Matrices y vectores aleatorios

Felipe Yépez

2022-10-04

```
X = matrix(c(1,6,8,4,2,3,3,6,3), nrow=3)
bX = matrix(c(1,1,1), nrow=1)
bX = t(bX%*%t(X))
bX
```

```
##      [,1]
## [1,]    8
## [2,]   14
## [3,]   14
```

```
cX = matrix(c(1,2,-3), nrow=1)
cX = t(cX%*%t(X))
cX
```

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

```
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

```
mean(bX)
```

```
## [1] 12
```

```
mean(cX)
```

```
## [1] -1
```

```r
var(bX)
```

```
##      [,1]
## [1,]   12
```

```r
var(cX)
```

```
##      [,1]
## [1,]   43
```

```r
S = cov(A)
S
```

```
##      [,1] [,2]
## [1,]   12   -3
## [2,]   -3   43
```

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

```r
det(S)
```

```
## [1] 507
```

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

```r
S
```

```
##      [,1] [,2]
## [1,]   12   -3
## [2,]   -3   43
```

c) Hallar los valores y vectores propios de S

```r
e = eigen(S)
e
```

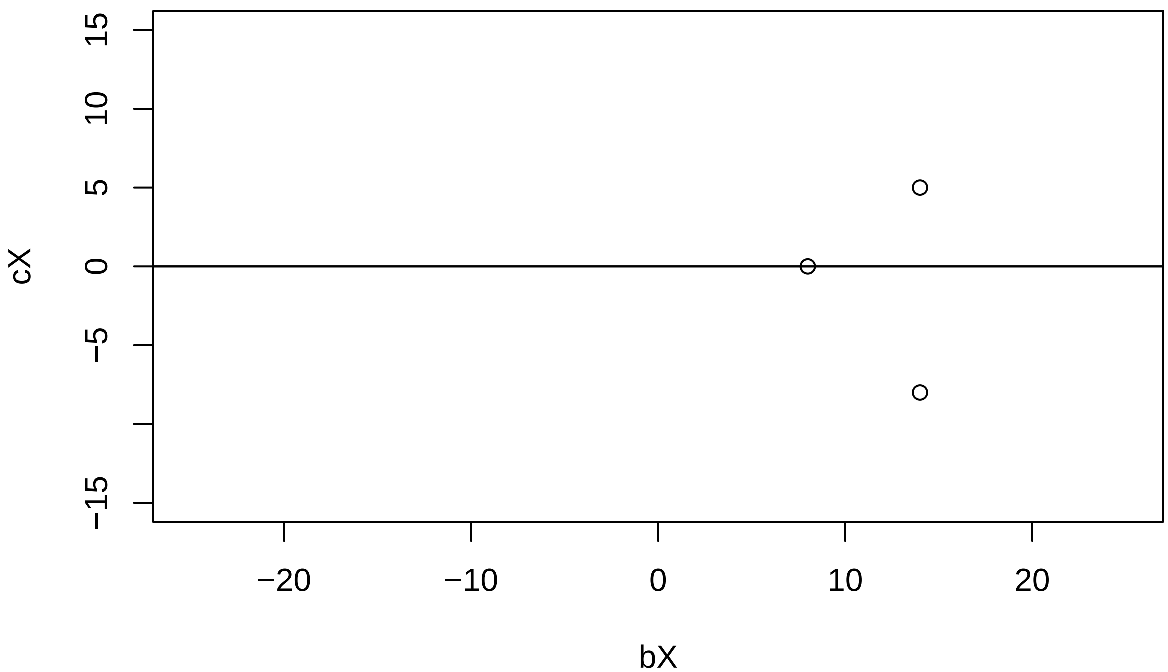
```
## eigen() decomposition
## $values
## [1] 43.28765 11.71235
##
## $vectors
##             [,1]        [,2]
## [1,] -0.09544671 -0.99543454
## [2,]  0.99543454 -0.09544671
```

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

Se puede afirmar que son independientes dado que los coeficientes obtenidos para cada variable son de tamaño considerable, es decir ambas variables contribuyen para obtener Y.

e) 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)
```

