A2-Matrices y vectores aleatorios

Edgar Antonio Galarza López A00828688

2022-10-04

Ejercicio 1.

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

```
##      [,1] [,2] [,3]
## [1,]    1    4    3
## [2,]    6    2    6
## [3,]    8    3    3
```

```
bX = X %*% c(1, 1, 1)
cX = X %*% c(1, 2, -3)
A = cbind(bX, cX)
A
```

```
##      [,1] [,2]
## [1,]    8    0
## [2,]   14   -8
## [3,]   14    5
```

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

```
bX_mu = mean(bX)
cX_mu = mean(cX)

bX_var = var(bX[,1])
cX_var = var(cX)
cov = cov(bX[,1], cX[,1])
cat("Media de b'X:", bX_mu, "\n")
```

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

```
cat("Media de c'X:", cX_mu, "\n", "\n")
```

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

```
cat("Varianza de b'X:", bX_var, "\n")
```

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

```
cat("Varianza de c'X:", cX_var, "\n", "\n")
```

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

cat("Covarianza:", cov, "\n")

## Covarianza: -3
```

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

```
covA = cov(A)
bX_det = det(covA)
cat("Determinante de cov_A(S):", bX_det, "\n")

## Determinante de cov_A(S): 507
```

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

```
cat("Matriz de Varianzas-Covarianzas:", covA, "\n")

## Matriz de Varianzas-Covarianzas: 12 -3 -3 43
```

d) Hallar los valores y vectores propios de S.

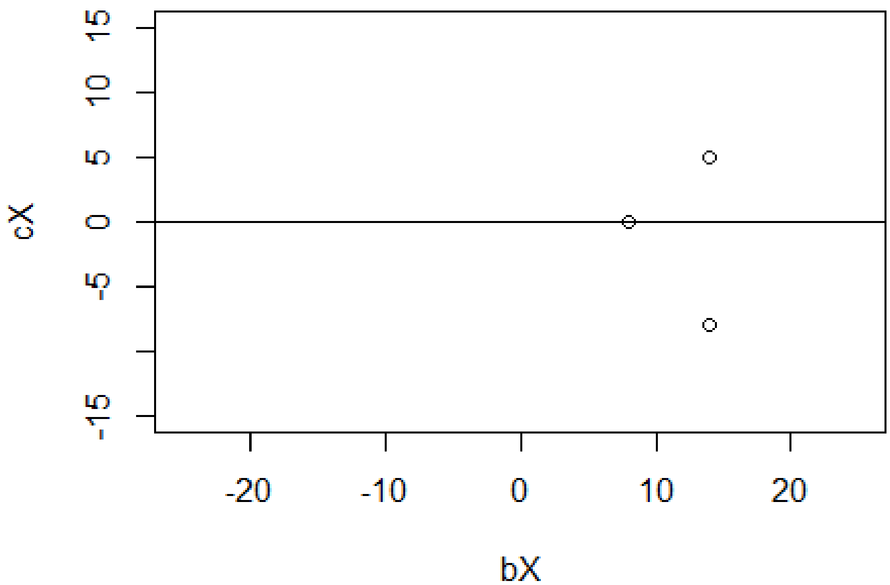
```
e = eigen(covA)
e

