

Momento de Retroalimentación: Módulo 5

Procesamiento de datos multivariados (Portafolio Implementación)

Significado de variables

- 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 = estimacion (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)

3) 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.

```
## [1] "-----MARDIA-----"
```

```
## $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   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   X8          8.6943 <0.001        NO
## 7 Anderson-Darling   X9          1.9770 <0.001        NO
## 8 Anderson-Darling  X10          0.6585 0.081         YES
## 9 Anderson-Darling  X11          1.0469 0.0086        NO
##
## $Descriptives
##           n      Mean      Std.Dev Median  Min    Max  25th  75th      Skew
## X3  53  37.5301887  38.2035267  19.60  1.20  128.00  6.60  66.50  0.9679170
## X4  53   6.5905660  1.2884493   6.80  3.60   9.10  5.80   7.40 -0.2458771
## X5  53  22.2018868  24.9325744  12.60  1.10  90.70  3.30  35.60  1.3045868
## X6  53  23.1169811  30.8163214  12.80  0.70  152.40  4.60  24.70  2.4130571
## X7  53   0.5271698  0.3410356   0.48  0.04   1.33  0.27  0.77  0.5986343
## X8  53  13.0566038  8.5606773  12.00  4.00  44.00 10.00  12.00  2.5808773
## X9  53   0.2798113  0.2264058   0.25  0.04   0.92  0.09  0.33  1.0729099
## X10 53   0.8745283  0.5220469   0.84  0.06   2.04  0.48  1.33  0.4645925
## X11 53   0.5132075  0.3387294   0.45  0.04   1.53  0.25  0.70  0.9449951
##           Kurtosis
## X3  -0.4705349
## X4  -0.6239638
## X5   0.6130359
## X6   6.1042185
## X7  -0.6312607
## X8   6.0089455
## X9   0.4060828
## X10 -0.6692490
## X11  0.5733500
```

```
## [1] "----- ANDERSON DARLING-----"
```

```
## $multivariateNormality
##           Test      HZ p value MVN
## 1 Henze-Zirkler 1.421134      0 NO
##
## $univariateNormality
##           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   X8      8.6943 <0.001      NO
## 7 Anderson-Darling   X9      1.9770 <0.001      NO
## 8 Anderson-Darling  X10      0.6585  0.081      YES
## 9 Anderson-Darling  X11      1.0469  0.0086      NO
##
## $Descriptives
##      n      Mean   Std.Dev Median  Min    Max  25th  75th      Skew
## X3  53 37.5301887 38.2035267  19.60 1.20 128.00  6.60 66.50  0.9679170
## X4  53  6.5905660  1.2884493   6.80 3.60   9.10  5.80  7.40 -0.2458771
## X5  53 22.2018868 24.9325744  12.60 1.10  90.70  3.30 35.60  1.3045868
## X6  53 23.1169811 30.8163214  12.80 0.70 152.40  4.60 24.70  2.4130571
## X7  53  0.5271698  0.3410356   0.48 0.04   1.33  0.27  0.77  0.5986343
## X8  53 13.0566038  8.5606773  12.00 4.00  44.00 10.00 12.00  2.5808773
## X9  53  0.2798113  0.2264058   0.25 0.04   0.92  0.09  0.33  1.0729099
## X10 53  0.8745283  0.5220469   0.84 0.06   2.04  0.48  1.33  0.4645925
## X11 53  0.5132075  0.3387294   0.45 0.04   1.53  0.25  0.70  0.9449951
##      Kurtosis
## X3  -0.4705349
## X4  -0.6239638
## X5   0.6130359
## X6   6.1042185
## X7  -0.6312607
## X8   6.0089455
## X9   0.4060828
## X10 -0.6692490
## X11  0.5733500
```

B. Realiza la prueba de Mardia y Anderson Darling de las variables que sí tuvieron normalidad en los incisos anteriores. Interpreta los resultados obtenidos con base en ambas pruebas y en la interpretación del sesgo y la curtosis de cada una de ellas.

