Datos multivariados

Ileana Parra

2022-10-19

```
M=read.csv("C:/Users/ilean/Desktop/ITESM/7to Semestre/Parte
1/mercurio.csv") #leer la base de datos
D=matrix(c(M$X3,M$X4,M$X5,M$X6,M$X7,M$X8,M$X9,M$X10,M$X11),ncol=9)
```

##Prueba de normalidad

```
library(MVN)
## Test de Multinomalidad: Método Sesgo y kurtosis de Mardia
mvn(D,subset = NULL,mvn = "mardia", covariance = FALSE,
alpha=0.05,showOutliers = FALSE)
```

```
## $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  Column1     3.6725  <0.001      NO
## 2 Anderson-Darling  Column2     0.3496  0.4611      YES
## 3 Anderson-Darling  Column3     4.0510  <0.001      NO
## 4 Anderson-Darling  Column4     5.4286  <0.001      NO
## 5 Anderson-Darling  Column5     0.9253  0.0174      NO
## 6 Anderson-Darling  Column6     8.6943  <0.001      NO
## 7 Anderson-Darling  Column7     1.9770  <0.001      NO
## 8 Anderson-Darling  Column8     0.6585   0.081      YES
## 9 Anderson-Darling  Column9     1.0469  0.0086      NO
##
## $Descriptives
##    n       Mean    Std.Dev Median  Min    Max  25th  75th       Skew
Kurtosis
## 1 53 37.5301887 38.2035267  19.60 1.20 128.00  6.60 66.50  0.9679170 -
0.4705349
## 2 53  6.5905660  1.2884493   6.80 3.60   9.10  5.80  7.40 -0.2458771 -
0.6239638
## 3 53 22.2018868 24.9325744  12.60 1.10  90.70  3.30 35.60  1.3045868
0.6130359
## 4 53 23.1169811 30.8163214  12.80 0.70 152.40  4.60 24.70  2.4130571
6.1042185
## 5 53  0.5271698  0.3410356   0.48 0.04   1.33  0.27  0.77  0.5986343 -
0.6312607
```

```
## 6 53 13.0566038  8.5606773  12.00 4.00  44.00 10.00 12.00  2.5808773
6.0089455
## 7 53  0.2798113  0.2264058   0.25 0.04   0.92  0.09  0.33  1.0729099
0.4060828
## 8 53  0.8745283  0.5220469   0.84 0.06   2.04  0.48  1.33  0.4645925 -
0.6692490
## 9 53  0.5132075  0.3387294   0.45 0.04   1.53  0.25  0.70  0.9449951
0.5733500
```

La prueba de Mardia indica un sesgo de 410.214 lo que significa que los datos presentan un sesgo a la derecha, también nos indica que contamos con una distribución platicútica ya que la curtosis es mayor a 3. La prueba de Anderson-Darling nos indica que no tenemos normalidad multivariada ya que solo dos variables son normales.

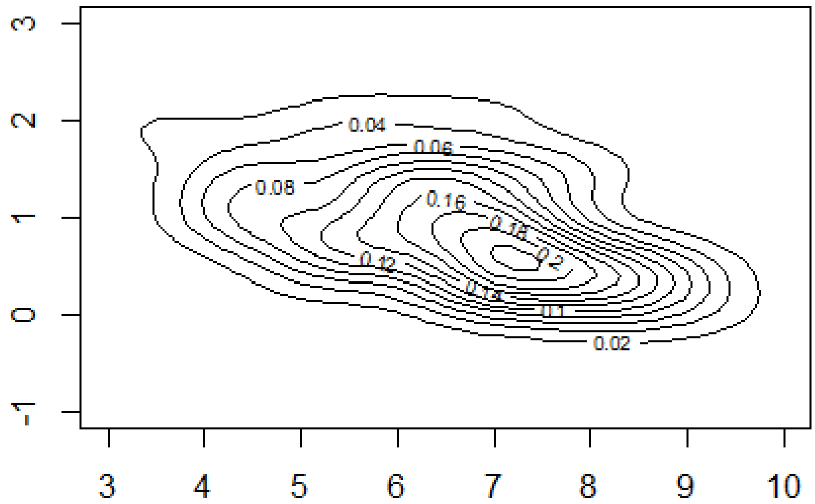
```
B=matrix(c(D[,2],D[,8]),ncol=2)
library(MVN)
## Test de Multinomalidad: Método Sesgo y kurtosis de Mardia
mvn(B,subset = NULL,mvn = "mardia", covariance = FALSE,
alpha=0.05,showOutliers = FALSE)
```

```
## $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  Column1     0.3496    0.4611    YES
## 2 Anderson-Darling  Column2     0.6585    0.0810    YES
##
## $Descriptives
##    n      Mean   Std.Dev Median  Min  Max 25th 75th       Skew
Kurtosis
## 1 53 6.5905660 1.2884493   6.80 3.60 9.10 5.80 7.40 -0.2458771 -
0.6239638
## 2 53 0.8745283 0.5220469   0.84 0.06 2.04 0.48 1.33  0.4645925 -
0.6692490
```

La prueba de Mardia indica un sesgo a la derecha y una distribución leptocúrtica. La prueba de Anderson-Darling nos indica que sí hay normalidad multivariada ya que ambas variables son normales.