## eigen() decomposition
## $values
## [1] 43.28765 11.71235
##
## $vectors
##             [,1]        [,2]
## [1,] -0.09544671 -0.99543454
## [2,]  0.99543454 -0.09544671
```

e) Argumentar si b'X y c'X son independientes o no.

Son independientes puesto que ambos pesos en los coeficientes contribuyen para poder obtener Y.

f) Hallar la varianza generalizada. Explicar el comportamiento de los datos de X basándose en los la variable generalizada, en los valores y vectores propios.

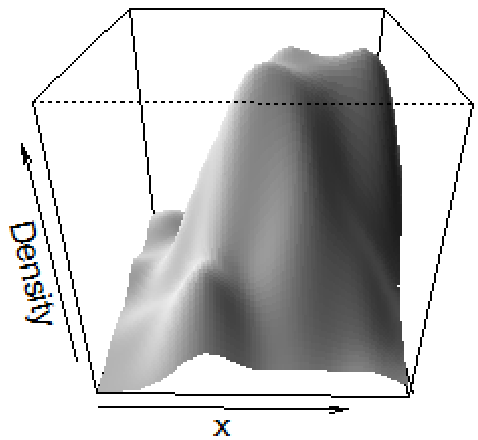
```
plot(bX, cX, xlim = c(-25,25), ylim= c(-15,15))
x11 = seq(0, 100, 100)
x12 = e$vectors[1,1] / e$vectors[2,1] * x11
x21 = seq(0,100,100)
x22 = e$vectors[1,1] / e$vectors[1,2] * x21
abline(x11,x12)
abline(x21,x22)
```



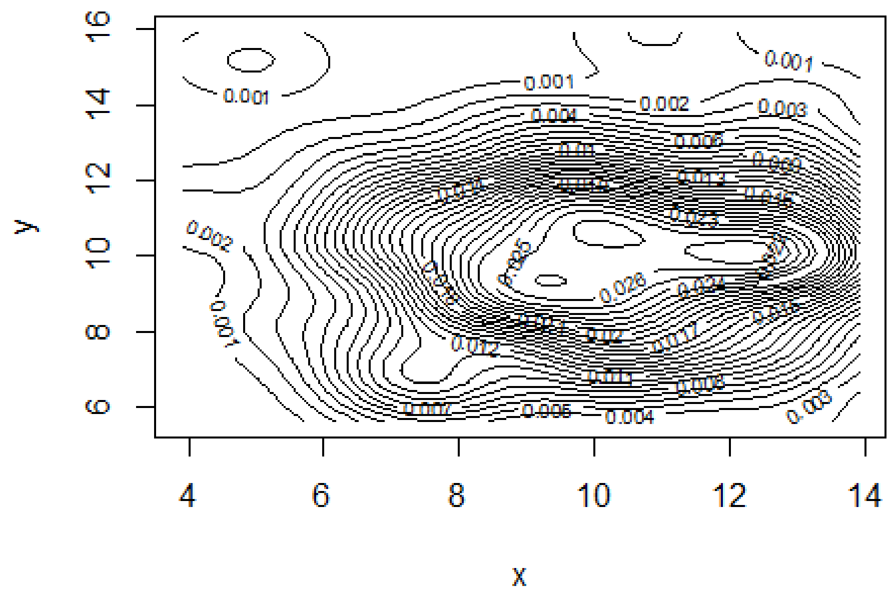
Ejercicio 2.

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

```
library(MVN)
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 0.4231184 0.7661876 YES
##
## $univariateNormality
##                Test  Variable Statistic   p value Normality
## 1 Anderson-Darling      x        0.5628    0.1418    YES
## 2 Anderson-Darling      y        0.2205    0.8294    YES
##
## $Descriptives
##     n      Mean  Std.Dev    Median      Min      Max     25th     75th
## x 100 10.074114 2.230546 10.169805 3.908418 13.90780 8.510032 12.11936
## y 100  9.885466 1.965596  9.952288 5.633022 15.90477 8.644656 11.15001
##          Skew   Kurtosis
## x -0.2897178 -0.5925475
## y  0.2409851  0.1983794

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




```
## $multivariateNormality
##              Test        HZ   p value MVN
## 1 Henze-Zirkler 0.4231184 0.7661876 YES
##
## $univariateNormality
##                 Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.5628    0.1418    YES
## 2 Anderson-Darling     y        0.2205    0.8294    YES
##
## $Descriptives
##     n      Mean  Std.Dev    Median      Min      Max     25th     75th
## x 100 10.074114 2.230546 10.169805 3.908418 13.90780 8.510032 12.11936
## y 100  9.885466 1.965596  9.952288 5.633022 15.90477 8.644656 11.15001
##         Skew   Kurtosis
## x -0.2897178 -0.5925475
## y  0.2409851  0.1983794
```

Analizando las gráficas obtenidas tenemos que el diagrama de perspéctiva ofrece una extensión de la curva de distribución de probabilidad univariada en una superficie de distribución tridimensional. Aquí se tiene información de como es que se relacionan las variables del set de datos entre sí. Dos dimensiones se refieren a los valores de las dos variables y la tercera dimensión es el valor de la función de densidad de probabilidad multivariable.

