

# Reporte - Los Peces y el Mercurio

Nicolas Cardenas A01114959

2022-10-19

Módulo 5: Estadística Avanzada para ciencia de datos y nombre de la concentración

Grupo 502

## 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. Las variables que se midieron se encuentran en mercurio.csv 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)

```
## [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
##   Min.      :0.0400
##   1st Qu.:0.2500
##   Median :0.4500
##   Mean   :0.5132
##   3rd Qu.:0.7000
##   Max.    :1.5300
```

```
## [1] "Desviacion Estandar"
```

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

## Resumen

Se retomaron los diferentes temas vistos en clase para el analisis de los datos. Se realizo primeramente un analisis de normalidad de las variables (para determinar las significantes), asi como un analisis de componentes principales (con el objetivo de reducir dimensionalidad y encontrar variables relacionadas). Todo esto con el mismo objetivo que nuestro reporte pasado: determinar los factores principales de los niveles de mercurio en el agua.

## Introducción

Tras haber hecho un reporte analizando las diferentes variables a detalle, vamos a omitirlo para este reporte. Este reporte busca ser una extension de nuestro reporte pasado pero con el mismo objetivo, responder la pregunta: ¿Cuáles son los principales factores que influyen en el nivel de contaminación por mercurio en los peces de los lagos de Florida?

