# Actividad1_5

2024-09-30

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

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

## a) Hallar la media, varianza y covarianza de X

Media

```
medias <- colMeans(X)
medias
```

```
## [1] 5 3 4
```

Varianza

```
varianzas <- apply(X, 2, var)
varianzas
```

```
## [1] 13  1  3
```

Covarianza

```
covarianza <- cov(X)
covarianza
```

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

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

Se crea b y c

```
b <- matrix(c(1, 1, 1), nrow = 1, ncol = 3, byrow = TRUE)
c <- matrix(c(1, 2, -3), nrow = 1, ncol = 3, byrow = TRUE)
```

Se crea b'x y c'x, y se guarda la Y transpuesta

```
bx <- b %*% t(X)
cx <- c %*% t(X)
Y <- t(rbind(bx,cx))
Y
```

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

```
## [3,]   14    5
```

Obtener la media

```
medias_cx <- colMeans(Y)
medias_cx
```

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

Obtener la varianza

```
varY <- apply(Y, 2, var)
varY
```

```
## [1] 12 43
```

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

```
S <- det(cov(X))
S
```

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

### d) Hallar los valores y vectores propios de S

```
eigen(cov(X))
```

```
## eigen() decomposition
## $values
## [1]  1.379150e+01  3.208497e+00 -7.859007e-17
##
## $vectors
##            [,1]       [,2]       [,3]
## [1,]  0.9645458 -0.2295697 -0.1301889
## [2,] -0.2076189 -0.3555080 -0.9113224
## [3,]  0.1629288  0.9060418 -0.3905667
```

### e) Argumentar si hay independencia entre b'X y c'X , ¿y qué ocurre con X1, X2 y X3? ¿son independientes?

Dado que el determinante es 0, sabemos que existe dependencia lineal entre las variables $Y_1$ y $Y_2$.

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

```
print("La varianza generalizada")
```

```
## [1] "La varianza generalizada"
```

```
print(S)
```

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

```
print("Eigen valores y vectores")
```

```
## [1] "Eigen valores y vectores"
```

```r
print(eigen(cov(X)))
```

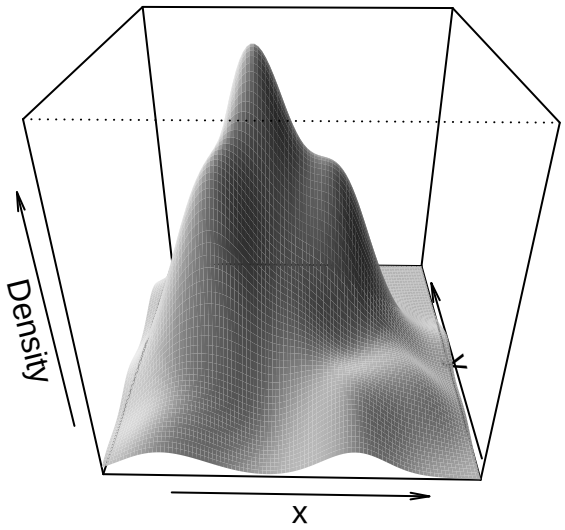
```
## eigen() decomposition
## $values
## [1]  1.379150e+01  3.208497e+00 -7.859007e-17
##
## $vectors
##            [,1]       [,2]       [,3]
## [1,]  0.9645458 -0.2295697 -0.1301889
## [2,] -0.2076189 -0.3555080 -0.9113224
## [3,]  0.1629288  0.9060418 -0.3905667
```

La varianza generalizada nos da una idea de la dispersión conjunta de todas las variables, como el determinante es 0, esto indica que las variables están correlacionadas linealmente y que existe dependencia lineal entre ellas.

Los valores propios de $S$ explican la varianza en las direcciones principales. Dados que los vectores son cercanos a cero, indican poca variación lo que refuerza la idea de dependencia lineal entre las variables.

