A2-Matrices y vectores aleatorios

Héctor San Román Caraza

4/10/2022

1.- Considere la matriz de datos siguiente: X = que consta de 3 observaciones (filas) y 3 variables (columnas)

```
X = matrix(c(1,6,8,4,2,3,3,6,3), ncol =3)

b = matrix(c(1,1,1), ncol =3)

X1 = X[,1]
X2 = X[,2]
X3 = X[,3]


a1 = X1 + X2 + X3

a2 = X1 + (2*X2) - (3* X3)
```

a) Hallar la media, varianza y covarianza de b'X y c'X

```
# Media
cat("Media b'X: ",mean(a1),'\n')
```

```
## Media b'X:  12
```

```
# Media
cat("Media c'X: ",mean(a2),'\n')
```

```
## Media c'X:  -1
```

```
# varianza
cat("Varianza b'X: ",var(a1),'\n')
```

```
## Varianza b'X:  12
```

```
cat("Varianza c'X: ",var(a2),'\n')
```

```
## Varianza c'X:  43
```

```
#Covarianza
cat("Covarianza b'X:",cov(matrix(c(8,14,14), ncol =1)),"\n")
```

```
## Covarianza b'X: 12
```

```
cat("Covarianza c'X:",cov(matrix(c(0,-8,5), ncol =1)))
```

```
## Covarianza c'X: 43
```

b) Hallar el determinante de S (matriz de var-covarianzas de X)

```
det(cov(X))
```

```
## [1] 0
```

c) Hallar la matriz de varianzas-covarianzas (o porqué no se puede hallar)

```
cat("Covarianzas")
```

```
## Covarianzas
```

```
S = cov(X)
cat(S)
```

```
## 13 -2.5 1.5 -2.5 1 -1.5 1.5 -1.5 3
```

```
cat("Varianzas")
```

```
## Varianzas
```

```
var(X)
```

```
##      [,1] [,2] [,3]
## [1,] 13.0 -2.5  1.5
## [2,] -2.5  1.0 -1.5
## [3,]  1.5 -1.5  3.0
```

d) Hallar los valores y vectores propios de S

```
lambda <- eigen(cov(S))
cat("Vectores propios")
```

```
## Vectores propios
```

```
lambda$vectors
```

```
##            [,1]       [,2]      [,3]
## [1,]  0.9718588  0.1963193 0.1301889
## [2,] -0.1952911  0.3624264 0.9113224
## [3,]  0.1317262 -0.9111014 0.3905667
```

```
cat("Valores propios","\n")
```

```
## Valores propios
```

```
lambda$values
```

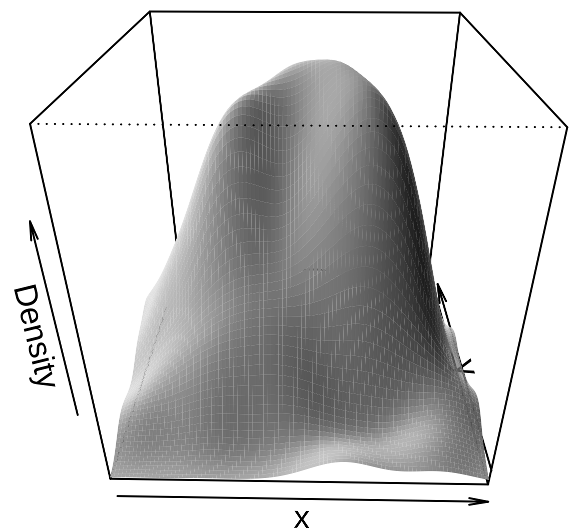
```
## [1]  6.835432e+01  4.895676e+00 -1.215444e-15
```

2.- Explore los resultados del siguiente código y dé una interpretación (se sugiere intersertarlo en un trozo de R en Rmarkdown para que dé varias ventanas de salida de resutados):

```
library(MVN)
```