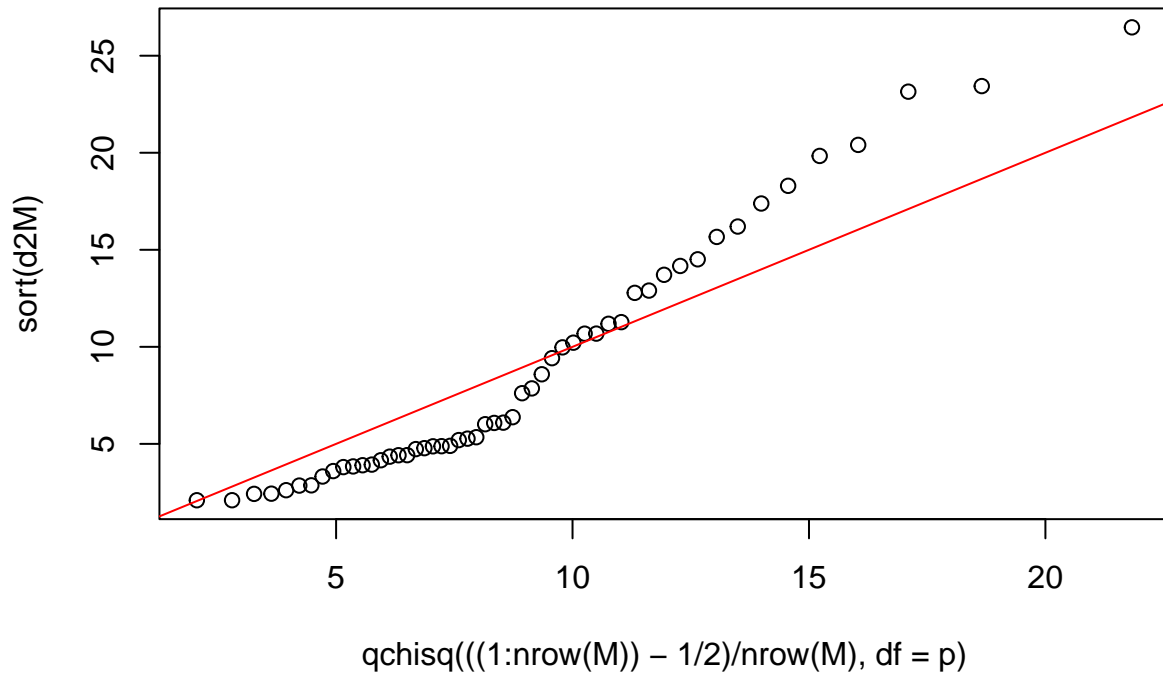
## Análisis de los resultados

## TEST MULTIVARIADO

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

Primero checamos la Distancia de Mahalanobis y Multinormalidad Test gráfico Q-Q Plot para observar datos atipicos:

```
## [1] 23.148248 10.214173 17.385270 2.613110 12.781132 4.880536 3.317725
## [8] 3.598373 8.585982 2.850207 4.342272 11.186573 4.414540 15.664149
## [15] 9.971855 6.093529 18.296407 13.711814 6.078138 16.197452 10.681426
## [22] 2.863256 2.096404 26.461606 4.730028 7.858117 2.423587 6.374972
## [29] 3.899854 4.418718 3.802860 2.434265 20.404938 4.873017 11.272724
## [36] 3.838086 14.166717 23.435759 4.779110 19.836807 9.415256 7.611127
## [43] 5.195259 3.932293 10.678490 4.897695 14.509979 12.898648 5.343472
## [50] 4.156698 2.095601 5.269222 6.012521
```



Utilizando el QQplot multivariado y la distancia de Mahalanobis podemos detectar que hay varios datos atípicos.

Seguiremos con un MVN test:

Prendemos showOutliers porque queremos los datos sin datos atípicos

Apagamos también covariance porque no queremos los datos estandarizados ya que nos importan las relaciones naturales de los factores.

```
## $multivariateNormality
##           Test           Statistic           p value Result
## 1  Mardia Skewness 410.214790601478 7.04198777815398e-23    NO
## 2  Mardia Kurtosis 4.59612555772731 4.30419392238868e-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  numPeces      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 estimacion     1.0469 0.0086         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
## numPeces     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
## estimacion   53 0.5132075 0.3387294 0.45 0.04 1.53 0.25 0.70
##           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
## numPeces     2.5808773 6.0089455
## minMercurio  1.0729099 0.4060828
## maxMercurio  0.4645925 -0.6692490
## estimacion   0.9449951 0.5733500
##
## $multivariateOutliers
## NULL
```

Lo que podemos observar el la curtosis es que tenemos buena curtosis, no es 3 pero es suficientemente cerca para lo que buscamos, va a ser leptocurtica, es decir un poco mas “flaca”. El cesgo es pesimo y es positivo lo cual indica que la moda es mas pequena que la mediana y la mediana es menor que el promedio. No nos indica normalidad.

Utilizaremos las variables que resultaron normales y volveremos a realiar el test.

## Variables Normales

```
## $multivariateNormality
##           Test      Statistic      p value Result
## 1 Mardia Skewness 6.17538668676458 0.186427564928852 YES
## 2 Mardia Kurtosis -1.12820795824432 0.25923210375991 YES
## 3              MVN              <NA>              <NA> YES
##
## $univariateNormality
##           Test      Variable Statistic      p value Normality
## 1 Anderson-Darling      ph      0.3496      0.4611 YES
## 2 Anderson-Darling maxMercurio 0.6585      0.0810 YES
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min     Max   25th   75th      Skew
## ph          53 6.5905660 1.2884493 6.80 3.60 9.10 5.80 7.40 -0.2458771
## maxMercurio 53 0.8745283 0.5220469 0.84 0.06 2.04 0.48 1.33 0.4645925
##           Kurtosis
## ph          -0.6239638
## maxMercurio -0.6692490
##
```

```
## $multivariateOutliers
## NULL
```

Nos da muchos mejores resultado, esto nos indica que ph y el max de mercurio cuentan con una dsitribucion normal multivariada.

