

RetoMercurio2

Omar Uresti

2022-11-28

La problemática presentada en este reto se describe con un dataset de la presencia de mercurio en peces de distintos lagos de Florida. Este dataset presenta distintos estudios y distintas variables que nos pueden indicar las posibles razones por las cuales esto se presente.

La solución que se realiza es que se desarrollará un modelo que pueda explicar de qué forma las distintas sustancias en los lagos puedan influenciar la presencia del mercurio en peces por medio de un modelo de regresión.

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)

#Análisis de los resultados

Variables = 12

Rows = 53

##Variables cualitativas

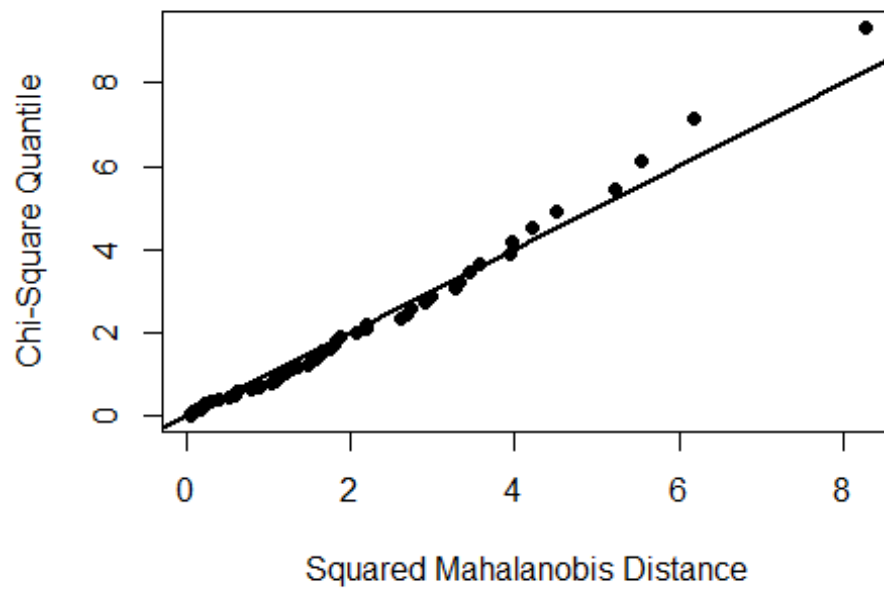
| | | | | |
|----|-------------------|--------------|--------------|--------------|
| ## | | | | |
| ## | Alligator | Annie | Apopka | Blue Cypress |
| ## | 1 | 1 | 1 | 1 |
| ## | Brick | Bryant | Cherry | Crescent |
| ## | 1 | 1 | 1 | 1 |
| ## | Deer Point | Dias | Dorr | Down |
| ## | 1 | 1 | 1 | 1 |
| ## | East Tohopekaliga | Eaton | Farm-13 | George |
| ## | 1 | 1 | 1 | 1 |
| ## | Griffin | Harney | Hart | Hatchineha |
| ## | 1 | 1 | 1 | 1 |
| ## | Iamonia | Istokpoga | Jackson | Josephine |
| ## | 1 | 1 | 1 | 1 |
| ## | Kingsley | Kissimmee | Lochloosa | Louisa |
| ## | 1 | 1 | 1 | 1 |
| ## | Miccasukee | Minneola | Monroe | Newmans |
| ## | 1 | 1 | 1 | 1 |
| ## | Ocean Pond | Ocheese Pond | Okeechobee | Orange |
| ## | 1 | 1 | 1 | 1 |
| ## | Panasoffkee | Parker | Placid | Puzzle |
| ## | 1 | 1 | 1 | 1 |
| ## | Rodman | Rousseau | Sampson | Shipp |
| ## | 1 | 1 | 1 | 1 |
| ## | Talquin | Tarpon | Tohopekaliga | Trafford |
| ## | 1 | 1 | 1 | 1 |
| ## | Trout | Tsala Apopka | Weir | Wildcat |
| ## | 1 | 1 | 1 | 1 |
| ## | Yale | | | |
| ## | 1 | | | |