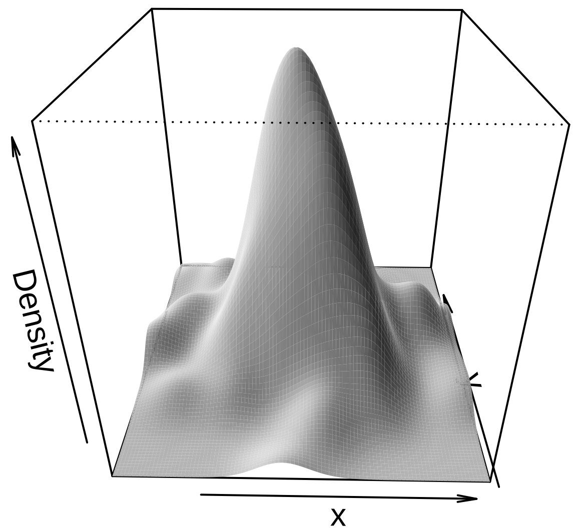


2. Explore los resultados del siguiente código y dé una interpretación (se sugiere intersertarlo en un trozo de R en Rmarkdown para que dé varias ventanas de salida de resutados):

```
library(MVN)
```

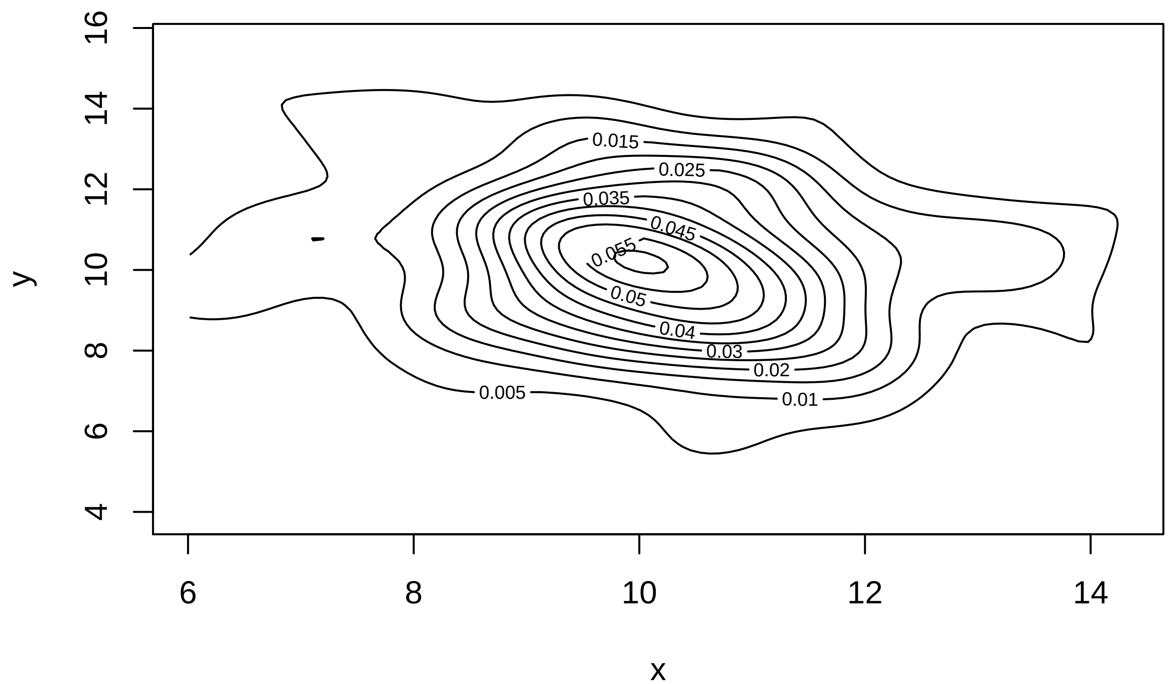
```
## Warning: package 'MVN' was built under R version 4.0.5
```

```
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 0.6144247 0.3646288 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x         0.5050    0.1984    YES
## 2 Anderson-Darling     y         0.2364    0.7823    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
## x 100 10.15634 1.654599 10.12678 6.021296 14.31289 9.289565 11.10671
## y 100 10.12814 1.934576 10.20096 3.914064 15.63254 8.789300 11.14190
##           Skew  Kurtosis
## x  0.006748517 0.2892304
## y -0.015967666 0.6132246
```

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




```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.6144247 0.3646288 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.5050    0.1984    YES
## 2 Anderson-Darling     y        0.2364    0.7823    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
## x 100 10.15634 1.654599 10.12678 6.021296 14.31289 9.289565 11.10671
## y 100 10.12814 1.934576 10.20096 3.914064 15.63254 8.789300 11.14190
##           Skew  Kurtosis
## x  0.006748517 0.2892304
## y -0.015967666 0.6132246
```

El diagrama de perspectiva muestra la curva de distribución de probabilidad. El eje z, es decir el de densidad muestra la densidad de probabilidad multivariable.

Como se puede observar en el gráfico de contornos el máximo no es alcanzado en el origen (0, 0) por lo que se puede interpretar que tienen correlacion ambas variables.

En el diagrama de contorno se puede analizar la normalidad y correlación de las variables. Se puede observar correlación entre las variables dado que los contornos que se generan no son completamente circulares, algo que sugeriría que existe correlación de 0, caso que no se da.