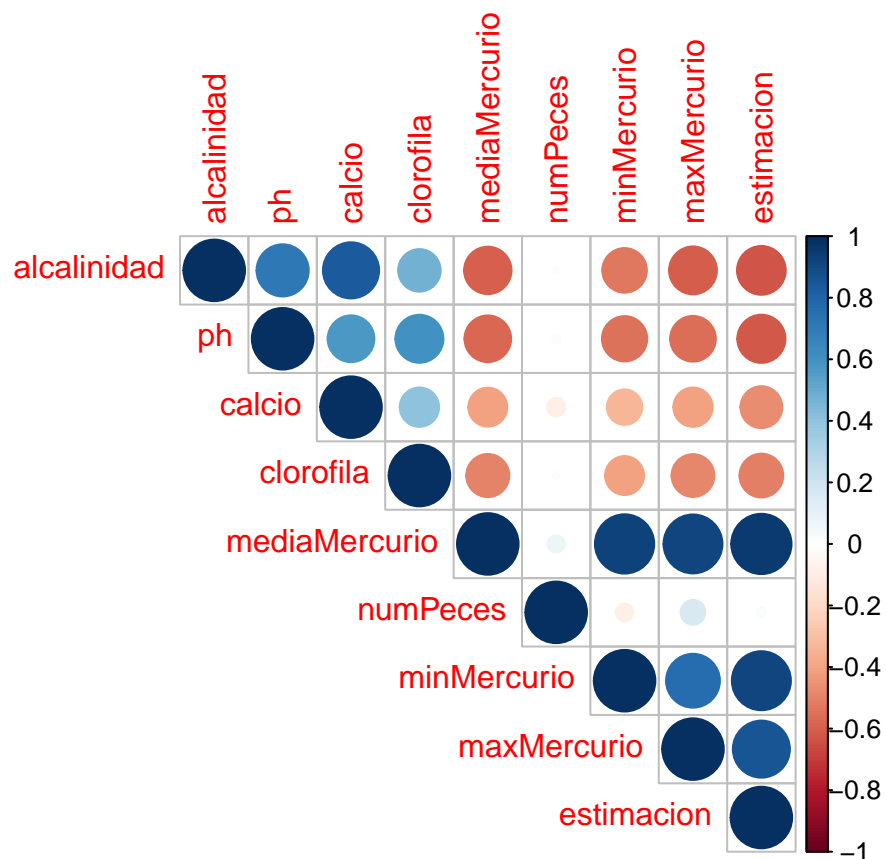
## COMPONENTES PRINCIPALES

Realizaremos un análisis de componentes principales con la base de datos completa para identificar los factores principales que intervienen en el problema de la contaminación por mercurio de los peces en agua dulce.

## CORRELACION

Para empezar a justificar el uso de componentes principales para este analisis, primero debemos observar la matriz de correlaciones.

```
## corrplot 0.92 loaded
```



Como podemos observar, todas las medidas de mercurio (min, max, media, estimacion) son grandemente relacionadas). Esto nos abre la puerta a realizar una buena reduccion de dimensiones y podemos “singularizar” nuestra variable objetivo (niveles de mercurio).

# COMPONENTES

A continuacion sacaremos los componentes principales:

```
## C.P. Covarianza: 0.7264164 0.9300767 0.9775266 0.9996963 0.9999067 0.9999882 0.9999979 0.9999994 1
```

```
## C.P. Correlacion (Estandarizados): 0.5939898 0.7296462 0.8454831 0.9196901 0.9569915 0.9802068 0.99
```

Podemos ver que con cuatro componentes tenemos mas de 99% de la varianza explicada, y cuando esta estandarizado, necesitamos 7 para explicar mas del 99%

Podriamos conformarnos con menos varianza explicada pero por la naturaleza del problema y para no perder informacion, dejaremos nuestro estandar a arriba de 99%

Ahora veremos los pesos para cada uno de los componentes para ver cuales estan relacionados:

