

# Mercurio

# Gilberto Ramos Salinas

13 de septiembre de 2022

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. Por ello, se decide realizar el siguiente estudio el cual pueda determinar las principales causas de esta problemática explicando las razones principales por las que se propicia la concentración de mercurio.

Estas son las variables que se utilizan para realizar dicho estudio:

- X1 = número de identificación
- 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)

Alrededor de la principal pregunta de investigación que surge en este estudio: ¿Cuáles son los principales factores que influyen en el nivel de contaminación por mercurio en los peces de los lagos de Florida? pueden surgir preguntas paralelas que desglosan esta pregunta general:

¿Hay evidencia para suponer que la concentración promedio de mercurio en los lagos es dañino para la salud humana? Considera que las normativas de referencia para evaluar los niveles máximos de Hg (Reglamento 34687-MAG y los reglamentos internacionales CE 1881/2006 y Codex Standard 193-1995) establecen que la concentración promedio de mercurio en productos de la pesca no debe superar los 0.5 mg de Hg/kg. ¿Habrá diferencia significativa entre la concentración de mercurio por la edad de los peces? Si el muestreo se realizó lanzando una red y analizando los peces que la red encontraba ¿Habrá influencia del número de peces encontrados en la concentración de mercurio en los peces? ¿Las concentraciones de alcalinidad, clorofila, calcio en el agua del lago influyen en la concentración de mercurio de los peces?

Por medio de la siguiente instrucción podemos conocer datos generales sobre las variables que se tienen que analizar.

12 Variables										53 Observations										
										Id										
										n missing distinct										
										Info										
										Mean										
Gmd	.05	.10	53	0	53	1	27	18	3.6	6.2										
.25	.50	.75	.90	.95																
14.0	27.0	40.0	47.8	50.4	lowest :	1	2	3	4	5	highest:	49	50	51	52	53				
					Lake					n missing distinct					53		0		53	
lowest :			Alligator		Annie		Apopka		Blue Cypress			Brick		highest: Trout						
Tsala Apopka		Weir		Wildcat		Yale														

----- Alcalinity      n missing distinct      Info  
 Mean    Gmd    .05    .10    53    0    51    1    37.53    40.84    3.74    4.84  
       .25    .50    .75    .90    .95  
       6.60    19.60    66.50    104.70    114.80    lowest : 1.2    2.5    3.5    3.9    4.5, highest: 108.5    114.0    116.0    119.1  
 128.0 -----

----- Ph      n missing distinct      Info  
 Mean    Gmd    .05    .10    53    0    34    0.999    6.591    1.475    4.40    4.64  
       .25    .50    .75    .90    .95  
       5.80    6.80    7.40    8.18    8.52    lowest : 3.6    4.3    4.4    4.5    4.6, highest: 8.3  
 8.4    8.7    8.9    9.1

----- Calcium      n missing distinct      Info  
 Mean    Gmd    .05    .10    53    0    48    1    22.2    25.66    1.82    2.18  
       .25    .50    .75    .90    .95  
       3.30    12.60    35.60    58.36    77.76    lowest : 1.1    1.6    1.7    1.9    2.0, highest: 66.5    72.6  
 85.5    86.5    90.7

----- Chlorophyl  
       n missing distinct      Info    Mean    Gmd    .05    .10    53    0    43    0.999    23.12  
 27.72    1.60    1.96  
       .25    .50    .75    .90    .95  
       4.60    12.80    24.70    63.60    79.20    lowest : 0.7    1.6    1.8    2.6    3.2, highest: 71.1    78.6    80.1    128.3  
 152.4 -----

Mercury\_con  
       n missing distinct      Info    Mean    Gmd    .05    .10    53    0    41    0.999    0.5272  
 0.387    0.080    0.162  
       .25    .50    .75    .90    .95  
       0.270    0.480    0.770    1.060    1.176    lowest : 0.04    0.05    0.10    0.15    0.16, highest: 1.10  
 1.16    1.20    1.23    1.33