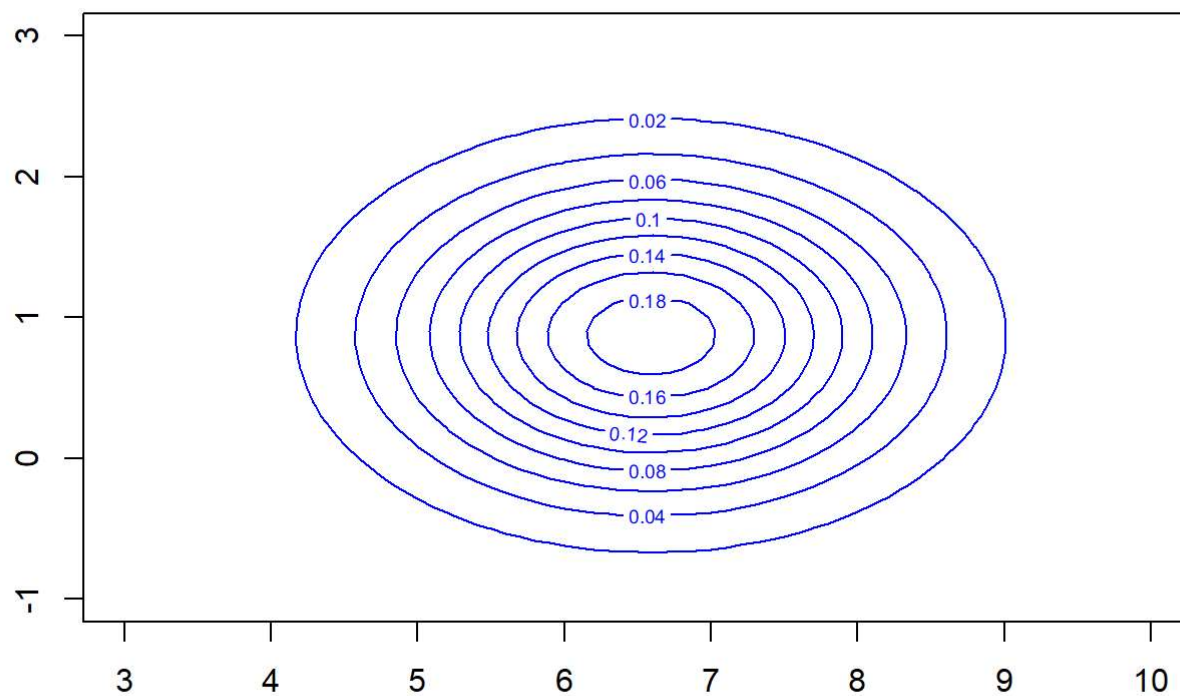
```
## [1] "-----MARDIA PARA VARIABLES X4 Y X10-----"
```

```
## $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   X4      0.3496    0.4611    YES
## 2 Anderson-Darling   X10     0.6585    0.0810    YES
##
## $Descriptives
##      n      Mean  Std.Dev Median  Min  Max 25th 75th      Skew  Kurtosis
## X4  53 6.5905660 1.2884493   6.80 3.60 9.10 5.80 7.40 -0.2458771 -0.6239638
## X10 53 0.8745283 0.5220469   0.84 0.06 2.04 0.48 1.33 0.4645925 -0.6692490
```