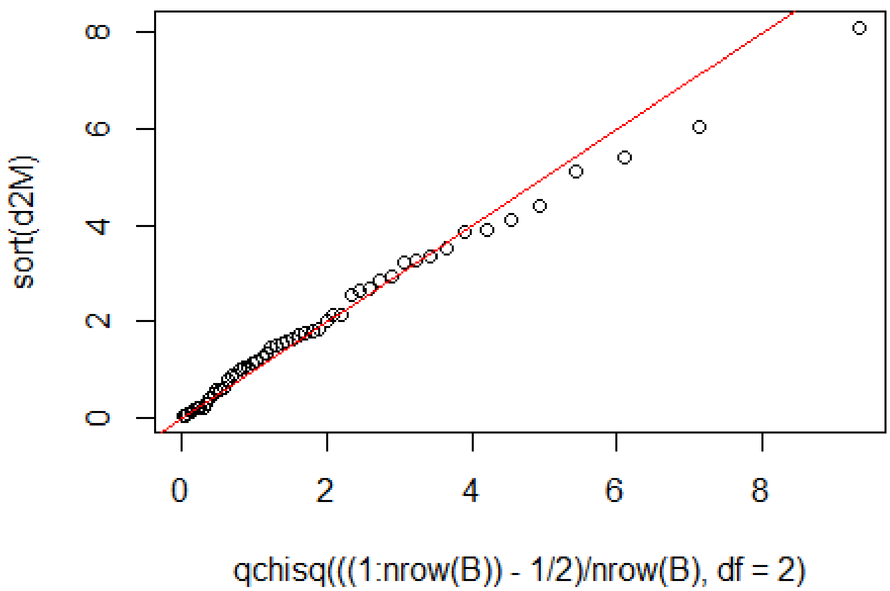
##Gráfica de contornos

```
library(MASS)
B.kde <- kde2d(B[,1], B[,2], n = 53,lims = c(range(3,10),range(-1,3)))
contour(B.kde)
```



##Datos atípicos o influyentes

```
#Distancia de Mahalanobis
X=colMeans(B)
S=cov(B)
d2M =  mahalanobis(B,X,S)
plot(qchisq(((1:nrow(B)) - 1/2)/nrow(B),df=2),sort( d2M ) )
abline(a=0, b=1,col="red")
```



De acuerdo con la gráfica, los datos no se comportan como una normal, ya que presentan un sesgo. Los puntos que se alejan de la gráfica se identifican como outliers.

#Componentes principales. ##Matriz de correlaciones

```
cor(D)

##              [,1]        [,2]        [,3]        [,4]        [,5]
[,6]
##  [1,]  1.00000000  0.71916568  0.83260419  0.47753085 -0.59389671
0.01029074
##  [2,]  0.71916568  1.00000000  0.57713272  0.60848276 -0.57540012 -
0.01860607
##  [3,]  0.83260419  0.57713272  1.00000000  0.40991385 -0.40067958 -
0.08937901
##  [4,]  0.47753085  0.60848276  0.40991385  1.00000000 -0.49137481 -
0.01182027
##  [5,] -0.59389671 -0.57540012 -0.40067958 -0.49137481  1.00000000
0.07903426
##  [6,]  0.01029074 -0.01860607 -0.08937901 -0.01182027  0.07903426
1.00000000
##  [7,] -0.52535654 -0.54196524 -0.33247623 -0.40045856  0.92720506 -
0.08165278
##  [8,] -0.60479558 -0.55181523 -0.40791663 -0.48497215  0.91586397
0.16109174
##  [9,] -0.62795845 -0.61284905 -0.46440947 -0.50644193  0.95921481
0.02580046
```

```
##               [,7]       [,8]        [,9]
##  [1,] -0.52535654 -0.6047956 -0.62795845
##  [2,] -0.54196524 -0.5518152 -0.61284905
##  [3,] -0.33247623 -0.4079166 -0.46440947
##  [4,] -0.40045856 -0.4849721 -0.50644193
##  [5,]  0.92720506  0.9158640  0.95921481
##  [6,] -0.08165278  0.1610917  0.02580046
##  [7,]  1.00000000  0.7653532  0.91908939
##  [8,]  0.76535319  1.0000000  0.85975810
##  [9,]  0.91908939  0.8597581  1.00000000
```