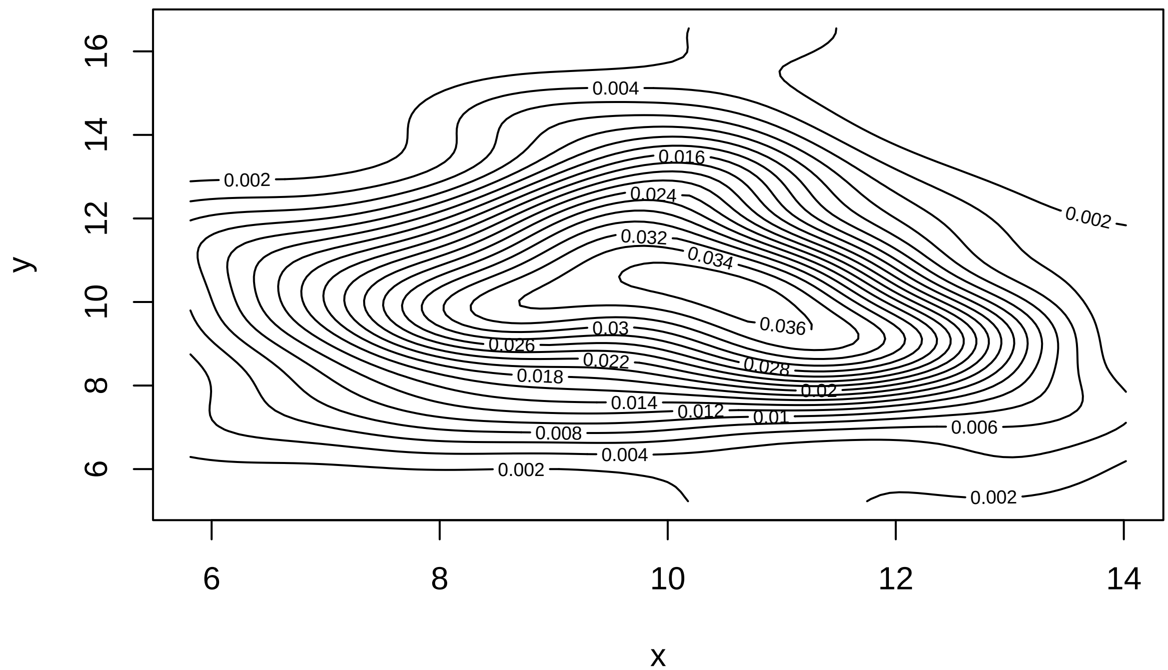
```
## Warning: package 'MVN' was built under R version 4.1.3
```

```
x = rnorm(100, 10, 2)
y = rnorm(100, 10, 2)
datos = data.frame(x,y)
mvn(datos, mvnTest = "hz", multivariatePlot = "persp")
```




```
## $multivariateNormality
##           Test        HZ   p value MVN
## 1 Henze-Zirkler 0.6205304 0.3540006 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.2066    0.8655    YES
## 2 Anderson-Darling     y        0.2595    0.7064    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
## x 100  9.81899 1.807965 9.863153 5.814083 14.01850 8.596829 10.99964
## y 100 10.13663 1.956214 9.962392 5.230179 16.55053 8.812105 11.20066
##          Skew   Kurtosis
## x -0.09778648 -0.4842697
## y  0.36282715  0.4209356
```

```
mvn(datos, mvnTest = "hz", multivariatePlot = "contour")
```




```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.6205304 0.3540006 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.2066    0.8655    YES
## 2 Anderson-Darling     y        0.2595    0.7064    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
## x 100  9.81899 1.807965 9.863153 5.814083 14.01850 8.596829 10.99964
## y 100 10.13663 1.956214 9.962392 5.230179 16.55053 8.812105 11.20066
##          Skew   Kurtosis
## x -0.09778648 -0.4842697
## y  0.36282715  0.4209356
```