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

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
##
```

```
## Loadings:
```

```
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
```

```
## alcalinidad 0.770 0.360 0.512 0.121
```

```
## ph          -0.989 0.140
```

```
## calcio      0.459 0.261 -0.825 -0.203
```

```
## clorofila   0.442 -0.896
```

```
## mediaMercurio          0.472 -0.282 0.307 0.773
```

```
## numPeces    0.237 -0.971
```

```
## minMercurio          0.295 -0.466 0.587 -0.589
```

```
## maxMercurio          0.694 0.693 -0.182
```

```
## estimacion    0.435 -0.471 -0.749 -0.147
```

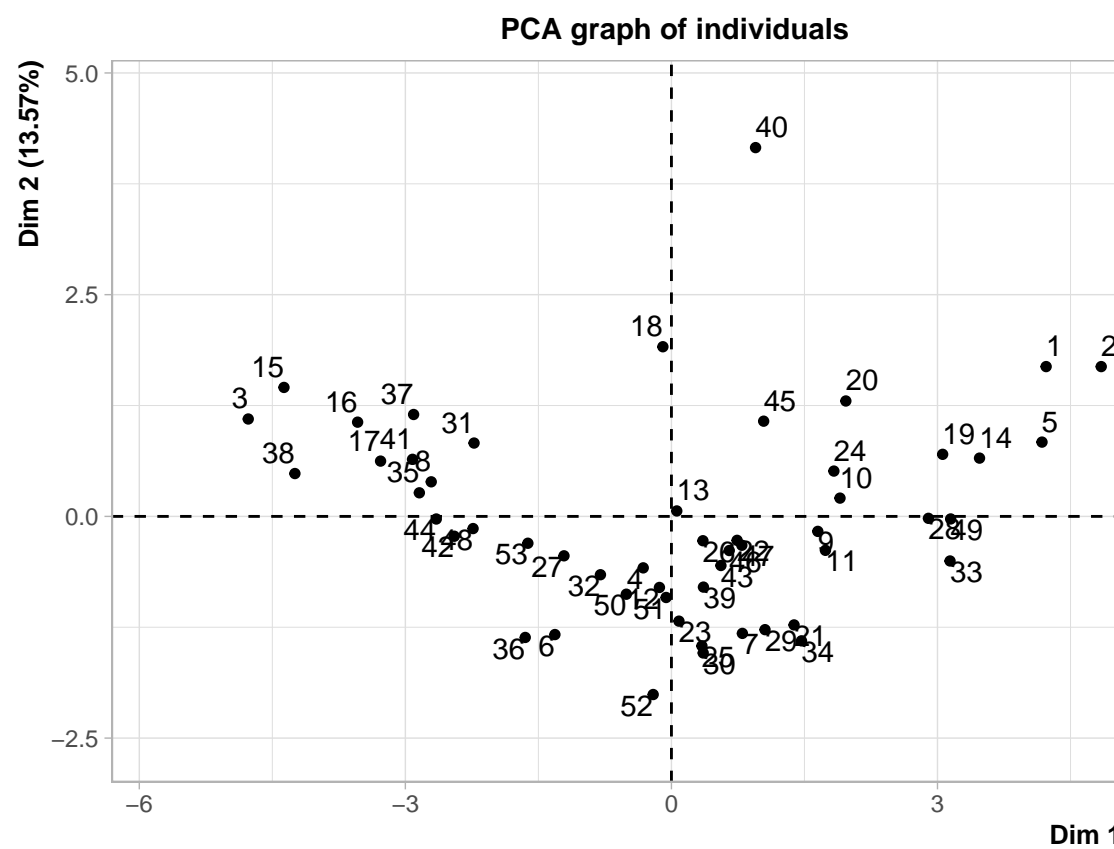
```
##
```

```
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
```

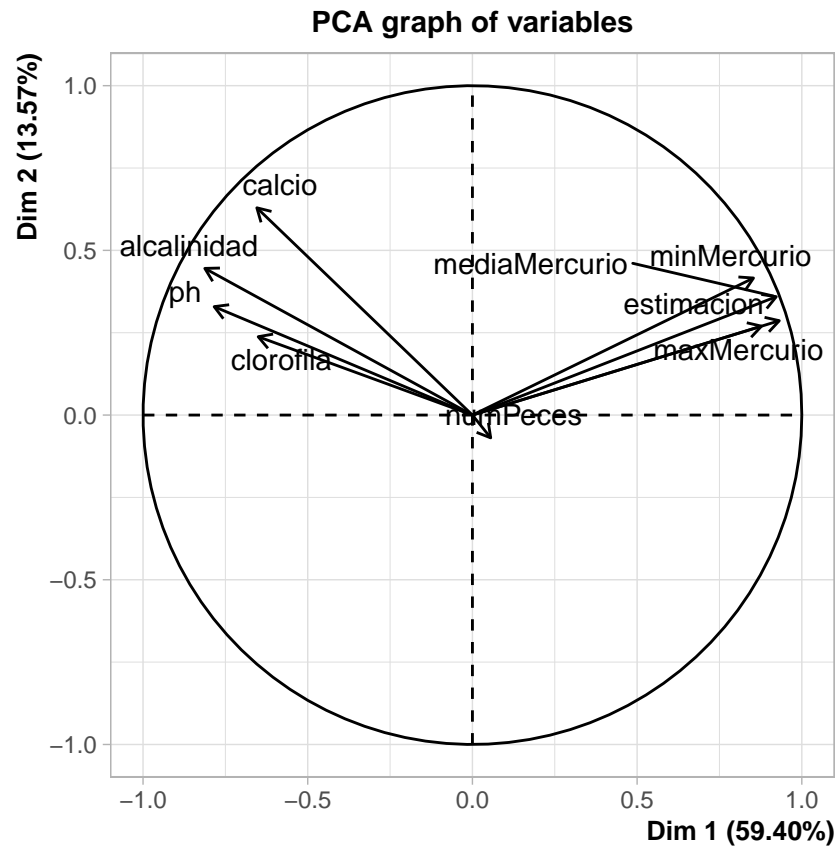
```
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
```