Como se puede ver en la matriz de correlaciones, hay una fuerte correlación entre la mayoría de las variables, por lo que el uso de componentes principales es adecuado.

```
cpa <- prcomp(D, scale=FALSE)

print("desviaciones estándar: ")

## [1] "desviaciones estándar: "

cpa$sdev

## [1] 47.50220005 25.15208924 12.14055875  8.29851209  0.80849910
0.50296710
## [7]  0.17379240  0.06820524  0.04412128

print("medias: ")

## [1] "medias: "

print("center y scale dan las medias y desv estándar previa
estandarización: ")

## [1] "center y scale dan las medias y desv estándar previa
estandarización: "

cpa$center

## [1] 37.5301887  6.5905660 22.2018868 23.1169811  0.5271698 13.0566038
0.2798113
## [8]  0.8745283  0.5132075

cpa$scale

## [1] FALSE

print("Los coeficientes de la combinación lineal normalizada de
componete")

## [1] "Los coeficientes de la combinación lineal normalizada de
componete"

cpa$rotation
```

```
##                  PC1           PC2          PC3           PC4
PC5
##  [1,]  0.770052707  0.3595628867  0.512208576  1.212294e-01
0.023208046
##  [2,]  0.020607444 -0.0064784700  0.013447171  7.835184e-03 -
0.989042929
##  [3,]  0.459104487  0.2605992620 -0.824602008 -2.030053e-01 -
0.006231571
##  [4,]  0.442395277 -0.8959627956 -0.034881281 -8.630735e-03
0.015103512
##  [5,] -0.004349946  0.0015154350 -0.006280700 -6.265597e-03
0.070456679
##  [6,] -0.003461124 -0.0017240484  0.236858719 -9.714347e-01 -
0.005749934
##  [7,] -0.002482186  0.0006039179 -0.004911710  6.851298e-05
0.064182335
##  [8,] -0.006732316  0.0020103775 -0.009275875 -1.488313e-02
0.073080876
##  [9,] -0.004611180  0.0012562489 -0.004970505 -3.214838e-03
0.080815413
##                  PC6           PC7           PC8           PC9
##  [1,]  0.011728199 -0.0017253286  0.0001564783  6.253514e-05
##  [2,]  0.140012239  0.0375235348 -0.0004638668 -8.577476e-03
##  [3,] -0.009417996  0.0016136975 -0.0011527025 -8.392504e-05
##  [4,]  0.004138600 -0.0006108642 -0.0003813870  3.630229e-04
##  [5,]  0.472098416  0.2823081524  0.3073272995  7.732471e-01
##  [6,] -0.010565321  0.0070248213  0.0006326070 -1.805388e-03
##  [7,]  0.295149427  0.4661481465  0.5865444157 -5.894136e-01
##  [8,]  0.693632901 -0.6928389806  0.0108480697 -1.817475e-01
##  [9,]  0.434654887  0.4706578821 -0.7492634949 -1.468724e-01
```

```
print("Los datos por sustituidos en la combinación lineal de vectores
propios:")
```

```
## [1] "Los datos por sustituidos en la combinación lineal de vectores
propios:"
```

```
cpa$x
```

```
##               PC1         PC2         PC3         PC4         PC5
PC6
##  [1,] -43.084877    3.7286132  -1.5177014   8.0637715 -0.21259430
1.06088173
##  [2,] -44.362410    0.3430139  -1.4712102   6.0168716  0.78358632
1.19924426
##  [3,] 117.098400  -60.3258141  16.8432357  11.0509784  0.68445651
0.50385250
##  [4,]  -9.890924   16.7358656   6.1821090   2.6039299 -0.54090044 -
0.09126671
##  [5,] -45.316492    1.4918771  -1.5759604   0.8503810  1.14561992
0.64276097
```

```
## [6,] -12.634479 -29.8680141 4.9218529 0.3362359 -0.77959659 -
0.44762194
## [7,] -42.538610 0.9973616 -0.6112599 3.1522516 0.24945538 -
0.56341800
## [8,] 45.954683 11.2855932 -10.4515781 -1.6352720 -0.84782217 -
0.36193877
## [9,] -24.118995 11.8761475 8.3446449 -9.1715195 0.29188903
0.26804122
## [10,] -33.565226 -15.7976743 -2.4908540 0.6240893 -0.36889541
0.47782653
## [11,] -36.429512 -8.8312410 0.2557589 1.2958050 0.50632880 -
0.13595908
## [12,] -28.491192 7.3743321 -3.4164658 0.3546602 -1.34540361 -
0.31768505
## [13,] -13.026799 6.7452584 -9.7111331 3.9067309 -1.03390384
0.01337051
## [14,] -39.032695 0.1006494 6.1007547 -29.2050607 -0.06959024
0.68132042
## [15,] 120.451410 6.2886469 -8.8146308 -0.4777808 1.29479504 -
0.21055584
## [16,] 80.488210 -21.5733920 -15.5059979 -0.8787155 -0.05771199 -
0.25330137
## [17,] 85.967334 -22.1070712 29.7346166 -20.7555348 0.08715085
0.28274930
## [18,] 30.430158 25.9833677 -18.1903972 2.6694979 -0.89898165
0.70328872
## [19,] -40.534693 0.6668664 -1.0365406 3.0322505 0.04707431
0.55480467
## [20,] -11.022059 1.2695179 1.1263590 7.5672244 -0.15335262
0.85816604
## [21,] -38.421592 -3.9353945 1.4674704 1.5852051 1.40684493 -
0.43906015
## [22,] -26.878217 1.9303452 -0.6530519 1.0258334 -0.68655968 -
0.03789898
## [23,] -28.632013 -11.8846696 2.3080661 1.7795595 -0.04183420 -
0.47687842
## [24,] -26.836625 -23.1671229 -3.0983471 0.4587599 -0.64707415
0.93134558
## [25,] -37.643171 5.4267464 -0.7163289 3.1030261 0.20175131 -
0.69092397
## [26,] -10.400333 -3.4632675 8.4811217 -21.5032877 -0.56488958 -
0.02337384
## [27,] 11.597402 3.3640836 13.5859715 6.4139444 -0.26874903 -
0.04790420
## [28,] -41.737794 -2.5533827 -2.3127197 4.7844889 1.29833342
0.22254647
## [29,] -37.786279 -9.3929011 0.2790723 2.3353288 1.03798982 -
0.47957021
## [30,] -42.647233 3.9393748 -0.3589747 3.2117902 -0.18202832 -
0.62107140
```

```
## [31,]  48.593919    1.5464413 -16.3423487  -1.0095242 -0.47627064 -
0.17375454
## [32,]  -7.962985  -14.8536743   4.3820955   4.2758537 -0.81169905 -
0.12839542
## [33,] -42.771601    1.0912829   1.1265447   1.4919939  2.18491263
0.36905188
## [34,] -43.980346    0.4839801   1.1325364   0.4867638  1.24581722 -
0.44924075
## [35,]  67.183613   39.8287190  28.4462226   7.7095077  0.28798133
0.18699275
## [36,]  -5.706088  -27.6451131   4.0714859   1.1305018 -0.47590403 -
0.63162101
## [37,]  81.642111   48.5748321 -20.9272008  -5.4200589  0.76878573 -
0.53374797
## [38,]  79.927960 -104.1720974 -17.9865770   4.8384033  0.28082684 -
0.22628869
## [39,] -35.950572   -6.3301461   1.4939340   1.6026721 -1.12042533 -
0.28704021
## [40,]  66.302363   37.2038944 -27.1690257  -3.7901999 -0.01719698
1.09443672