----- no\_fish      n missing distinct      Info  
 Mean    Gmd    .05    .10    53    0    15    0.937    13.06    6.821    6.0    6.2  
       .25    .50    .75    .90    .95  
       10.0    12.0    12.0    14.0    37.6    lowest : 4    5    6    7    8,  
 highest: 24    36    40    43    44

Value      4    5    6    7    8    10    11    12    13    14    24  
 Frequency    1    1    4    2    1    11    3    20    3    2    1  
 Proportion 0.019    0.019    0.075    0.038    0.019    0.208    0.057    0.377    0.057    0.038    0.019

Value      36    40    43    44  
 Frequency    1    1    1    1  
 Proportion 0.019    0.019    0.019    0.019

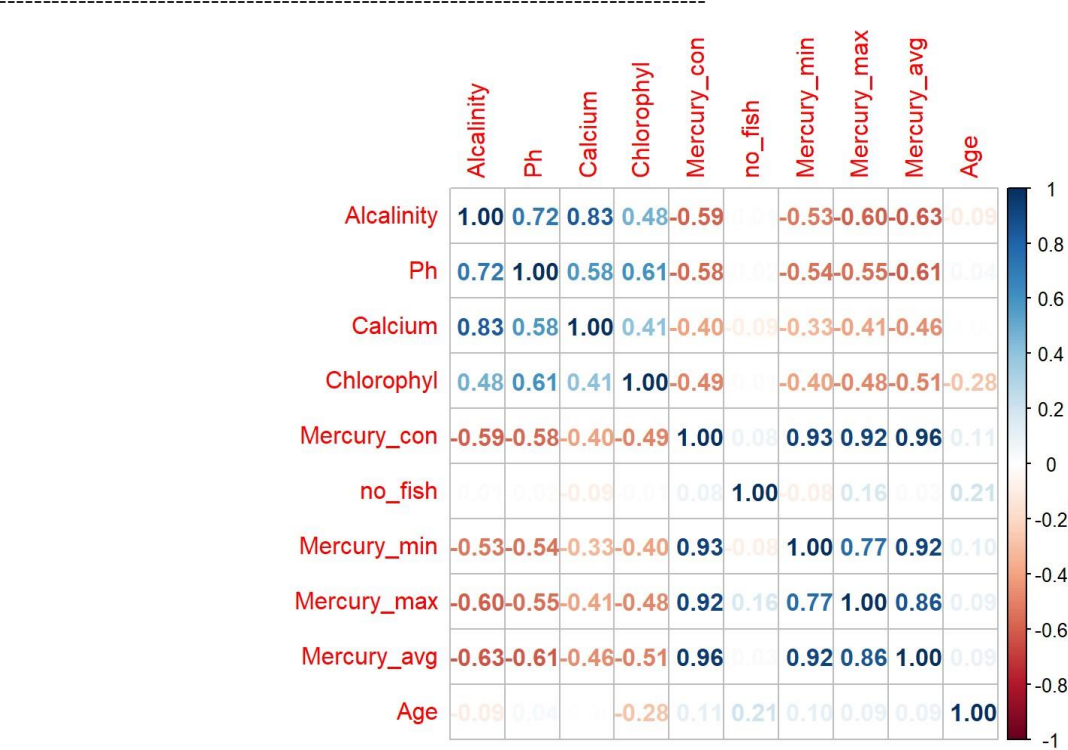
----- Mercury\_min      n missing distinct  
 Info    Mean    Gmd    .05    .10    53    0    32    0.998    0.2798    0.2448    0.040    0.042  
       .25    .50    .75    .90    .95  
       0.090    0.250    0.330    0.630    0.730    lowest : 0.04    0.05    0.07    0.08    0.09, highest: 0.67  
 0.69    0.79    0.85    0.92