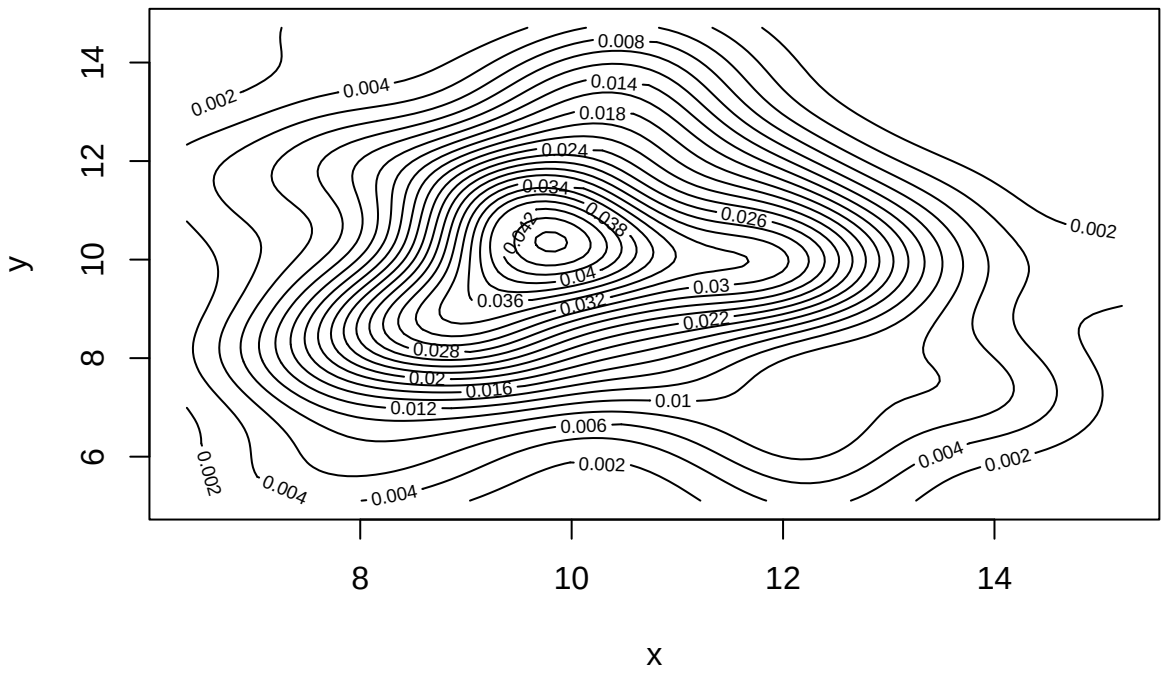
## 2. Explore los resultados del siguiente código y dé una interpretación

```r
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 0.7052116 0.2289475 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.3292    0.5118    YES
## 2 Anderson-Darling     y        0.1248    0.9858    YES
##
## $Descriptives
##     n      Mean  Std.Dev    Median      Min      Max    25th     75th
## x 100 10.177044 1.764563 10.001412 6.359761 15.20591 9.02911 11.53629
## y 100  9.886751 2.019447  9.920327 5.107353 14.70608 8.64010 11.13416
##         Skew   Kurtosis
## x 0.27556747 -0.2938956
## y 0.01596805 -0.2628866
```

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





```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.7052116 0.2289475 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.3292    0.5118    YES
## 2 Anderson-Darling     y        0.1248    0.9858    YES
```

```
##
## $Descriptives
##     n      Mean  Std.Dev    Median      Min      Max    25th     75th
## x 100 10.177044 1.764563 10.001412 6.359761 15.20591 9.02911 11.53629
## y 100  9.886751 2.019447  9.920327 5.107353 14.70608 8.64010 11.13416
##         Skew   Kurtosis
## x 0.27556747 -0.2938956
## y 0.01596805 -0.2628866
```