A continuación se observa el resultado de la prueba de normalidad univariante

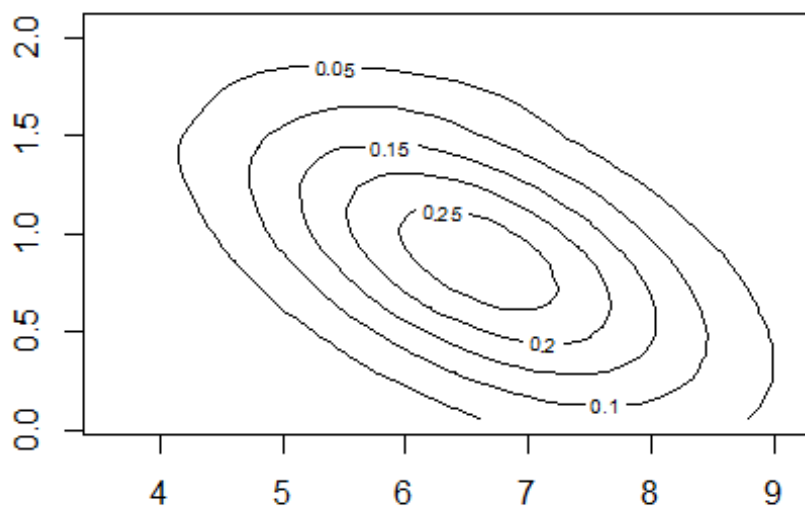
| Test | Variable | Statistic | p value | Normality | |
|------|------------------|-----------|---------|-----------|-----|
| ## 1 | Anderson-Darling | X3 | 3.6725 | <0.001 | NO |
| ## 2 | Anderson-Darling | X4 | 0.3496 | 0.4611 | YES |
| ## 3 | Anderson-Darling | X5 | 4.0510 | <0.001 | NO |
| ## 4 | Anderson-Darling | X6 | 5.4286 | <0.001 | NO |
| ## 5 | Anderson-Darling | X7 | 0.9253 | 0.0174 | NO |
| ## 6 | Anderson-Darling | X9 | 1.9770 | <0.001 | NO |
| ## 7 | Anderson-Darling | X10 | 0.6585 | 0.081 | YES |
| ## 8 | Anderson-Darling | X11 | 1.0469 | 0.0086 | NO |

X4 y X10 fueron quienes presentaron normalidad.