## [41,]  74.641310   55.6005761  -1.5669846  -1.7171546  0.68469832 -
0.41483029
## [42,]  49.408220   42.7894125  11.8591357   3.7273771  0.67415146 -
0.38184527
## [43,] -28.419535   10.5574274 -14.8105204  -0.3773763 -0.23105185 -
0.40930253
## [44,]  44.041646  -28.9975623   9.9153579   3.4006498 -0.44952539
0.09633345
## [45,]  -7.422373   -3.6683920 -26.9850506  -5.4584190 -0.62570415
0.21255060
## [46,] -30.393041    0.7829681 -17.6001430  -3.0886108 -0.56554510 -
0.42028300
## [47,] -11.685195  -10.9481942   8.9677957 -29.6034968  0.09236774 -
0.16275989
## [48,]  27.175925   27.4725243  22.7646242  12.6751502 -1.50449424
0.24662856
## [49,] -44.644342   -3.3014880  -2.2140799   2.7626215  1.42610023
0.14847895
## [50,] -15.075064   12.9479080   6.1009310   2.6102706 -0.74009867 -
0.15349448
## [51,] -27.681703   -6.4224603   2.4865639   2.8417763 -0.82974017 -
0.28531951
## [52,] -33.489801    6.1135142   5.9264711   2.6477342  0.64029544 -
0.92454327
## [53,]  19.310202   24.6979312  19.2303499   5.6681213 -0.77366912
0.02622294
##                 PC7           PC8           PC9
##  [1,]  0.520192214 -0.1845147190 -0.0326610526
##  [2,]  0.138652171  0.0438954496 -0.0468256342
##  [3,] -0.027411883 -0.0033662159  0.0146534275
```

```
##  [4,] -0.101392788 -0.0465333338  0.0306604734
##  [5,]  0.292430205 -0.1326818135  0.0555234469
##  [6,] -0.012124996 -0.0168989714  0.0530479953
##  [7,]  0.043778072  0.0664718250 -0.0028256104
##  [8,]  0.025625583  0.0003192099 -0.0090373781
##  [9,] -0.131905712 -0.0390523607  0.0995723637
## [10,] -0.118475881 -0.0378096873 -0.0071047945
## [11,]  0.241083006  0.0701206993 -0.0176212910
## [12,]  0.056985636 -0.1002479415  0.1008432791
## [13,] -0.087854868 -0.0467242804 -0.0451050845
## [14,] -0.073822790  0.0121852979  0.0240105819
## [15,]  0.006972898 -0.0285658449 -0.0079038377
## [16,]  0.126128548 -0.0127258171 -0.0068520020
## [17,]  0.095668746  0.0006548949 -0.0515248990
## [18,] -0.339757615  0.0840111932  0.0513631339
## [19,]  0.231438968  0.0279917563  0.0631178117
## [20,] -0.036225444  0.2393054977 -0.0002638699
## [21,] -0.077907910  0.1330918902  0.0647404929
## [22,]  0.063436324  0.0207285687 -0.0552247501
## [23,]  0.076344870  0.0387575481 -0.0044536935
## [24,] -0.561425570 -0.1086446156 -0.1262163372
## [25,]  0.037467294  0.0129468667 -0.0605020990
## [26,]  0.005716546  0.0035524079 -0.0135745453
## [27,]  0.009142616  0.0336724237  0.0159302107
## [28,] -0.025975606  0.0366760913 -0.0645121700
## [29,] -0.030267884  0.0402004398 -0.0148174799
## [30,] -0.001255303 -0.0559446215 -0.0481140609
## [31,] -0.033210223  0.0001608443 -0.0286936505
## [32,] -0.002961041 -0.0463985235 -0.0156286843
## [33,] -0.517676135 -0.1060414482  0.0285925408
## [34,] -0.130886464 -0.0131457510  0.0210122434
## [35,] -0.035416732  0.0210739769 -0.0043492297
## [36,]  0.056785389  0.0451329682  0.0056576425
## [37,]  0.016993899 -0.0974671047  0.0024984184
## [38,]  0.028666420 -0.0250853391  0.0289516885
## [39,]  0.227557992 -0.0101845905 -0.0144767430
## [40,]  0.236135991  0.1324527827 -0.0159925606
## [41,]  0.035093934 -0.0498531363  0.0073316205
## [42,] -0.009171953 -0.0423597842 -0.0399298577
## [43,] -0.160542533  0.0367674782 -0.0499765821
## [44,] -0.104232562  0.0095957721  0.0188882100
## [45,] -0.095043929  0.0111812495  0.0899490205
## [46,]  0.021635789 -0.0132789181 -0.0442591753
## [47,]  0.124313456  0.0292009902 -0.0185015224
## [48,] -0.003349499 -0.0341545072  0.0553733240
## [49,]  0.160658438 -0.0139813220  0.0199199922
## [50,] -0.231676409  0.0123945147  0.0368706678
## [51,]  0.049814098  0.0184930562 -0.0013124549
## [52,]  0.037101500  0.0350342398 -0.0105787312
## [53,] -0.015848876  0.0495907144 -0.0296688039
```

```
S=cov(D)
lambda=eigen(S)
lambda$values

## [1] 2.256459e+03 6.326276e+02 1.473932e+02 6.886530e+01 6.536708e-01
## [6] 2.529759e-01 3.020380e-02 4.651955e-03 1.946687e-03

lambda$vectors

##                [,1]          [,2]         [,3]          [,4]
[,5]
##  [1,]  0.770052707 -0.3595628867  0.512208576 -1.212294e-01 -
0.023208046
##  [2,]  0.020607444  0.0064784700  0.013447171 -7.835184e-03
0.989042929
##  [3,]  0.459104487 -0.2605992620 -0.824602008  2.030053e-01
0.006231571
##  [4,]  0.442395277  0.8959627956 -0.034881281  8.630735e-03 -
0.015103512
##  [5,] -0.004349946 -0.0015154350 -0.006280700  6.265597e-03 -
0.070456679
##  [6,] -0.003461124  0.0017240484  0.236858719  9.714347e-01
0.005749934
##  [7,] -0.002482186 -0.0006039179 -0.004911710 -6.851298e-05 -
0.064182335
##  [8,] -0.006732316 -0.0020103775 -0.009275875  1.488313e-02 -
0.073080876
##  [9,] -0.004611180 -0.0012562489 -0.004970505  3.214838e-03 -
0.080815413
##                [,6]          [,7]          [,8]          [,9]
##  [1,]  0.011728199  0.0017253286  0.0001564783  6.253514e-05
##  [2,]  0.140012239 -0.0375235348 -0.0004638668 -8.577476e-03
##  [3,] -0.009417996 -0.0016136975 -0.0011527025 -8.392504e-05
##  [4,]  0.004138600  0.0006108642 -0.0003813870  3.630229e-04
##  [5,]  0.472098416 -0.2823081524  0.3073272995  7.732471e-01
##  [6,] -0.010565321 -0.0070248213  0.0006326070 -1.805388e-03
##  [7,]  0.295149427 -0.4661481465  0.5865444157 -5.894136e-01
##  [8,]  0.693632901  0.6928389806  0.0108480697 -1.817475e-01
##  [9,]  0.434654887 -0.4706578821 -0.7492634949 -1.468724e-01

varianza=sum(diag(S))
v=sum(lambda$values)
print("La varianza total es:")

## [1] "La varianza total es:"

print(varianza)

## [1] 3106.289

print("La suma acumulada es:")
```

```
## [1] "La suma acumulada es:"
```

```r
cumsum(lambda$values/varianza)
```

```
## [1] 0.7264164 0.9300767 0.9775266 0.9996963 0.9999067 0.9999882
0.9999979
## [8] 0.9999994 1.0000000
```