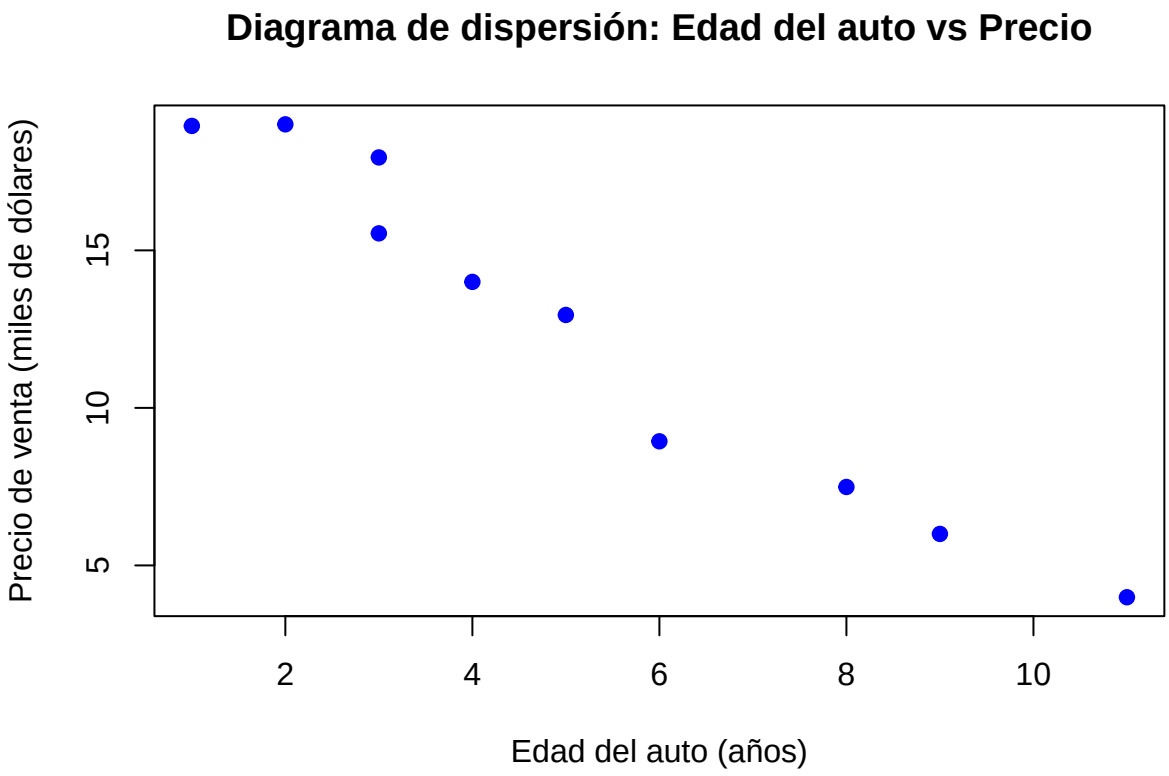
Segun el valor $p = 0.8130 > 0.05$ entonces, no hay suficiente evidencia para rechazar la hipotesis nula de Henze-Zirkler, por lo que los datos son consistentes para una distribución multivariada. De igual manera obtenido valores de $p > 0.05$ para el test de Anderson-Darling en cada variable,lo cual significa que no hay suficiente evidencia para demostrar que los datos no siguen una desviación estandar, por lo que las variables siguen una distribución estandar. Los gráficos nos ayudan a visualizar como se distribuyen los datos en 3d, asi como en sus capas de nivel

## 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)
```

### a) Construya un diagrama de dispersión

```
plot(x1, x2,
     main = "Diagrama de dispersión: Edad del auto vs Precio",
     xlab = "Edad del auto (años)",
     ylab = "Precio de venta (miles de dólares)",
     pch = 19,
     col = "blue")
```




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

La covarianza será negativa, pues mientras la edad de los carros aumenta, el precio disminuye.

### c) Calcular el cuadrado de las distancias de Mahalanobis