La correlación sugiere ser un tanto negativa dado que se forman elipses cercanas a la diagonal y = -x.

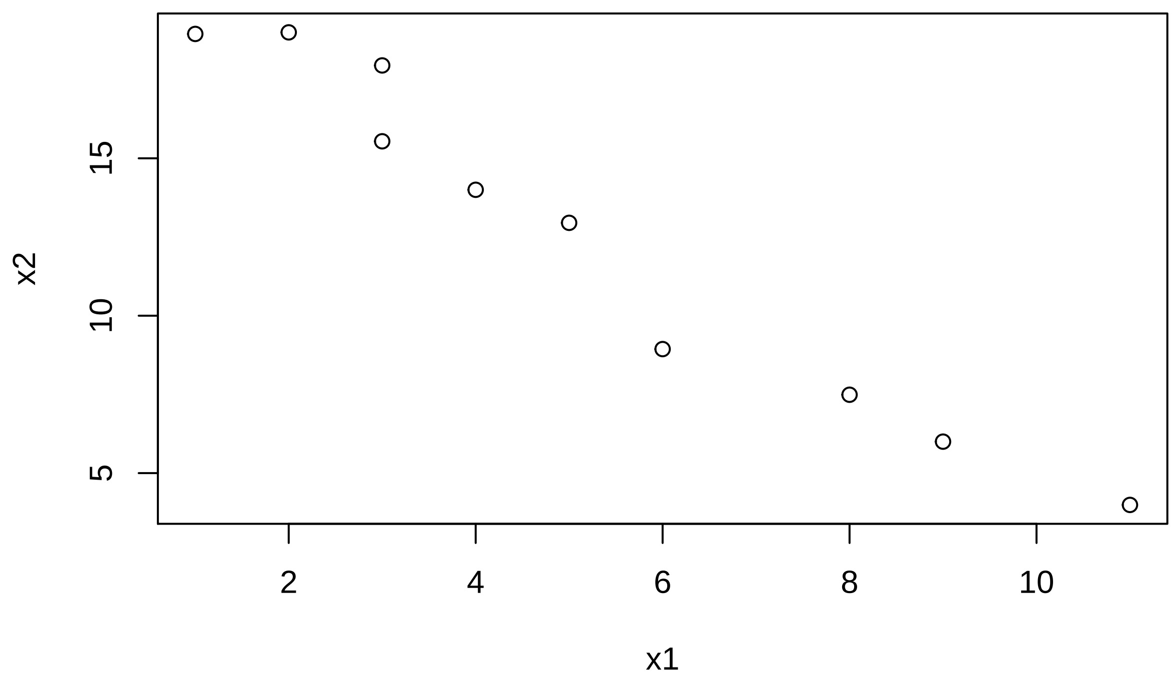
Info: https://datasciencegenie.com/3d-contour-plots-of-bivariate-normal-distribution/

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 = 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)
data = data.frame(x1 = x1, x2=x2)
```

a) Construya un diagrama de dispersión

```
plot(data)
```



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

La covarianza muestral según el gráfico sugiere ser negativa ya que según aumenta x1, x2 decrece.

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

```r
S = cov(data)
dist = mahalanobis(data, colMeans(data), S)
dist
```

```
##  [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.

```r
pvalue = pchisq(dist, df = 1)
pvalue
```

```
##  [1] 0.8291312 0.8447942 0.9114704 0.6088184 0.4226382 0.1055900 0.9466493
##  [8] 0.6338063 0.7591031 0.9599385
```

```r
dentro = pvalue > 0.5
dentro
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
```