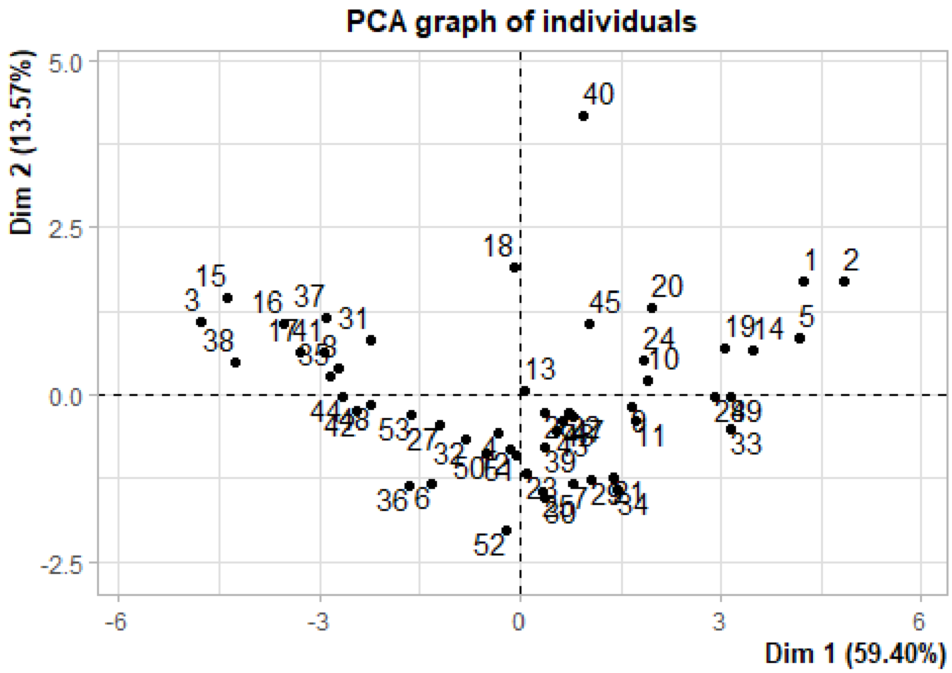
*El componente uno explica el 72.64% de la varianza. De acuerdo con el porcentaje de varianza explicada, el número ideal de componentes es 3, ya que con estos se logra explicar el 97.753%.