Por otro lado la gráfica de contorno muestra un espacio bidimensional para verificar la suposición de normalidad multivariado. Aquí podemos observar información sobre la normalidad y correlación de los datos.

Dicho esto podemos decir que existe una correlación positiva entre nuestra variable x y y en cierto nivel. A medida de que la correlación se acerca a 0, debería hacer que las líneas en la gráfica de contorno se vayan volviendo circulares y en este caso las vemos más elípticas.

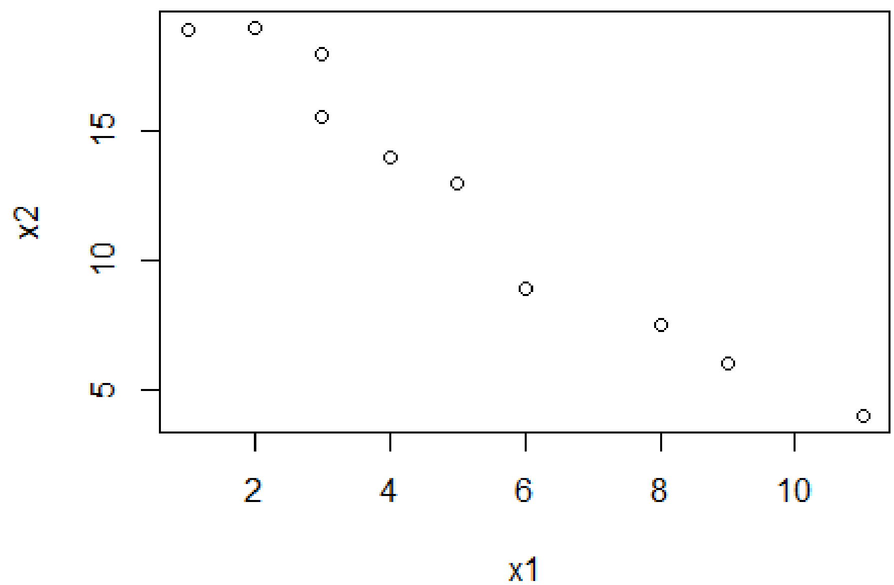
Ejercicio 3.

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

```
n <- 10 #Cantidad de datos
x1 <- c(1,2,3,3,4,5,6,8,9,11) #Años
x2 <- c(18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99)
#Precio
X <- cbind(x1,x2) #DF
X_mu <- colMeans(X) #Media
S <- cov(X) # Covarianza
Sinv <- solve(S)
```

a) Construya un diagrama de dispersión

```
plot(x = x1, y = x2) # Diagrama de dispersión
```



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

Podemos apreciar que la covarianza es negativa debido a que cuando x1 crece, el valor de x2 disminuye.