```
## Proportion Var 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
```

```
## Cumulative Var 0.111 0.222 0.333 0.444 0.556 0.667 0.778 0.889 1.000
```

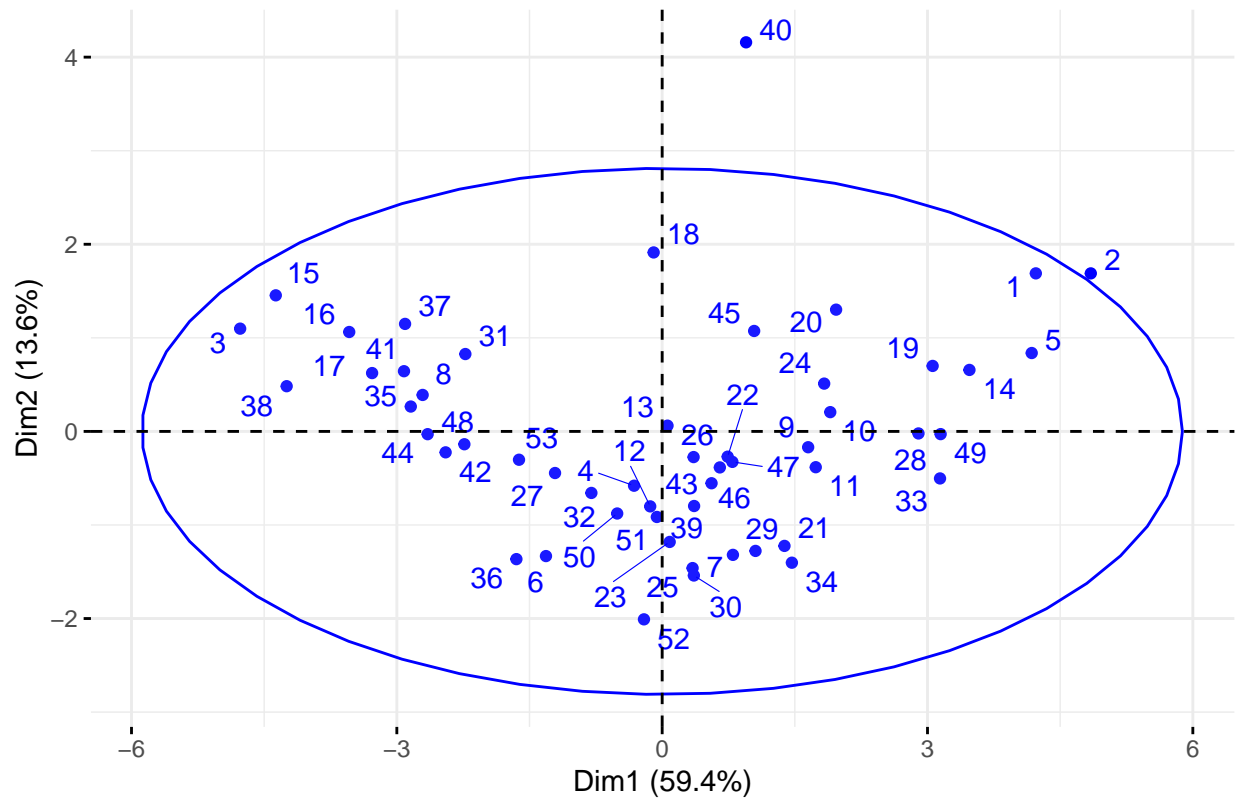


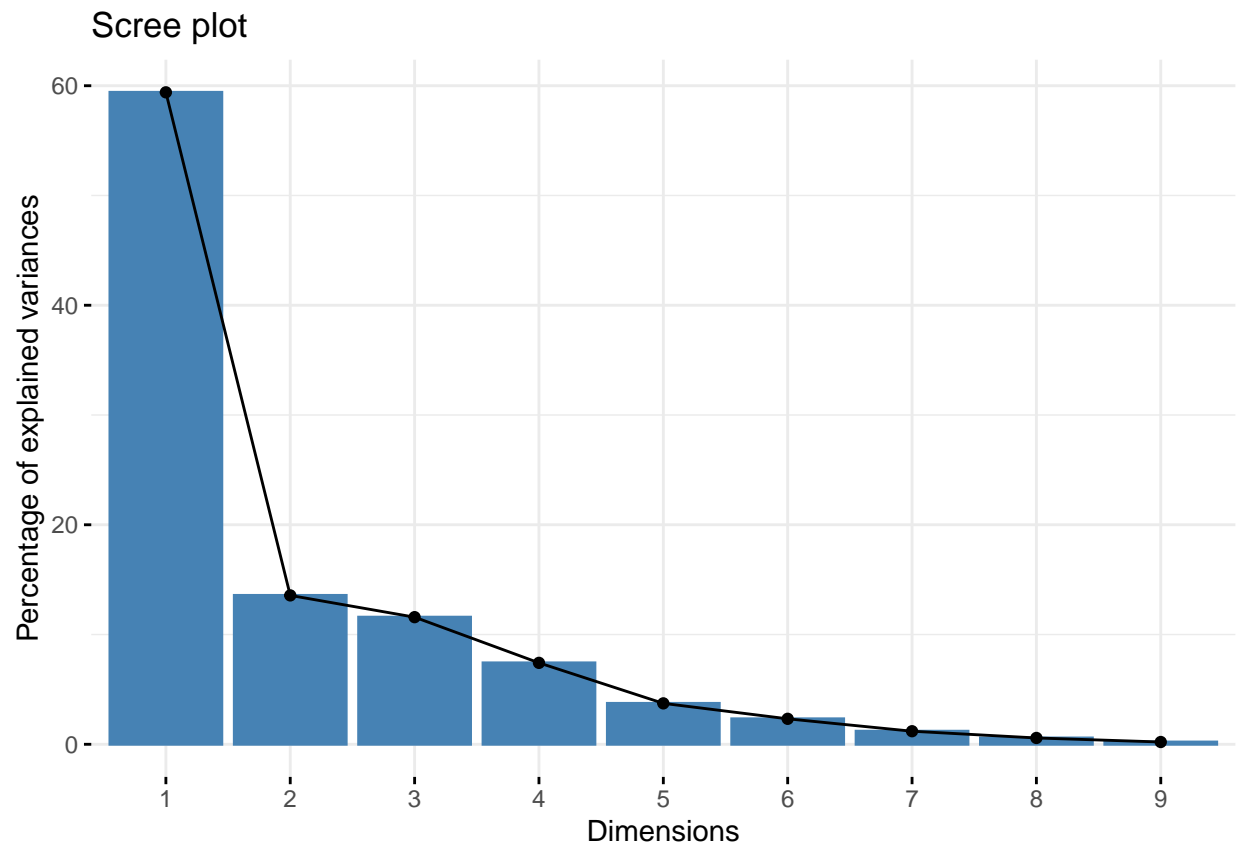
Observemoslo graficamente:

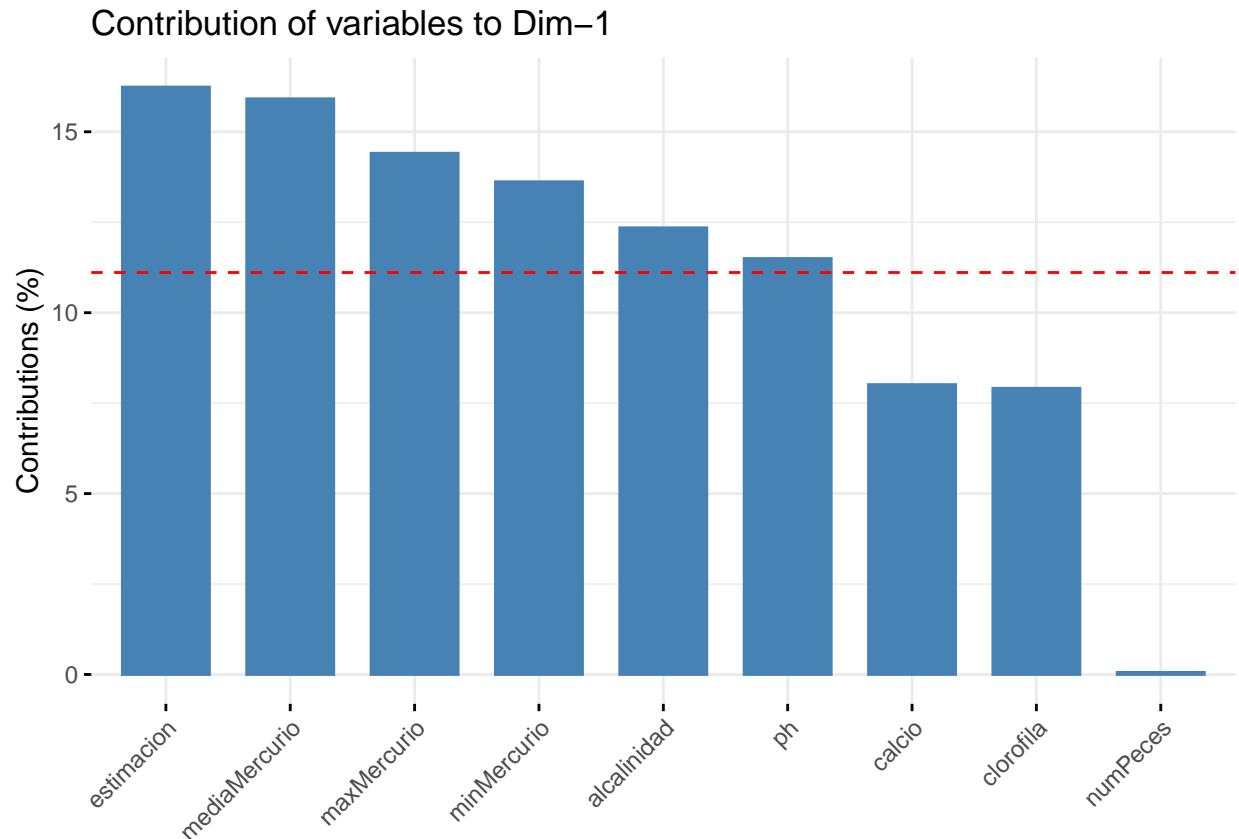




Individuals – PCA







Esta prueba nos sirvió para ver las relaciones entre variables. Hay que tomar en consideración que esto es para el primer y segundo componente solamente (los que más contribuyen). En cuanto a variabilidad explicada vemos que el primero es el que más nos da por una buena cantidad y sus pesos son mayormente alcalinidad, ph, calcio y clorofila. Esto cuadra