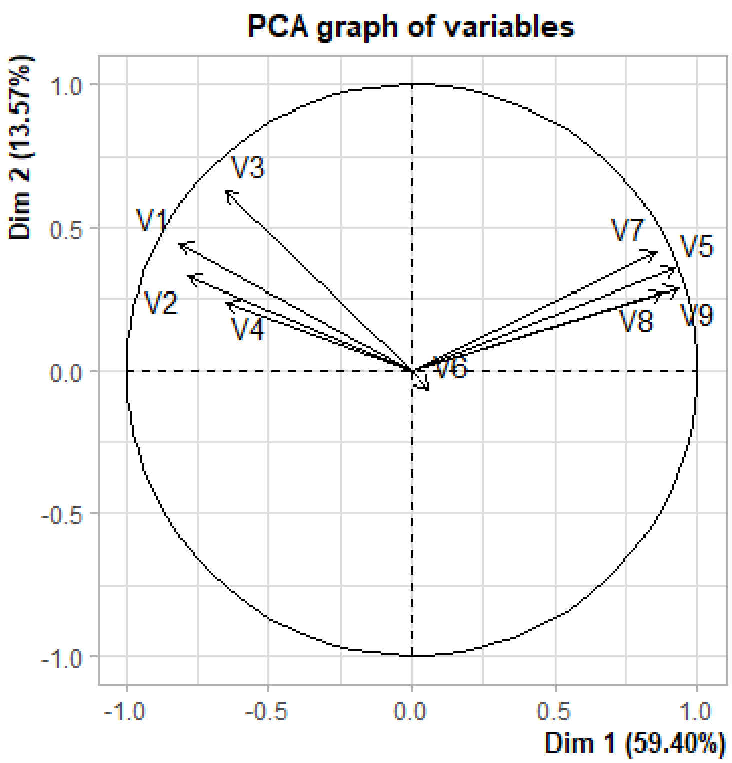
```r
library(stats)
library(FactoMineR)
library(factoextra)
```