----- Mercury\_max      n missing distinct  
 Info    Mean    Gmd    .05    .10    53    0    44    1    0.8745    0.5964    0.152    0.266  
       .25    .50    .75    .90    .95  
       0.480    0.840    1.330    1.494    1.900    lowest : 0.06    0.11    0.18    0.26    0.29, highest: 1.47  
 1.50    1.90    2.03    2.04

----- Mercury\_avg      n missing distinct  
 Info    Mean    Gmd    .05    .10    53    0    38    0.999    0.5132    0.3727    0.110    0.160  
       .25    .50    .75    .90    .95

0.250 0.450 0.700 0.962 1.144 lowest : 0.04 0.05 0.15 0.16 0.18, highest: 0.98  
1.00 1.02 1.33 1.53

-----										Age	n missing	distinct	Info	Sum
Mean	Gmd	53	0	2	0.459	43	0.8113	0.312						



A través de esta información sabemos que existen varios datos que tienen una alta correlación con Mercury\_con como lo son: Mercury\_Avg, Mercury\_max, Mercury\_min. Sin embargo, no podemos utilizar estas variables ya que son datos con una alta relación. Por ello, se decidió agregar otra variable con menos relación para conocer los parámetros reales que producen una alta concentración del mercurio como: Mercury\_avg con otras variables como Chlorophyl que no tienen una correlación tan grande entre ellas.

Aplicación del modelo lineal

Call:

```
lm(formula = Mercury_con ~ Alkalinity + Calcium + Mercury_avg,
    data = dfdrop)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.27066	-0.06213	-0.00497	0.02858	0.27959

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.0279556	0.0411953	0.679	0.5006
Alkalinity	-0.0008615	0.0007251	-1.188	0.2405
Calcium	0.0016916	0.0009764	1.733	0.0895 .
Mercury_avg	0.9625544	0.0511470	18.819	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09633 on 49 degrees of freedom

Multiple R-squared: 0.9248, Adjusted R-squared: 0.9202

F-statistic: 200.9 on 3 and 49 DF, p-value: < 2.2e-16

## Selección del mejor modelo

Start: AIC=-244.2

Mercury\_con ~ Alkalinity + Calcium + Mercury\_avg

	Df	Sum of Sq	RSS	AIC
- Alkalinity	1	0.0131	0.4678	-244.69
<none>			0.4547	-244.20
- Calcium	1	0.0279	0.4826	-243.04
- Mercury_avg	1	3.2865	3.7412	-134.50

Step: AIC=-244.69

Mercury\_con ~ Calcium + Mercury\_avg

	Df	Sum of Sq	RSS	AIC
- Calcium	1	0.0155	0.4833	-244.97
<none>			0.4678	-244.69
+ Alkalinity	1	0.0131	0.4547	-244.20
- Mercury_avg	1	4.6091	5.0769	-120.32

Step: AIC=-244.97

Mercury\_con ~ Mercury\_avg

	Df	Sum of Sq	RSS	AIC
<none>			0.4833	-244.97
+ Calcium	1	0.0155	0.4678	-244.69
+ Alkalinity	1	0.0007	0.4826	-243.04
- Mercury_avg	1	5.5646	6.0479	-113.04

```
Call:
lm(formula = Mercury_con ~ Mercury_avg, data = dfdrop)

Coefficients:
(Intercept)  Mercury_avg
    0.03154      0.96575
```

```
Call:
lm(formula = Mercury_con ~ Mercury_avg, data = dfdrop)

Residuals:
    Min       1Q   Median       3Q      Max
-0.27913 -0.04133 -0.01606  0.01402  0.27244

