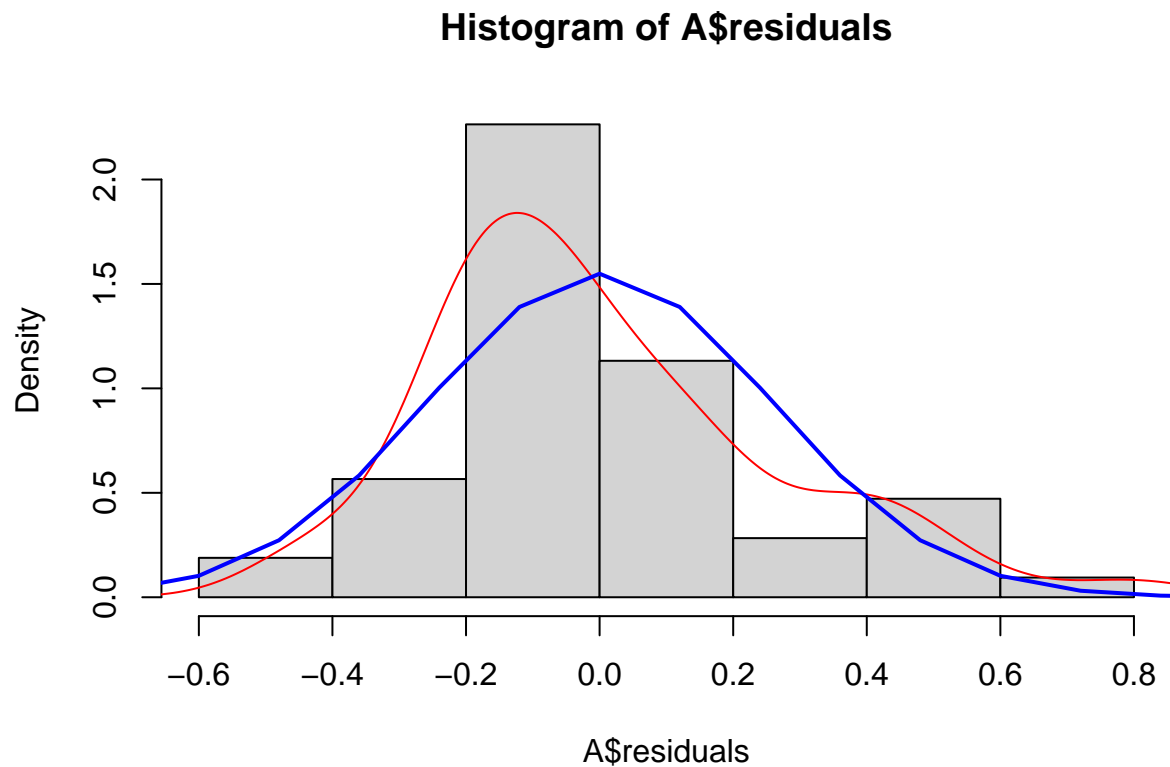
Antes de proceder, valdria la pena evaluar el modelo con clorofila y alcalinidad interactuando, para observar si es un mejor modelo:

```
##
## Call:
## lm(formula = mediaMercurio ~ clorofila * alcalinidad)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44742 -0.19201 -0.04896  0.12296  0.79908
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.020e-01  6.448e-02  12.439  < 2e-16 ***
## clorofila      -6.310e-03  2.823e-03  -2.235  0.030012 *
## alcalinidad    -5.221e-03  1.352e-03  -3.860  0.000332 ***
## clorofila:alcalinidad  4.718e-05  3.505e-05   1.346  0.184453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2653 on 49 degrees of freedom
## Multiple R-squared:  0.4297, Adjusted R-squared:  0.3948
## F-statistic: 12.31 on 3 and 49 DF,  p-value: 4.029e-06
```