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

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

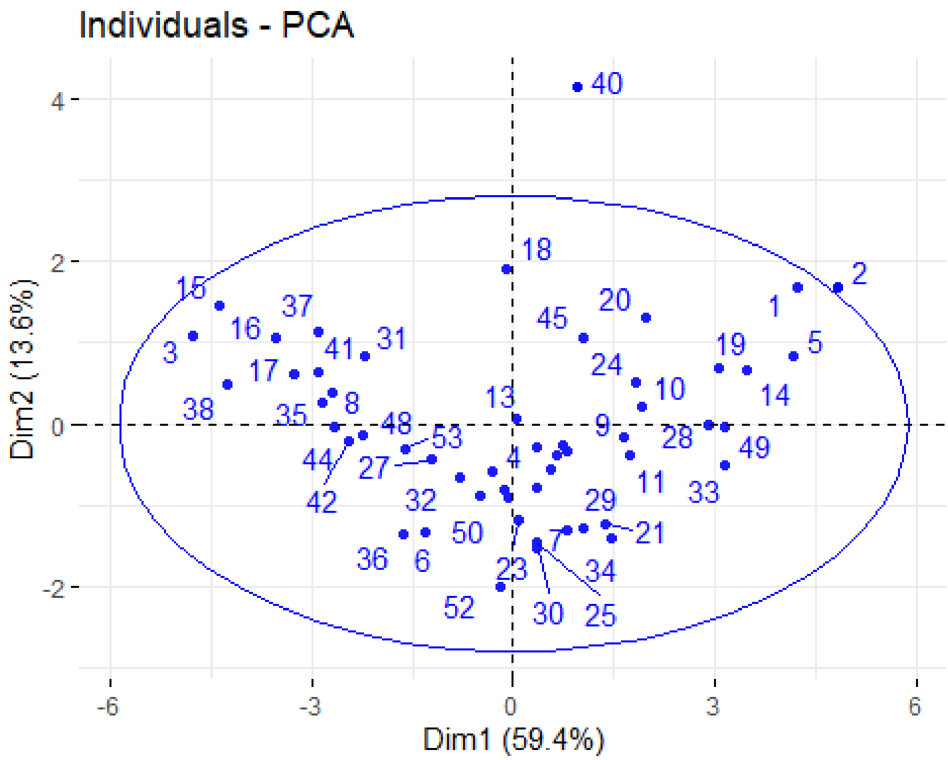
```r
library(ggplot2)
datos=D
cp3 = PCA(datos)
```