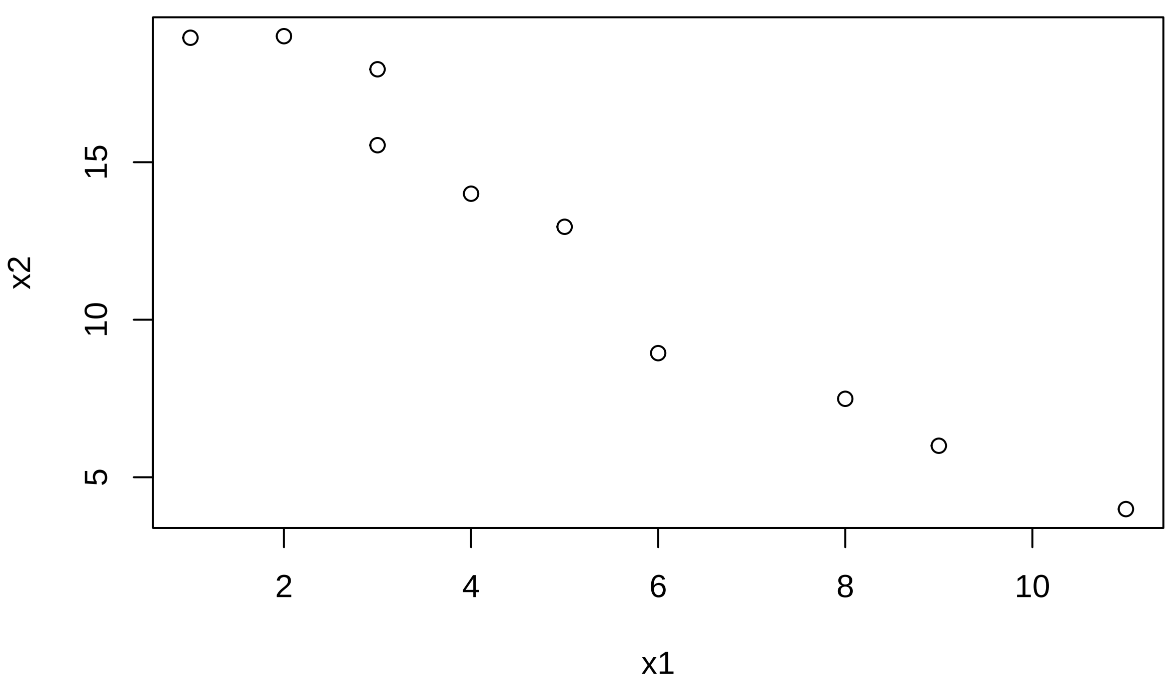
Aquí estamos haciendo un análisis de normal multivariable en nuestros datos. Estamos haciendo la prueba de Henze-Zirkler y de acuerdo a nuestro P value, sí tenemos una distribución normal multivariada. Vemos que el test Anderson-Darling también nos arroja que se trata de una normal multivariada, si observamos el p-value. Vemos que el código nos arrjoa 2 distintas gráficas, una de densidad y otra de contorno. La gráfica de densidad nos ayuda a ver la distribución de nuestros datos, mientras que la de contorno nos ayuda a ver las probabilidades dependiendo del nivel de los datos.

3.- Un periódico matutino enumera los siguientes precios de autos usados para un compacto extranjero con edad medida en años y precio en venta medido en miles de dólares.

- x1: 1, 2, 3, 3, 4, 5, 6, 8, 9, 11
- x2: 18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99

a) Construya un diagrama de dispersión

```
x1 <- c(1, 2, 3, 3, 4, 5, 6, 8, 9, 11)
x2 <- c(18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99)
plot(x1,x2)
```



b) Inferir el signo de la covarianza muestral a partir del gráfico.

El signo que podemos inferir a partir de la gráfica será negativo, ya que muestra una tendencia lineal hacia abajo, al cuadrante 4 de nuestra gráfica.

c) Calcular el cuadrado de las distancias estadísticas con . Nota: para el cálculo de la distancia de Mahalanobis, usa: mahalanobis(A,medias,S).

```
X3 = matrix(c(x1,x2), ncol =2)
S = cov(X3)
```

```r
mahalanobis(X3, colMeans(X3), S)
```

```
##  [1] 1.8753045 2.0203262 2.9009088 0.7352659 0.3105192 0.0176162 3.7329012
##  [8] 0.8165401 1.3753379 4.2152799
```