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.03154    0.02444   1.291   0.203
Mercury_avg   0.96575    0.03985  24.233 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09734 on 51 degrees of freedom
Multiple R-squared:  0.9201,    Adjusted R-squared:  0.9185
F-statistic: 587.2 on 1 and 51 DF,  p-value: < 2.2e-16
```

A través de esta información podemos inferir que Mercurio\_avg es significativa para la predicción ya que \*\*\* tiene un valor de 0.001 estando cerca del cero.

Concentrado del Mercurio es =

Pruebas de hipótesis

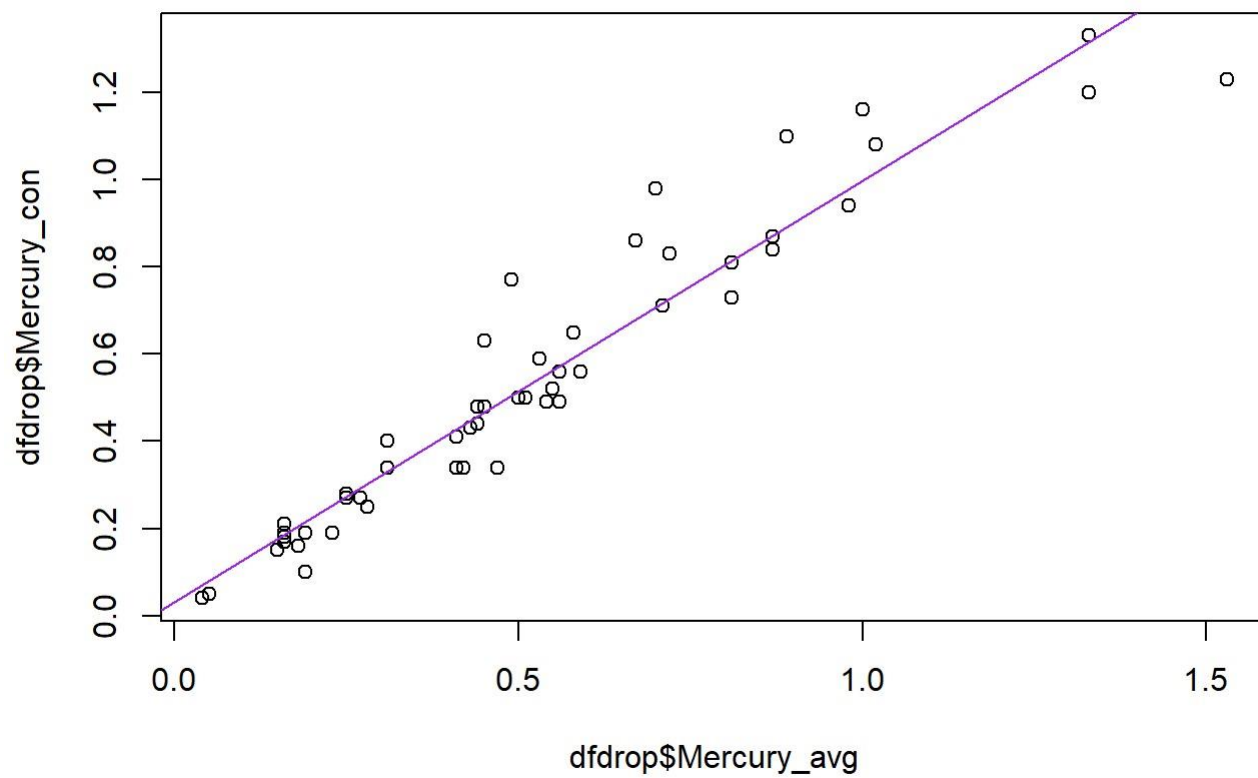
$H_0: \beta_1 = 0$   $H_1: \beta_1 \neq 0$