Normal Q-Q Plot



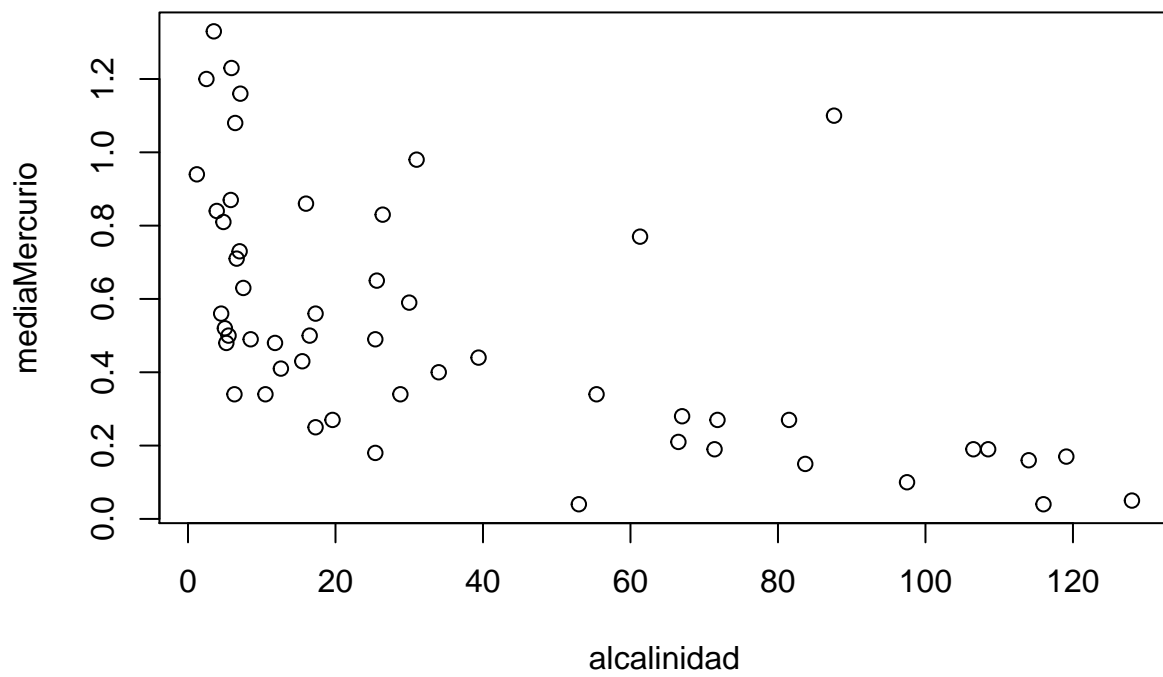


Igualmente es un mal modelo. Los residuos se distribuyen de manera no normal. Todos nuestros tres modelos muestran sesgo a la derecha y una curtosis positiva.

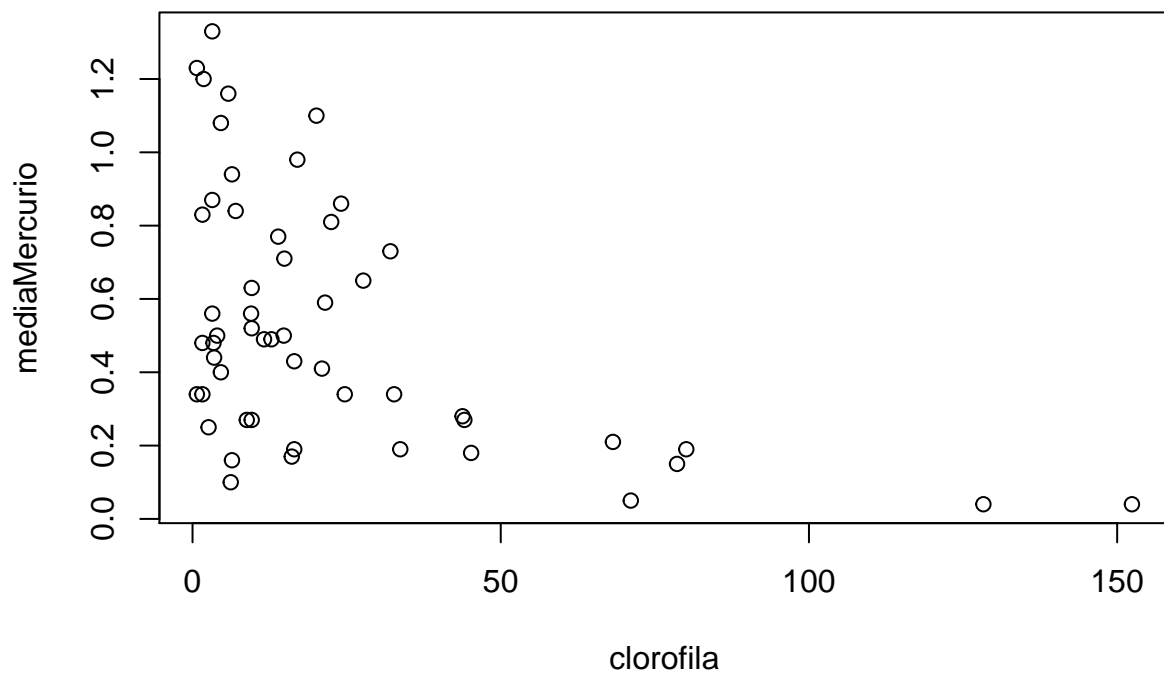
Como nos dio un modelo peor que cuando no interactuaban, nos quedaremos con el modelo de alcalinidad+clorofila.