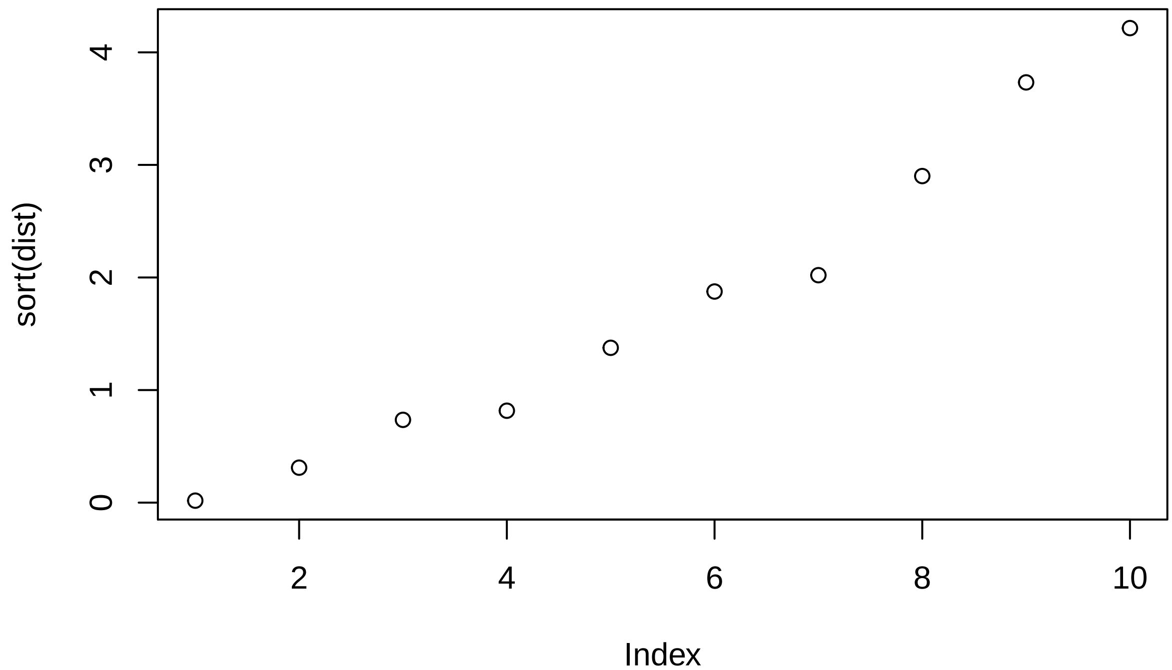
```r
table(dentro)
```

```
## dentro
## FALSE  TRUE
##     2     8
```

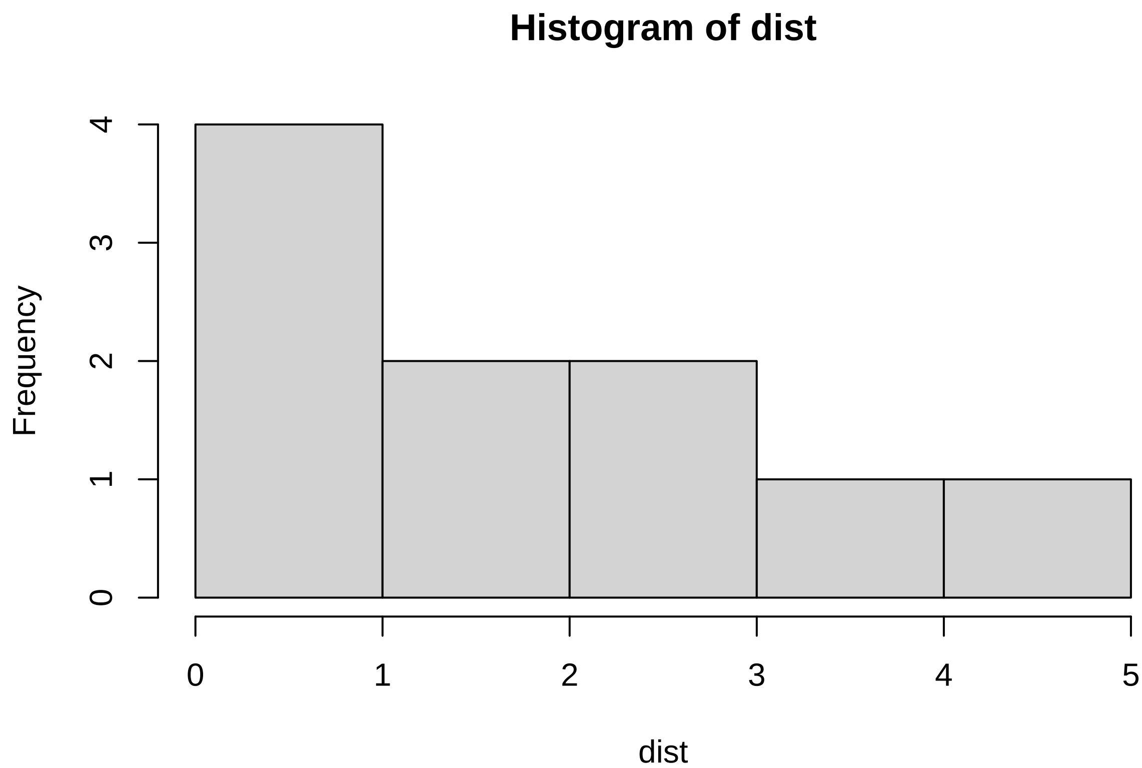
80% dentro, 20% fuera.

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

```r
plot(sort(dist))
```




```r
hist(dist)
```



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

Si, se puede decir que los daos son aproximadamente normales bivariados.