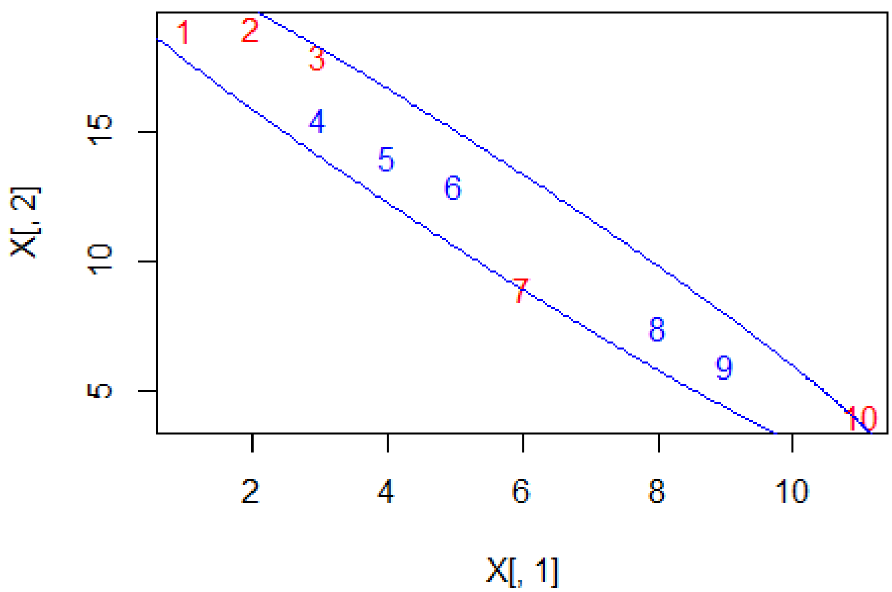
c) Calcular el cuadrado de las distancias estadísticas. Nota: para el cálculo de la distancia de Mahalanobis, usa: mahalanobis(A,medias,S).

```
d <- mahalanobis(X, colMeans(X), cov(X)) #Distancia de Mahalanobis
d

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

d) Usando las anteriores distancias, determine la proporción de las observaciones que caen dentro del contorno de probabilidad estimado del 50% de una distribución normal bivariada.

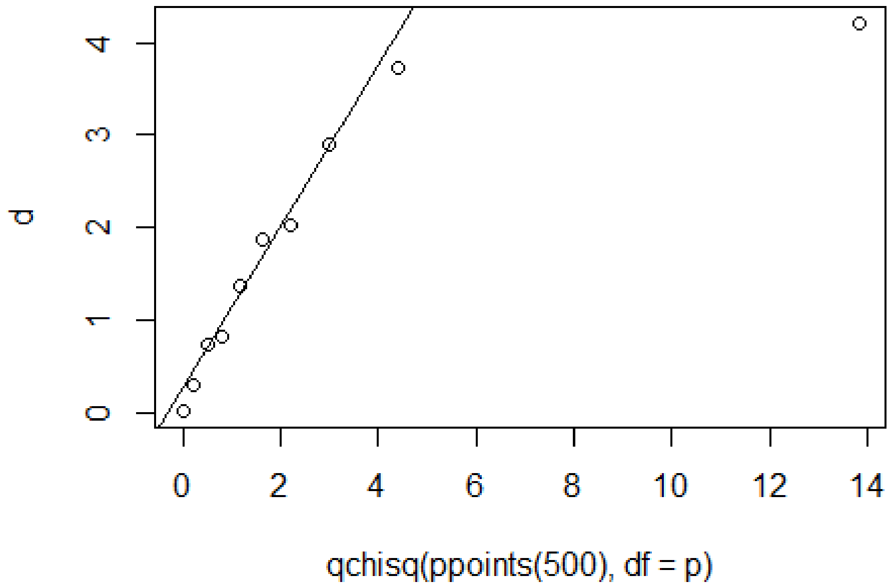
```
library(ellipse)

##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs
```

```
p <- 2
elps <- t(t(ellipse(S, level=0.85, npoints=1000))+X_mu)
plot(X[,1],X[,2],type="n")
index <- d < qchisq(0.5,df=p)
text(X[,1][index],X[,2][index],(1:n)[index],col="blue")
text(X[,1][!index],X[,2][!index],(1:n)[!index],col="red")
lines(elps,col="blue")
```



e) Ordene las distancias del inciso c y construya un diagrama chi-cuadrado.

```
names(d) <- 1:10
sort(d)

##         6         5         4         8         9         1         2
3
## 0.0176162 0.3105192 0.7352659 0.8165401 1.3753379 1.8753045 2.0203262
2.9009088
##         7        10
## 3.7329012 4.2152799

qqplot(qchisq(ppoints(500),df=p),d)
qqline(d,distribution=function(x){qchisq(x,df=p)})
```



f) Dados los resultados anteriores, ¿serían argumentos para decir que los datos son aproximadamente normales bivariados?

Según los resultados obtenidos para la parte b y c, concluimos que estos datos son aproximadamente normales bivariados. Esto debido a que la mayoría de los datos están alrededor de la línea teórica.