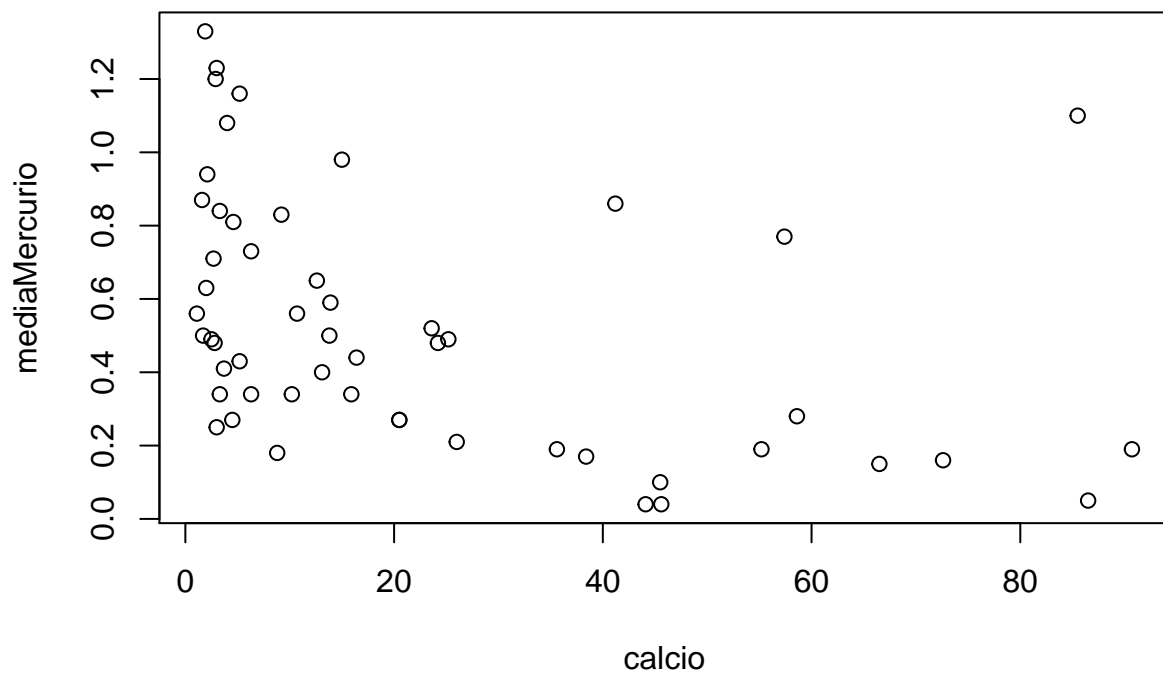
## Modelo

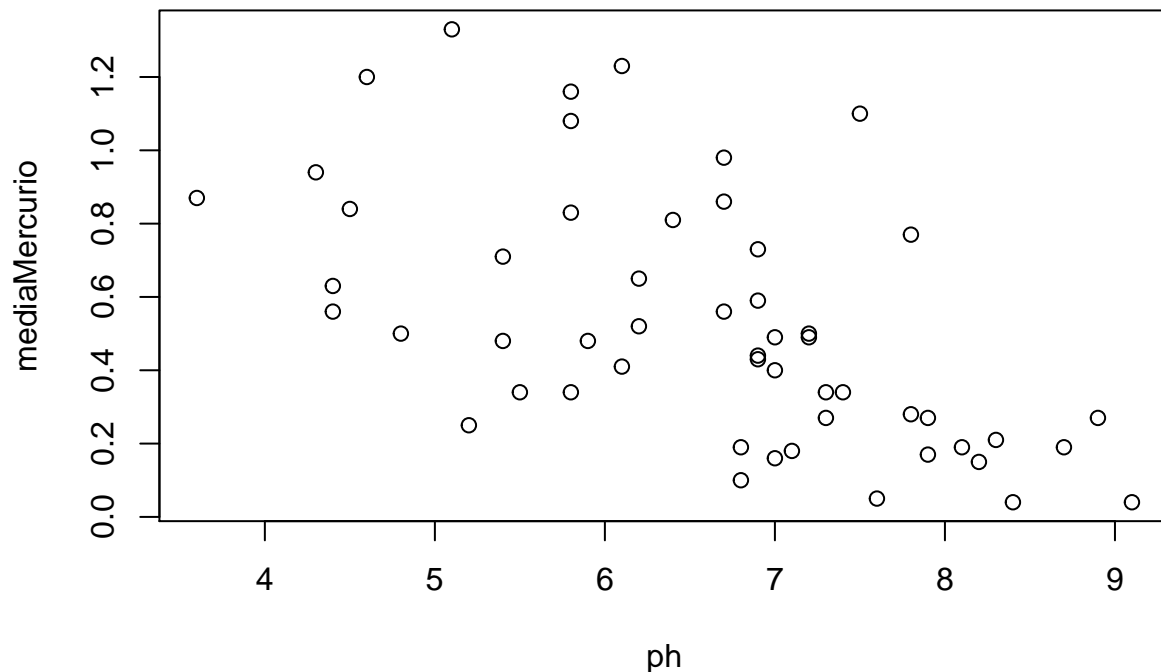
Este modelo sigue siendo malo, sin embargo, no hay duda que hay efecto de estas dos variables en la media del mercurio, para esto vamos a analizar como se comporta la media con cada una de estas variables por separado.