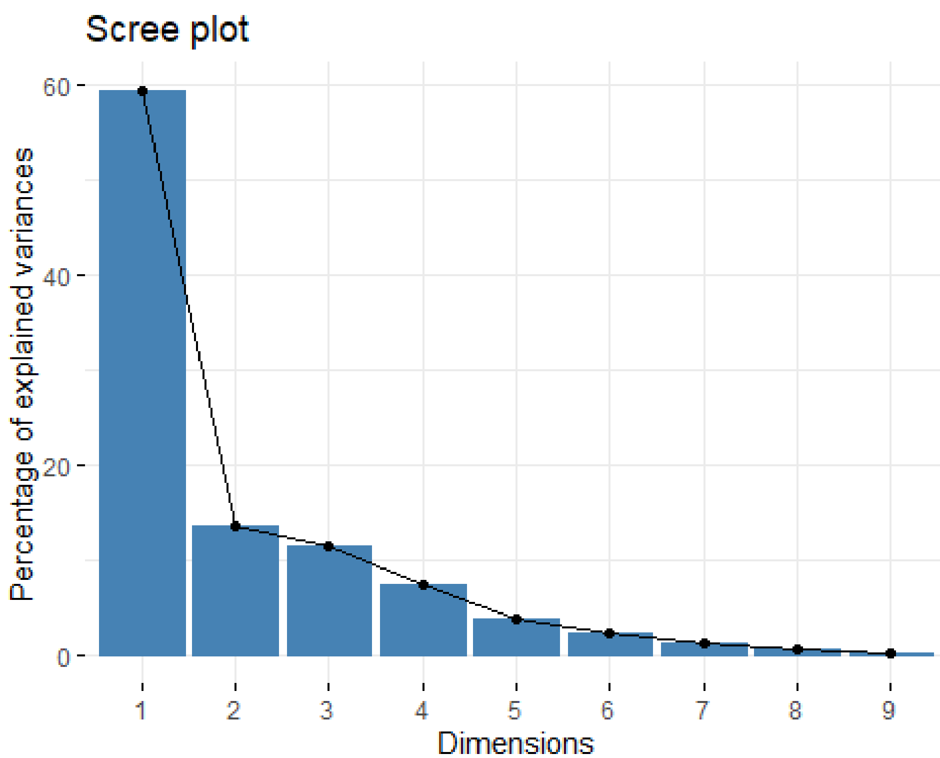



```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

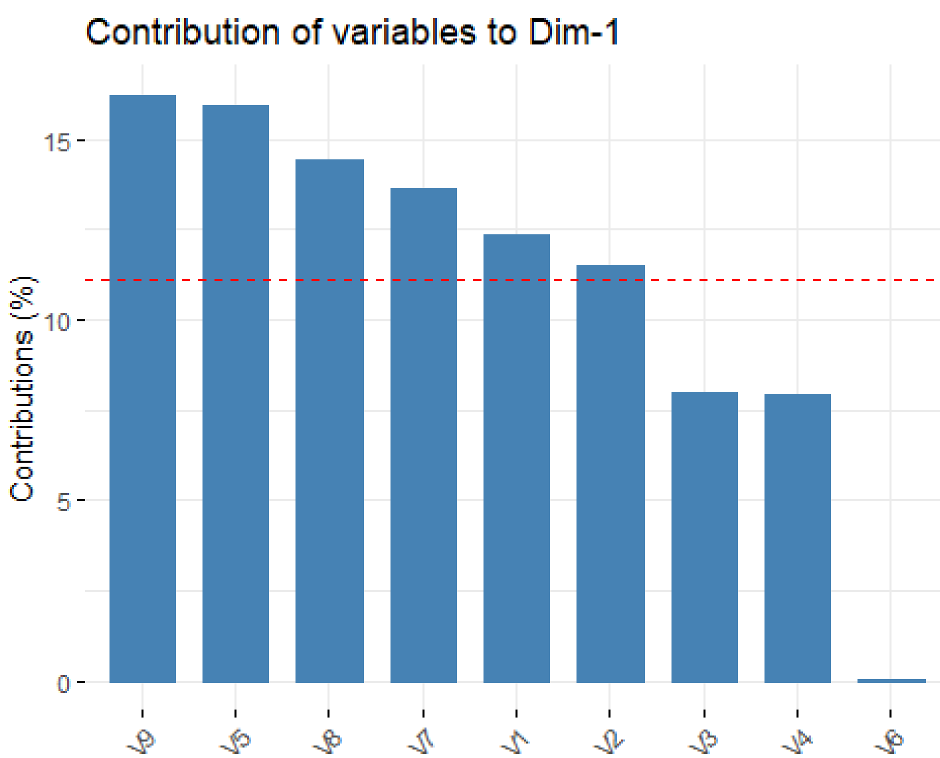
```
## Warning: ggrepel: 8 unlabeled data points (too many overlaps).
Consider
## increasing max.overlaps
```




```
fviz_screeplot(cp3)
```




```
fviz_contrib(cp3, choice = c("var"))
```



*En la primera gráfica se pueden ver algunos outliers como el 40 y un cluster de variables en el segundo, tercer y cuarto cuadrante.

*La segunda gráfica muestra las variables que tienen mayor influencia en los componentes. Se puede ver que el componente dos casi no tiene variables con correlación negativa, mientras que el componente uno la mitad de las variables tienen correlación positiva y la otra mitad negativa. De acuerdo con el gráfico, la variable 9 tiene mayor influencia en el componente uno y la variable 3 en el componente 2.

*El gráfico de sedimentación muestra que el primer componente explica la mayor parte de la varianza. Podemos usar este gráfico para justificar el número ideal de componentes el cual estaría entre 3 y 4, ya que después de estos componentes el porcentaje de varianza explicada es mínimo.

*El último gráfico nos muestra que las variables 9 y 5 son las que más contribuyen al componente uno.

#CONCLUSIONES

*Las variables que más influyen en el componente uno son:

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) X7 = concentración media de mercurio (parte por millón) en el tejido muscualar del grupo de peces estudiados en cada lago

*El análisis por componentes principales nos permite reducir la dimensionalidad del conjunto de datos, es decir, nos facilita el trabajo, ya que en este caso reducimos el conjunto de datos de 9 variables a 3.

*Debido a que se encontró normalidad en 2 variables facilita hacer cálculos o análisis donde se necesita normalidad para obtener una respuesta. La normalidad se encontró en las siguientes variables:

X4 = PH X10 = máximo de la concentración de mercurio en cada grupo de peces