```
[1] "H0 se rechaza "
```

Por lo tanto, la hipótesis 1 se acepta indicando que  $\beta_1$  se utiliza. Esto indica que Mercurio\_avg tiene un valor significativo para los resultados del modelo.

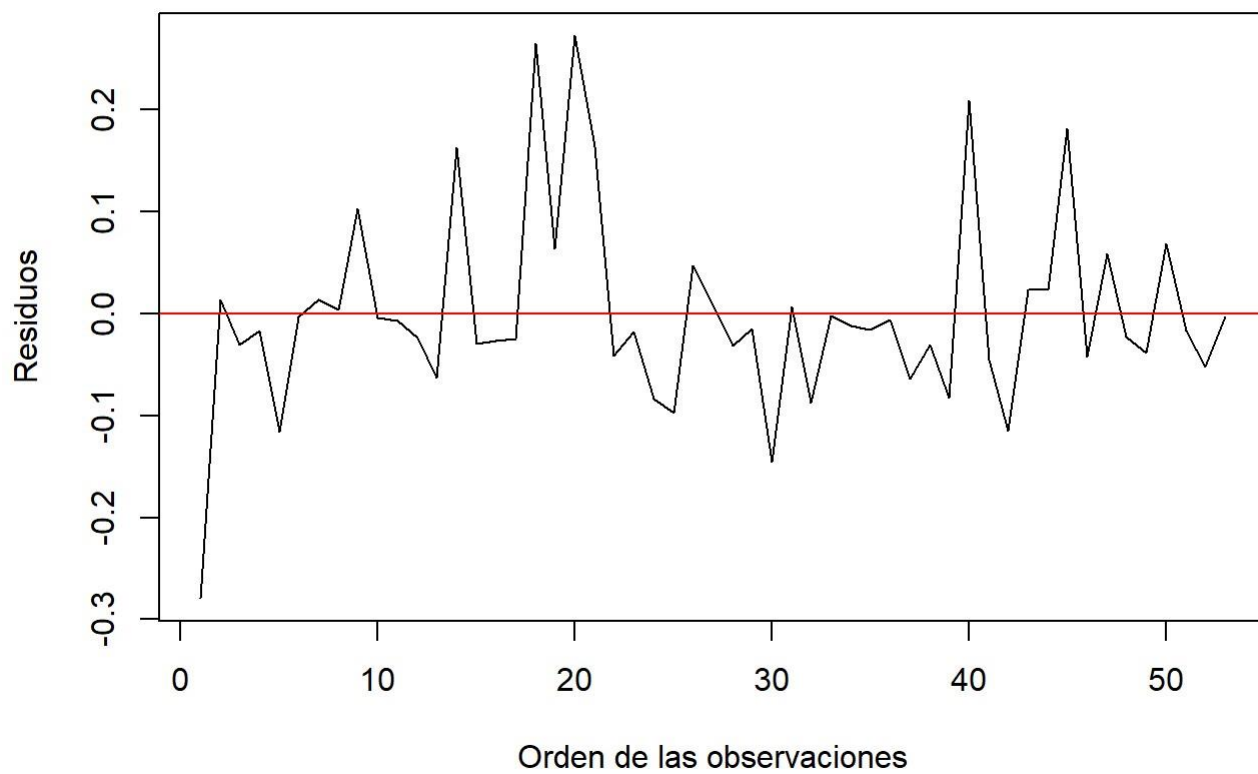
Verificación de supuestos:

Relacion Lineal



Por medio del siguiente diagrama se puede apreciar claramente como los datos poseen una relación lineal

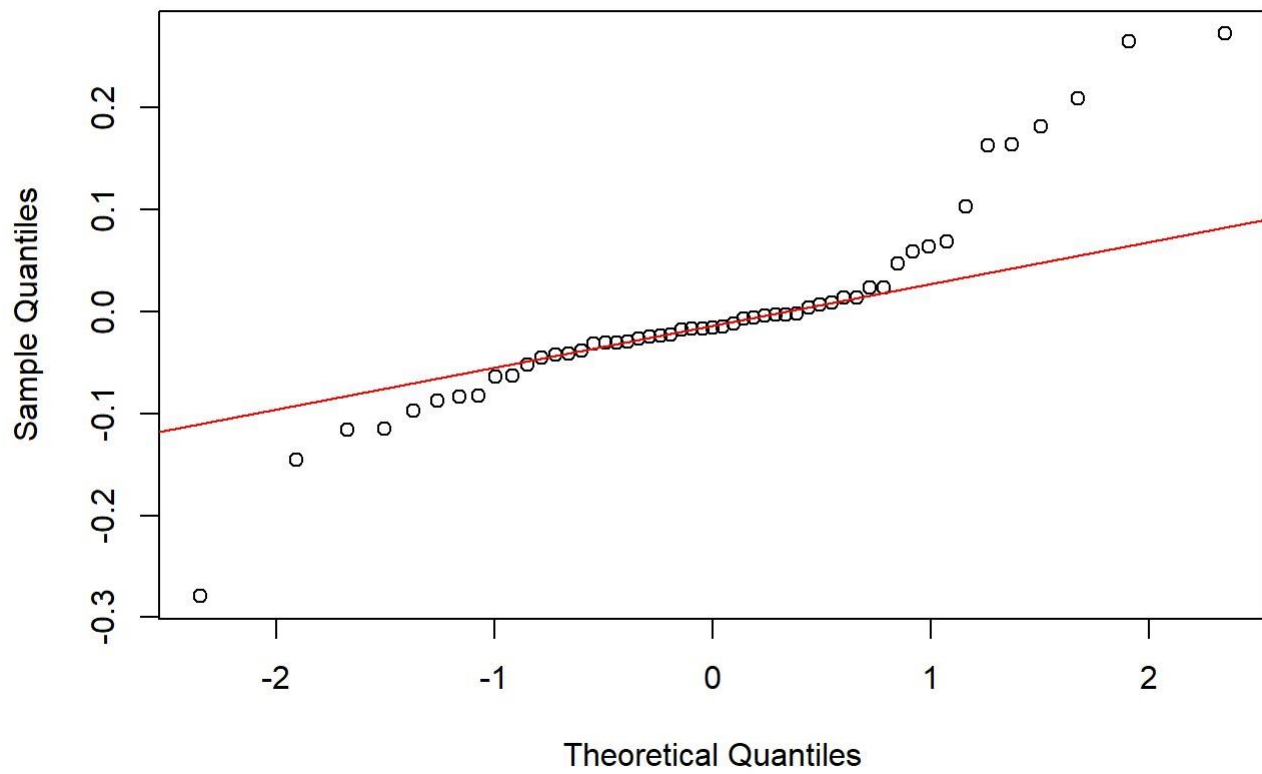
Independencia los residuos



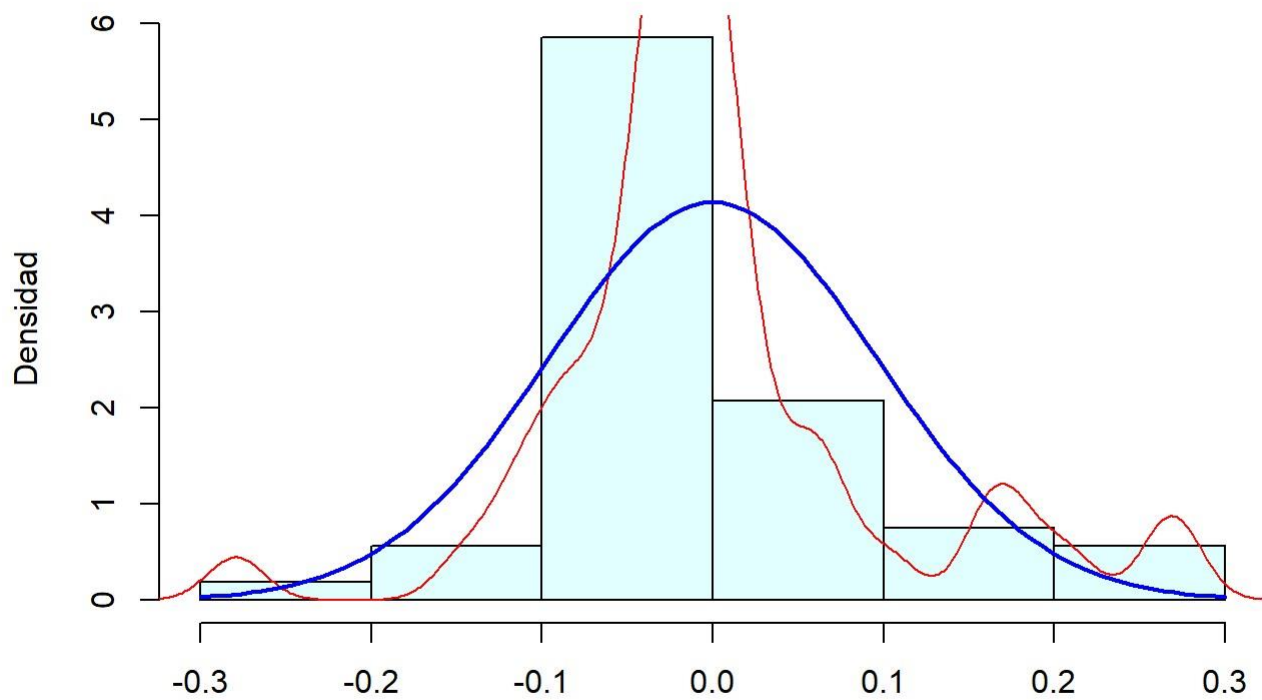
Los datos como pueden ser apreciados anteriormente poseen no tienen un tendencia defnida lo que nos indica que son totalmente independientes.

Normalidad