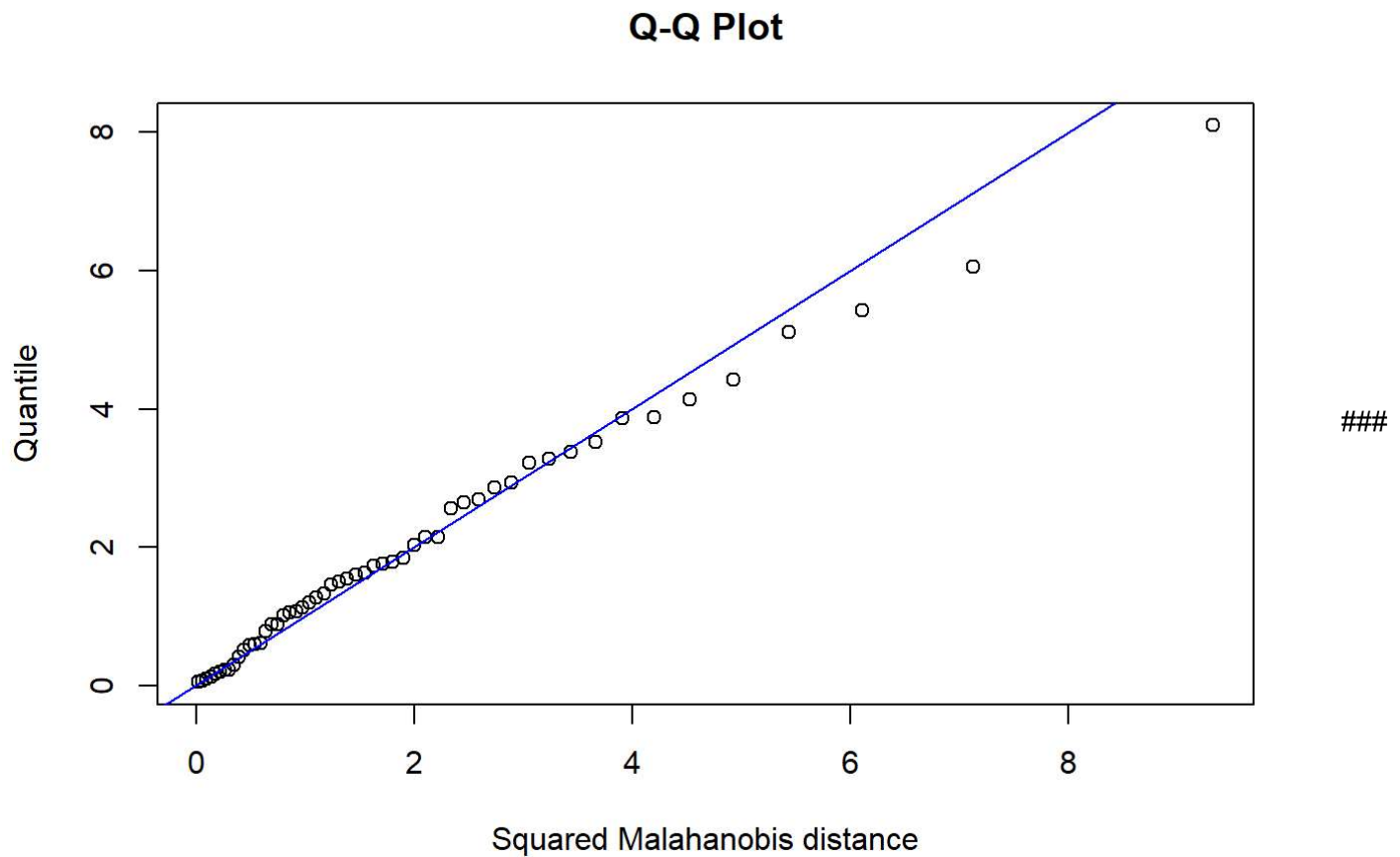
```
## [1] "----- ANDERSON DARLING PARA VARIABLES X4 Y X10-----"
```

```
## $multivariateNormality
##           Test      HZ    p value MVN
## 1 Henze-Zirkler 0.7543838 0.1124305 YES
##
## $univariateNormality
##           Test Variable Statistic  p value Normality
## 1 Anderson-Darling   X4      0.3496    0.4611    YES
## 2 Anderson-Darling   X10     0.6585    0.0810    YES
##
## $Descriptives
##      n      Mean  Std.Dev Median  Min  Max 25th 75th      Skew  Kurtosis
## X4  53 6.5905660 1.2884493   6.80 3.60 9.10 5.80 7.40 -0.2458771 -0.6239638
## X10 53 0.8745283 0.5220469   0.84 0.06 2.04 0.48 1.33 0.4645925 -0.6692490
```

C. Haz la gráfica de contorno de la normal multivariada obtenida en el inciso B.



D. Detecta datos atípicos o influyentes en la normal multivariada encontrada en el inciso B (auxíliate de la distancia de Mahalanobis y del gráfico QQplot multivariado)



Realice 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. Tome en cuenta los puntos que se sugieren a continuación (no son exhaustivos):

- A. Justifique por qué es adecuado el uso de componentes principales para analizar la base (haz uso de la matriz de correlaciones)
- B. Realiza el análisis de componentes principales y justifica el número de componentes principales apropiados para reducir la dimensión de la base
- C. Representa en un gráfico los vectores asociados a las variables y las puntuaciones de las observaciones de las dos primeras componentes
- D. Interprete los resultados. Explique brevemente a qué conclusiones llega con su análisis y qué significado tienen los componentes seleccionados en el contexto del problema