```
# Calcular el vector de medias
media <- colMeans(datos)

# Calcular la matriz de covarianza
S <- cov(datos)

# Calcular la distancia de Mahalanobis
distancias_mahalanobis <- mahalanobis(datos, center = media, cov = S)

# Mostrar las distancias de Mahalanobis
distancias_mahalanobis
```

```
##   [1] 0.50365552 0.70386763 1.77138759 1.50043194 1.56199179 7.00181621
##   [7] 0.24422081 1.99563907 0.75652738 3.26742265 2.74321995 2.21498087
##  [13] 3.30510765 5.61053383 0.04996745 1.13462498 0.30129740 1.02115394
##  [19] 3.39509090 0.36520698 6.65926982 0.02226155 1.19439973 1.13089893
##  [25] 5.58811011 4.24933227 0.81533598 5.49288702 0.12576136 0.74989746
##  [31] 0.54666786 0.26260647 1.63508042 4.12397753 0.88187062 4.10579699
##  [37] 0.33472475 0.51396610 0.26945770 1.05468428 2.45992538 0.14852663
##  [43] 3.05733017 1.29295190 0.75241575 1.52151034 1.81162950 0.66629162
```

```
##  [49] 1.82125377 1.42340852 0.98925939 5.94255145 0.07621753 3.52960417
##  [55] 1.07902483 0.02306631 7.82895073 0.76251864 3.30302825 2.03869819
##  [61] 1.26478147 0.80969655 3.74192906 0.43859770 0.95624134 1.15555707
##  [67] 1.06542184 0.04065986 0.02771329 2.77124998 2.92212391 4.30112007
##  [73] 3.46840678 2.33466055 1.51443395 0.29197866 0.04945765 6.53833143
##  [79] 0.81784399 0.35697981 1.36660141 2.49679042 3.96456449 8.14629572
##  [85] 0.67114860 0.22529003 0.65359047 3.99044613 1.52488320 0.18781064
##  [91] 2.76843881 4.31106007 1.49938324 0.90652169 0.99874226 2.28139453
##  [97] 0.03061561 2.41682205 3.63503512 1.32408604
```

**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
umbral_50 <- qchisq(0.5, df = 2)


observaciones_dentro <- sum(distancias_mahalanobis <= umbral_50)

# Calcular la proporción de observaciones dentro del contorno del 50%
proporcion <- observaciones_dentro / length(distancias_mahalanobis)
proporcion
```