**Normal Q-Q Plot**



**Histograma de Residuos**





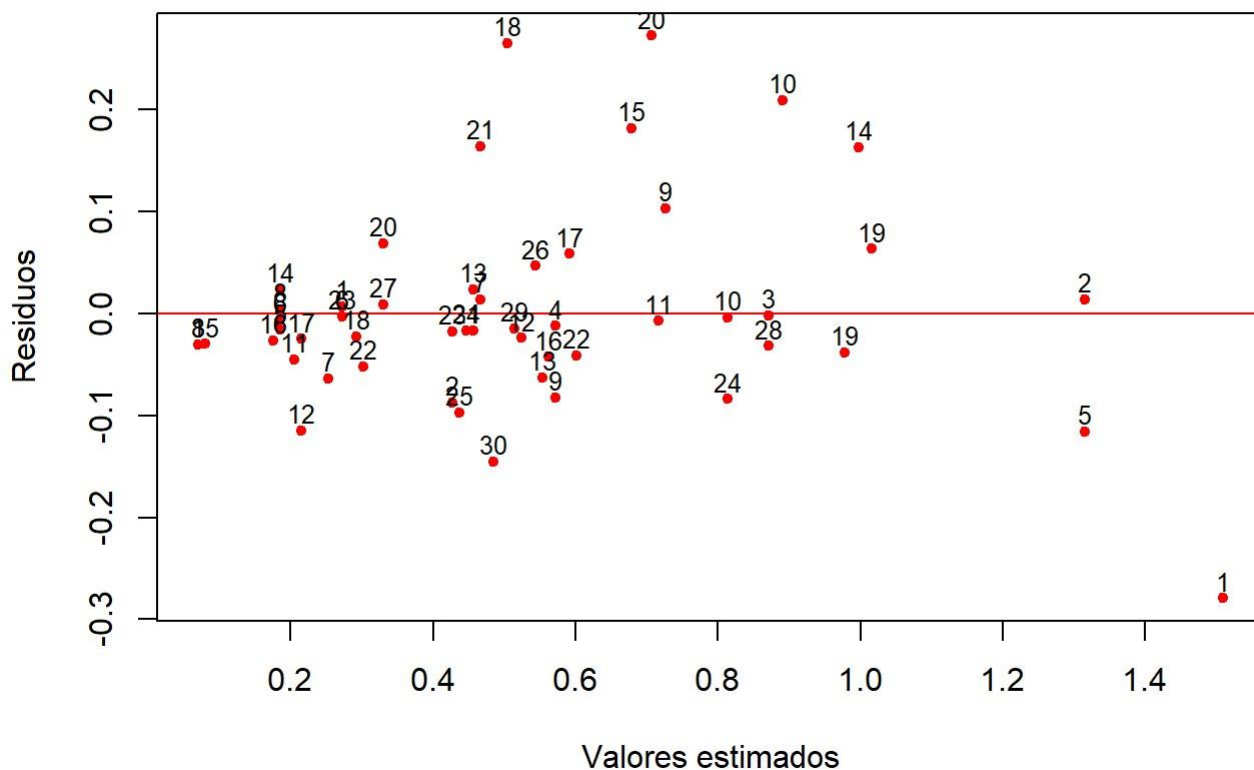
### Shapiro-Wilk normality test

data: E

W = 0.89095, p-value = 0.0001611

Los datos son distribuidos con normalidad.

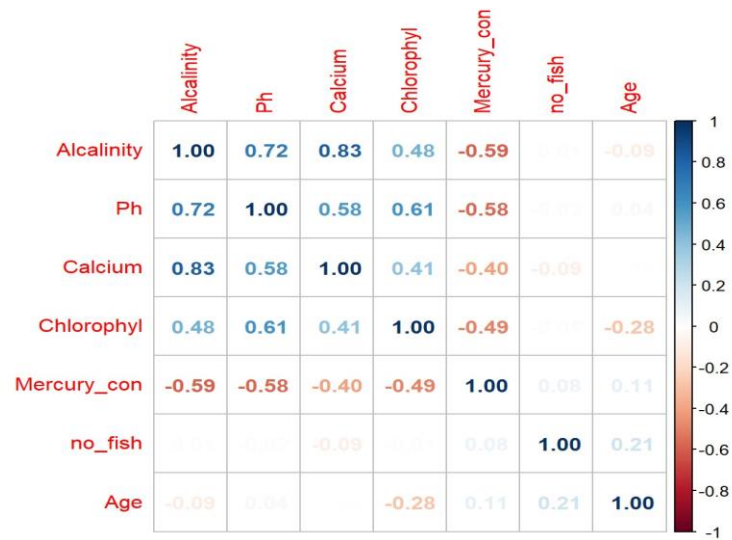
Homocedasticidad



Un modelo presenta homocedasticidad cuando existe una constancia en los errores de estimación. Sin embargo, este modelo presenta heterocedasticidad, dado que la dispersión del error (varianza) no es igual en la mayoría de las variables.

### Conclusión:

Acorde a los resultados anteriormente presentados podemos conocer que la variable Mercury\_avg siendo la estimación de mercurio que