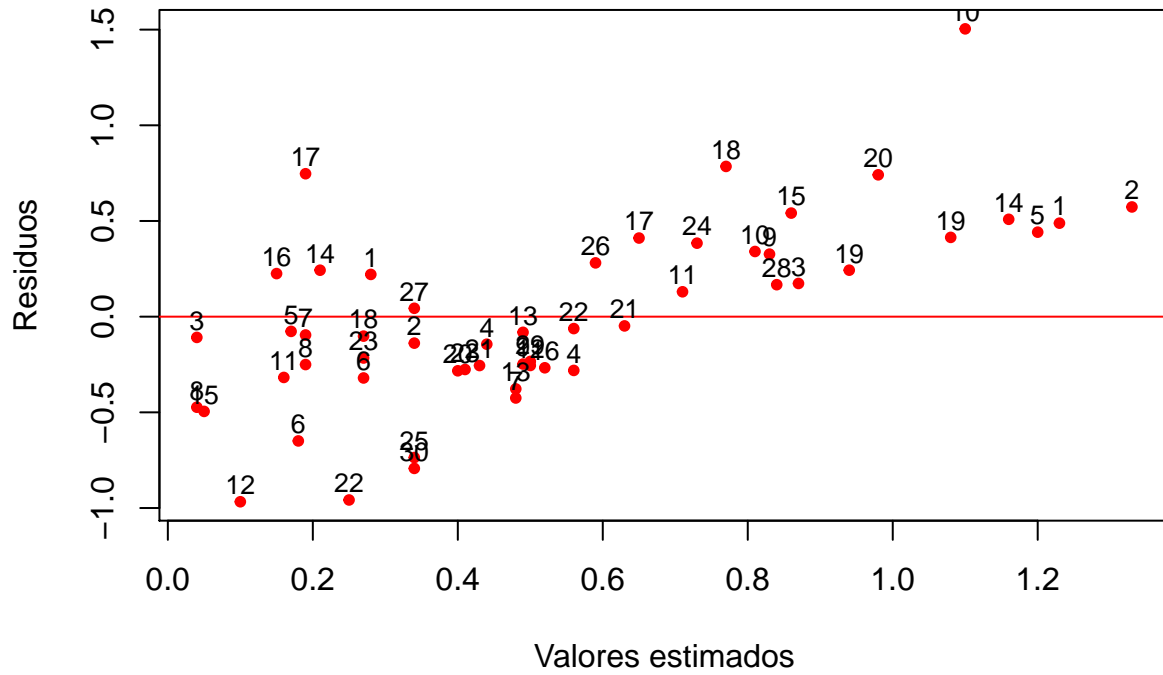
Como podemos ver la tendencia de estas dos variables no es linear. Se muestra como una exponencial decreciente con concava positiva ( $n^x$ ). Esto explica porque el modelo no se ajusta bien a pesar de que si muestran efecto.