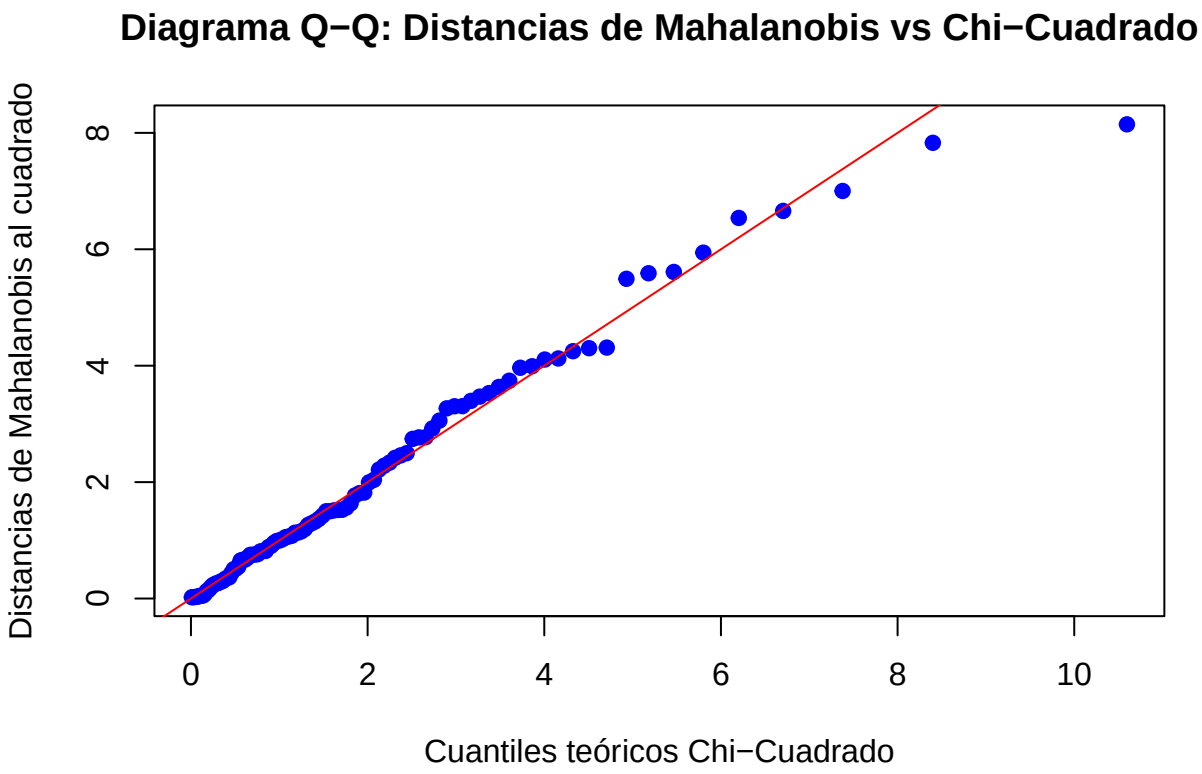
```
## [1] 0.52
```

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

```r
distancias_ordenadas <- sort(distancias_mahalanobis)

# Crear los cuantiles teóricos de la distribución chi-cuadrado
cuantiles_teoricos <- qchisq(ppoints(length(distancias_ordenadas)), df = 2)


qqplot(cuantiles_teoricos, distancias_ordenadas,
       main = "Diagrama Q-Q: Distancias de Mahalanobis vs Chi-Cuadrado",
       xlab = "Cuantiles teóricos Chi-Cuadrado",
       ylab = "Distancias de Mahalanobis al cuadrado",
       pch = 19, col = "blue")
abline(0, 1, col = "red")
```




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

Dado que las distancias de Mahalanobis siguen una distribución chi-cuadrado en el gráfico Q-Q.Asi como el 53% de los datos es parte del contorno del 50%. Podemos afirmar que los resultados son normales bivariados.

## 4. Ejercicio 4

Si X es un vector aleatorio con X1, X2, X3 son tres variables conjuntamente normales, no independientes, con b, un vector de 3 constantes, b1, b2 y b3, y c, otro vector de 3 constantes, c1, c2, c3, demuestra que las variables V1 = b'X y V2 = c'X son independientes si b´c = 0.

Dado que $X$ es un vector aleatorio normal multivariado, entonces cualquier combinación lineal de sus componentes también es una variable normal. Por lo que $V_1$ y $V_2$ son normales.

Dado que si no existe correlación entre las variables esto será independencia, entonces

Hay que conseguir la covarianza entre $V_1$ y $V_2$

$$cov(V_1, V_2) = Cov(b'X, c'X) = b'Cov(X)c$$

Dado que la matriz de covarianzas es simétrica, este producto debe dar 0. Una vez que vemos que las variables son conjuntamente normales, y la covarianza entre $V_1, V_2$ es cero, Esto demuestra que son independientes

Bien se sabe que la matriz de covarianzas es simétrica, por lo que el producto $b^T Cov(X)c$ debe dar 0. Recordando que las variables son conjuntamente normales, y observando que la covarianza entre $V_1, V_2$ es cero, hemos demostrado que son independientes.