con nuestro reporte anterior en el cual concluimos que alcalinidad y clorofila eran los más relacionados al nivel de mercurio.

## Conclusión

Utilizando las herramientas que aprendimos, pudimos extender nuestro reporte pasado a uno mucho más extenso, utilizando los análisis de normalidad y componentes principales. Este análisis en conjunto con los otros nos ayudaron a determinar que factores realmente afectan el nivel del mercurio en el agua de los lagos de Florida. Al igual que nuestro reporte pasado, pudimos observar que la alcalinidad y la clorofila son grandes factores, también contribuimos en este reporte que el ph y el calcio son factores. Sin embargo, en ambos reportes observamos que la alcalinidad es el factor #1 en el nivel de mercurio, por lo tanto, si se tuviera que tomar acción justo ahora sería atacar primeramente los niveles de alcalinidad. En el reporte pasado tocamos varios métodos para tratar justo esto en el agua.

## Bibliografía

Alcalinidad - Productos y descripción general de los parámetros de calidad del agua | Hach. (2022). Retrieved 27 November 2022, from <https://es.hach.com/parameters/alkalinity#:~:text=La%20alcalinidad%20es%20una%20medida,pH>

## **Anexos**

Drive: [https://drive.google.com/drive/folders/16Y6\\_cbbXaWo\\_AuxQGE0QRvrL4VdnImkS?usp=share\\_link](https://drive.google.com/drive/folders/16Y6_cbbXaWo_AuxQGE0QRvrL4VdnImkS?usp=share_link)