Como lo que se busca encontrar en esta investigacion es cuales factores afectan el agua y no predecir los niveles de mercurio, no es necesario encontrar un modelo de mejor ajuste. Contamos ya con las variables que tienen efecto sobre la media de nivel de mercurio en los peces y podemos dar como concluida la investigacion.

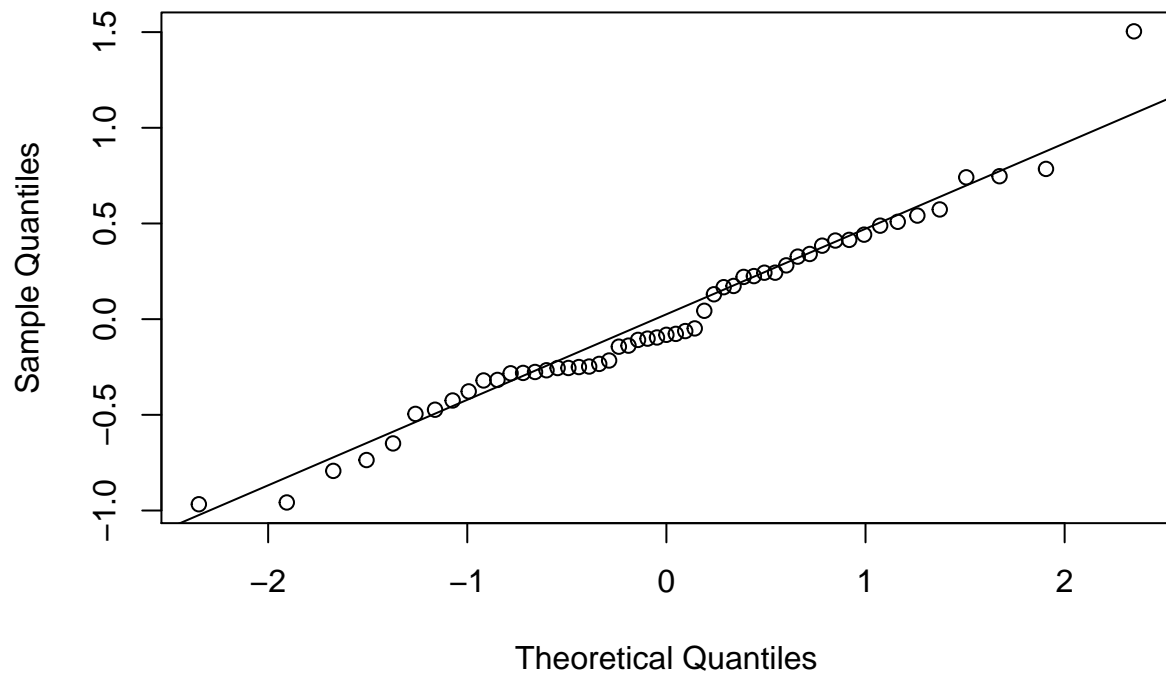
Sin embargo, si quisieramos construir un modelo para predecir la media del mercurio, podemos analizar como se comportarian estas variables con un modelo exponencial.