se encuentra en los peces en 3 años es un indicador fuerte para la concentración del mercurio en un lago. Esto fue verificado a lo largo de este documento haciendo uso de hipótesis y de verificación de supuestos. Sin embargo, podemos realizar nuevamente el estudio donde se descarte el máximo, mínimo y estimación del mercurio, ya que estas son variables que tienen una alta correlación, pudiendo afectar al modelo. Con ello, podemos conseguir un modelo que demuestre las variables reales que afecten.



Call:

```
lm(formula = Mercury_con ~ Alkalinity + Calcium, data = newdfdrop)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.46786	-0.19292	-0.04874	0.13716	0.68733

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.718575	0.053152	13.519	< 2e-16 ***
Alkalinity	-0.007574	0.001793	-4.225	0.000101 ***
Calcium	0.004182	0.002747	1.523	0.134163

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2735 on 50 degrees of freedom

Multiple R-squared: 0.3814, Adjusted R-squared: 0.3567

F-statistic: 15.41 on 2 and 50 DF, p-value: 6.1e-06

[1] "H1 se rechaza "

[1] "H1 se rechaza "

Por medio del siguiente análisis podemos apreciar que el modelo es bajo ya que r cuadrada tiene un valor de .35 y las variables de alcalinidad y calcio tienen correlación baja ya que H1 se rechaza. Por lo que, se rechaza el modelo.