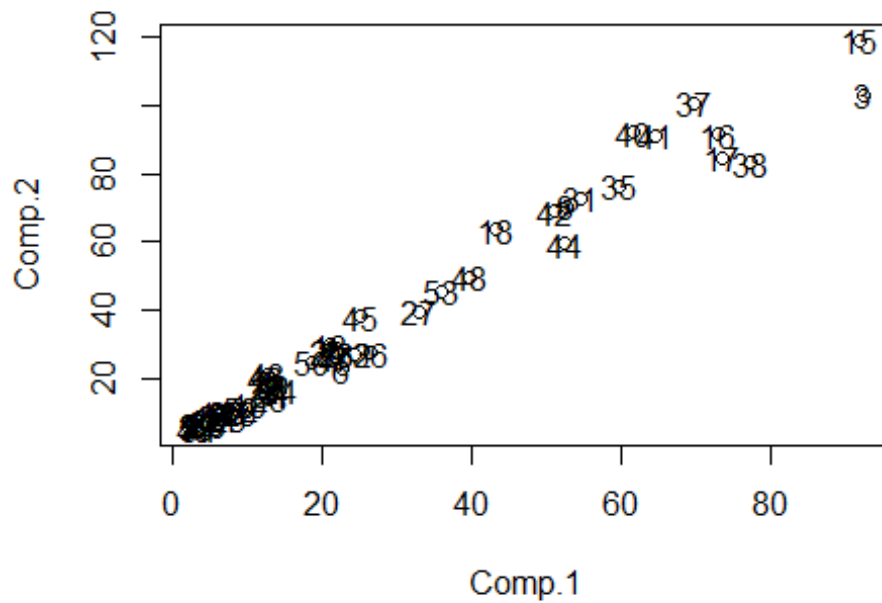
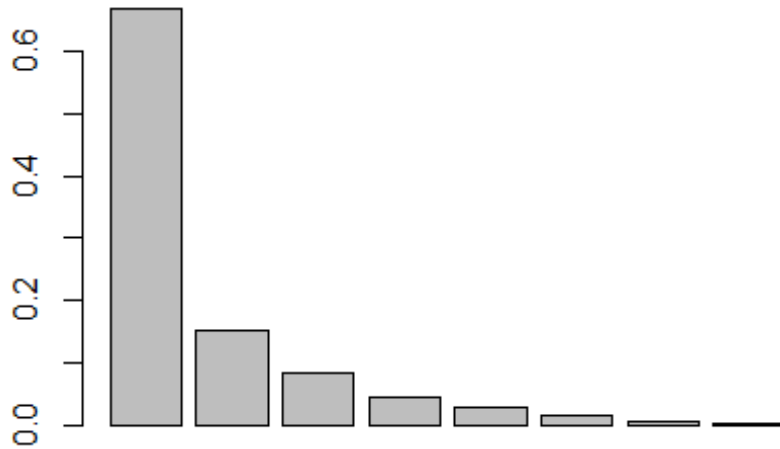
Chi-Square Q-Q Plot



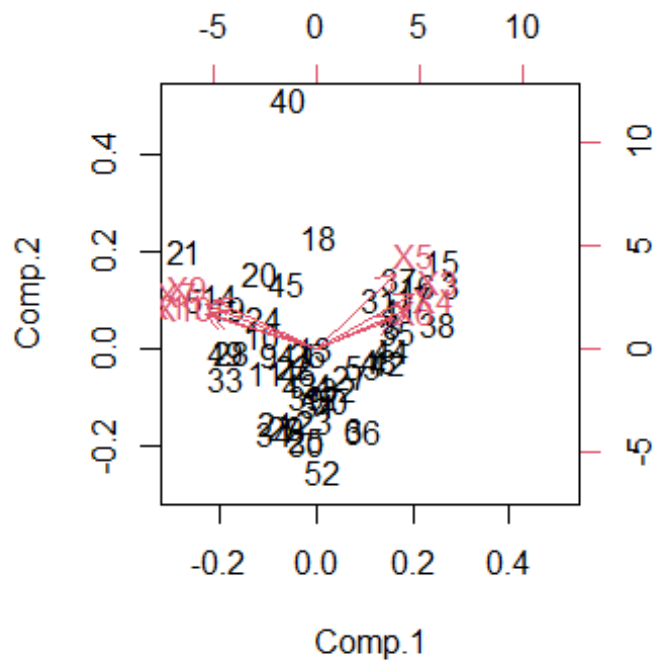
Contorno de la normal multivariada



Los componentes principales nos pueden dar indicio de la importancia de estos como factores a producir algún tipo de contaminación, haciendo que nos sea mas fácil identificar la ruta para poder solucionar esta problemática de contaminación.



Vectores asociados a las variables



El uso de las diferentes herramientas estadísticas utilizadas durante toda esta evidencia nos ayuda a conocer de mejor manera como las variables funcionan como factores para producir alguna anomalía o situación dentro de este dataset. Encontramos que la mayoría de la contaminación se debe en su mayoría a un único factor, lo que nos ayuda a poder actuar de mejor y más eficiente manera a la hora de intentar solucionar las problemáticas que se presentan.