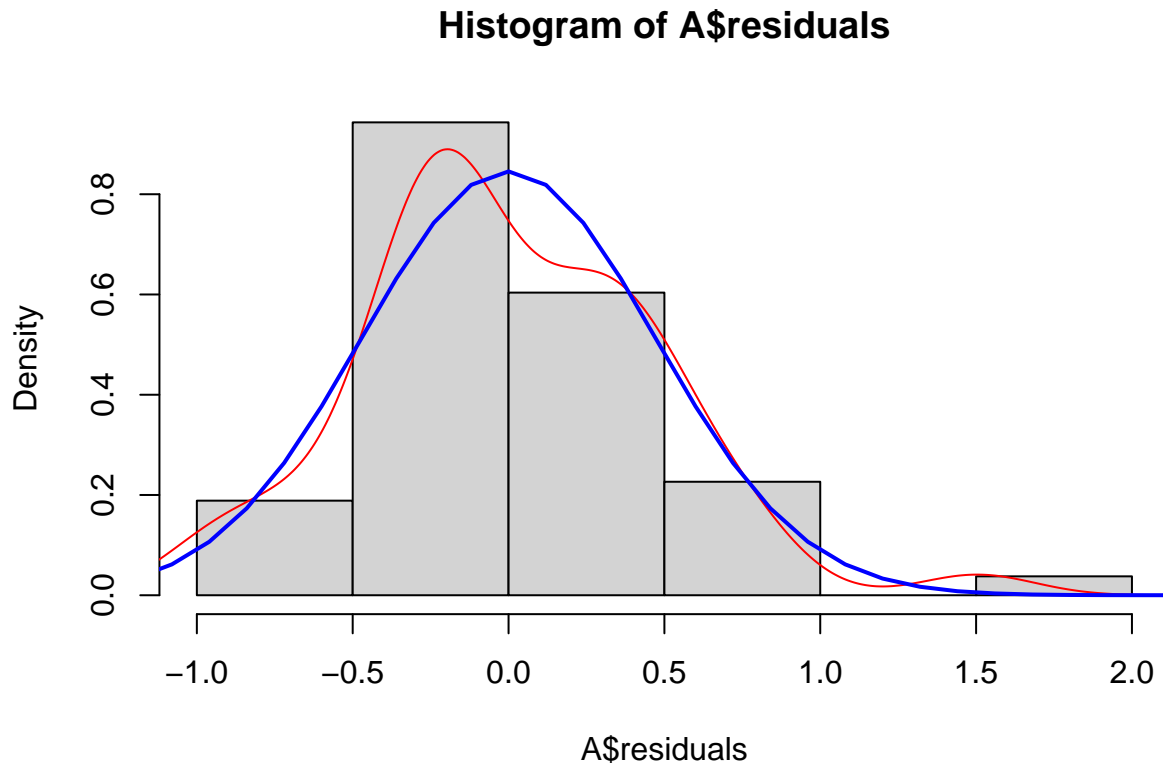
```
##
## Call:
## lm(formula = log(mediaMercurio) ~ alcalinidad + clorofila)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96716 -0.27599 -0.08178  0.32685  1.50399
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.209275  0.095504  -2.191   0.0331 *
## alcalinidad -0.010729  0.001988  -5.397 1.86e-06 ***
## clorofila   -0.012912  0.002464  -5.239 3.24e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4812 on 50 degrees of freedom
```

## Multiple R-squared: 0.6841, Adjusted R-squared: 0.6715  
## F-statistic: 54.14 on 2 and 50 DF, p-value: 3.083e-13



Normal Q-Q Plot





Exito! El modelo contiene puras variables significantes y tenemos un  $R^2$  de 0.68 lo cual es un modelo bueno, el mejor hasta ahora. Significa que podemos explicar 68.41% de la variación por el modelo. La fórmula quedaría:

$$\log(\text{MEDIA DE MERCURIO}) = -0.012912(\text{clorofila}) + -0.010729(\text{alcalinidad}) + -0.209275 \text{ MEDIA DE MERCURIO} = e^{\{-0.012912(\text{clorofila}) - 0.010729(\text{alcalinidad}) - 0.209275\}}$$

No solo tenemos la mejor  $R^2$  hasta ahora si no que también nuestros residuos parecen distribuirse de una manera más normal. No es perfecto pero indica nuevamente que el mejor modelo hasta ahorita.

