

## Mercurio\_2

Luis Cano Irigoyen A00827178

2022-10-26

### 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. 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)

¿Cuáles son los principales factores que influyen en el nivel de contaminación por mercurio en los peces de los lagos de Florida?

### Aproximación al problema

Buscaremos realizar una Regresión Lineal para determinar el mejor modelo que conteste la pregunta. X7 es la variable que buscaremos predecir con el modelo. Este problema lo atacaremos con Pruebas de Hipótesis

### Datos

```
D=read.csv("mercurio.csv")
N=nrow(D)
```

Cambiamos el nombre de las columnas para comprender mejor los análisis

```
colnames(D) <- c("ID", "Nombre", "Alcalinidad", "PH", "Calcio",
"Clorofila", "MediaMercurio",
"NumPez", "MinMercurio", "MaxMercurio",
"TresMercurio", "Edad")
head(D, 5)
```

##	ID	Nombre	Alcalinidad	PH	Calcio	Clorofila	MediaMercurio
## 1	1	Alligator	5.9	6.1	3.0	0.7	1.23
## 2	2	Annie	3.5	5.1	1.9	3.2	1.33
## 3	3	Apopka	116.0	9.1	44.1	128.3	0.04
## 4	4	Blue Cypress	39.4	6.9	16.4	3.5	0.44
## 5	5	Brick	2.5	4.6	2.9	1.8	1.20

  

##	MinMercurio	MaxMercurio	TresMercurio	Edad
## 1	0.85	1.43	1.53	1
## 2	0.92	1.90	1.33	0
## 3	0.04	0.06	0.04	0
## 4	0.13	0.84	0.44	0
## 5	0.69	1.50	1.33	1

ID y Nombre no son variables explicativas No es útil analizarlas por lo cual las eliminamos

```
D$ID <- NULL
D$Nombre <- NULL

Y=D[,5]
X=D[, -5]
```

## 1.Realice un análisis de normalidad de las variables continuas para identificar variables normales. Tome en cuenta los puntos que se sugieren a continuación (no son exhaustivos):

### A.

Realice la prueba de normalidad de Mardia y la prueba de Anderson Darling para identificar las variables que son normales y detectar posible normalidad multivariada de grupos de variables.

Normalidad de Mardia

```
library(MVN)
mvn(D, subset = NULL, mvn = "mardia")

## $multivariateNormality
##           Test           Statistic           p value Result
## 1 Mardia Skewness 502.667343452414 3.6277693977554e-24    NO
## 2 Mardia Kurtosis 4.83254138772002 1.34801075923896e-06    NO
## 3           MVN           <NA>           <NA>           NO
##
```

```
## $univariateNormality
##           Test      Variable Statistic    p value Normality
## 1 Anderson-Darling Alcalinidad    3.6725 <0.001      NO
## 2 Anderson-Darling      PH        0.3496 0.4611      YES
## 3 Anderson-Darling      Calcio    4.0510 <0.001      NO
## 4 Anderson-Darling    Clorofila    5.4286 <0.001      NO
## 5 Anderson-Darling MediaMercurio    0.9253 0.0174      NO
## 6 Anderson-Darling      NumPez     8.6943 <0.001      NO
## 7 Anderson-Darling MinMercurio     1.9770 <0.001      NO
## 8 Anderson-Darling MaxMercurio     0.6585 0.081      YES
## 9 Anderson-Darling TresMercurio     1.0469 0.0086      NO
## 10 Anderson-Darling      Edad      14.3350 <0.001      NO
##
## $Descriptives
##           n      Mean    Std.Dev Median   Min    Max   25th   75th
## Alcalinidad  53 37.5301887 38.2035267   19.60  1.20 128.00   6.60  66.50
## PH           53  6.5905660  1.2884493    6.80  3.60   9.10   5.80   7.40
## Calcio       53 22.2018868 24.9325744   12.60  1.10  90.70   3.30  35.60
## Clorofila    53 23.1169811 30.8163214   12.80  0.70 152.40   4.60  24.70
## MediaMercurio 53  0.5271698  0.3410356    0.48  0.04   1.33   0.27   0.77
## NumPez       53 13.0566038  8.5606773   12.00  4.00  44.00  10.00  12.00
## MinMercurio  53  0.2798113  0.2264058    0.25  0.04   0.92   0.09   0.33
## MaxMercurio  53  0.8745283  0.5220469    0.84  0.06   2.04   0.48   1.33
## TresMercurio 53  0.5132075  0.3387294    0.45  0.04   1.53   0.25   0.70
## Edad        53  0.8113208  0.3949977    1.00  0.00   1.00   1.00   1.00
##
##           Skew    Kurtosis
## Alcalinidad  0.9679170 -0.4705349
## PH          -0.2458771 -0.6239638
## Calcio       1.3045868  0.6130359
## Clorofila    2.4130571  6.1042185
## MediaMercurio 0.5986343 -0.6312607
## NumPez       2.5808773  6.0089455
## MinMercurio  1.0729099  0.4060828
## MaxMercurio  0.4645925 -0.6692490
## TresMercurio 0.9449951  0.5733500
## Edad        -1.5465748  0.4005116
```

## Prueba de Anderson Darling

```
library(nortest)
for (i in 1:9) {
  print(colnames(D)[i])
  print(ad.test(D[,i]))
}

## [1] "Alcalinidad"
##
## Anderson-Darling normality test
##
## data:  D[, i]
## A = 3.6725, p-value = 2.706e-09
```

```
##
## [1] "PH"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 0.34956, p-value = 0.4611
##
## [1] "Calcio"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 4.051, p-value = 3.193e-10
##
## [1] "Clorofila"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 5.4286, p-value = 1.4e-13
##
## [1] "MediaMercurio"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 0.92528, p-value = 0.0174
##
## [1] "NumPez"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 8.6943, p-value < 2.2e-16
##
## [1] "MinMercurio"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 1.977, p-value = 4.161e-05
##
## [1] "MaxMercurio"
##
## Anderson-Darling normality test
##
## data: D[, i]
## A = 0.65847, p-value = 0.08099
##
```

```
## [1] "TresMercurio"  
##  
## Anderson-Darling normality test  
##  
## data: D[, i]  
## A = 1.0469, p-value = 0.008637
```