Si quisiéramos mejorar nuestro modelo para predecir, usaríamos todas las variables, pero si utilizamos todas, las demás variables no tendrían efecto en la media pero mejoraría nuestra  $R^2$ . Solo como referencia haremos el análisis del modelo con todas las variables solo para tener el mejor modelo posible.

```
##
## Call:
## lm(formula = log(mediaMercurio) ~ alcalinidad + clorofila + ph +
##     madurez + calcio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89693 -0.31675  0.00029  0.30284  1.22458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.486670   0.470447  -1.034  0.306203
## alcalinidad -0.015170   0.003819  -3.972  0.000243 ***
## clorofila   -0.012911   0.002974  -4.341  7.49e-05 ***
```

```
## ph          0.027874    0.087686    0.318 0.751985
## madurez     0.125113    0.188551    0.664 0.510220
## calcio      0.007154    0.004922    1.453 0.152758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4801 on 47 degrees of freedom
## Multiple R-squared:  0.7043, Adjusted R-squared:  0.6729
## F-statistic: 22.39 on 5 and 47 DF,  p-value: 2.059e-11
```

Vemos que la  $R^2$  solo incremento por muy poco, el p-value para todas las que agregamos son muy bajos. Las incluimos en el modelo ya que en nuestro correlation chart vimos que ph y calcio tambien se mueven de manera similar a nuestras variables que si tienen efecto (exponencial). Este ultimo modelo puede ser ligeramente mejor que el otro pero con mas variables insignificantes lo cual tambien se tiene que considerar, por lo tanto el mejor modelo seria el anterior.

## CONCLUSION

Lo que podemos asumir es que entre mas baja la alcalinidad y mas baja la clorofila, la media del mercurio suele ser mas alta, sin embargo, si la incrementamos por un poco baja drasticamente y despues a como vamos incrementando estas dos variables, la media del mercurio baja semi-constantemente. Por lo tanto, si quisieramos bajar el mercurio en los peces por la contaminacion debemos encargarnos de incrementar la clorofila y la alcalinidad. Esto es una buena indicacion ya que aumentar la clorofila y alcalinidad por cantidades no muy grandes podemos disminuir mucho la media del mercurio. Podemos introducir quimicos para alterar estas propiedades en los lagos.

Tras una investigacion en internet, encontramos que los niveles de alcalinidad bajos en el agua hace que el ph fluctue altamente en el agua que causa danos a todo lo que hace contacto con ella. Niveles de clorofila bajos indican que hay muy poca vida microscopica en el agua, que creo que tambien es senal de altos niveles de mercurio ya que indica que hay poca vida ahi, probablemente por la contaminacion que ya hay. [2, 3] Es importante respaldar nuestra investigacion estadistica con una cientifica para no solo confirmar pero tambien razonar y comparar los resultados.

En conclusion, las características del agua que mas suelen afectar los niveles de mercurio en el agua son la alcalinidad y la clorofila.

## REPOSITORIO

Archivos Utilizados para este analisis: [https://drive.google.com/drive/folders/16Y6\\_cbbXaWo\\_AuxQGE0QRvrL4VdnlmkS?usp=sharing](https://drive.google.com/drive/folders/16Y6_cbbXaWo_AuxQGE0QRvrL4VdnlmkS?usp=sharing) Dentro de la carpeta “Mercurio”

## REFERENCIAS

- [1] ¿Por qué el mercurio supone un riesgo sanitario?. (2022). Retrieved 14 September 2022, from <https://www.mundodeportivo.com/vidae/salud/20220201/1001744736/mercurio-supone-riesgo-sanitario-act-pau.html>
- [2] Retrieved 14 September 2022, from: <https://hannainst.com.mx/blog/por-que-medir-la-alcalinidad-en-el-agua-potable/>
- [3] Analisis clorofila a y feopigmentos en agua - Phytocontrol. (2021). Retrieved 14 September 2022, from <https://www.phytocontrol.com/es/noticias/analisis-clorofila-a-y-feopigmentos-en-agua/>