Análisis Multivariado

Producto académico 01

Kevin Heberth Haquehua Apaza

14 de julio del 2025

Table of Contents

# Ejercicios Distribución Normal Multivariante e Inferencia

## Ejercicio 1:

Se tiene la matriz de datos X\_1:

X\_1 = data.frame(x1 = c(6,10,8),  
 x2 = c(9,6,3)) ; X\_1

## x1 x2  
## 1 6 9  
## 2 10 6  
## 3 8 3

Pruebe si el vector de medias muestrales igual al vector , usando el nivel de significancia

### Solución

Para este caso se tiene una prueba de hipótesis del vector de medias y con desconocido, nos planteamos las siguientes hipótesis:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.01 ; alpha

## [1] 0.01

#### 3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar

Donde y es la matriz de covarianza de la muestra, hallemos los resultados en R

#Covarianza de la muestra  
S <- cov(X\_1) ; S

## x1 x2  
## x1 4 -3  
## x2 -3 9

#Vector de medias muestrales  
xbar <- colMeans(X\_1) ; xbar

## x1 x2   
## 8 6

#Valor mu\_0  
mu0 <- matrix(c(9,5),2,1) ; mu0

## [,1]  
## [1,] 9  
## [2,] 5

#Vector de diferencia d  
d <- xbar - mu0 ; d

## [,1]  
## [1,] -1  
## [2,] 1

#Cantidad de datos en la muestra  
n=length(X\_1$x1) ; n

## [1] 3

#Cantidad de variables  
p=dim(X\_1)[2] ; p

## [1] 2

#Hallar el valor F calculado  
fis\_c <- ((n - p) \* t(d) %\*% solve(S) %\*% d )/p  
fis\_c

## [,1]  
## [1,] 0.1296296

#### 4) Hallar el valor de la región crítica

fis\_q <- qf(1-alpha, df1 = p, df2 = n - p)  
fis\_q

## [1] 4999.5

De la misma forma con el pvalor

pf(fis\_c, df1 = p, df2 = n-p, lower.tail = F)

## [,1]  
## [1,] 0.8911328

#### 5) Conclusión

Como , no se rechaza por lo tanto, a un nivel de significancia el vector de medias muestrales es igual al vector

## Ejercicio 2:

Se tiene la matriz de datos X\_2:

X\_2 = data.frame(x1 = c(2, 8, 6, 8),  
 x2 = c(12, 9, 9, 10)) ; X\_2

## x1 x2  
## 1 2 12  
## 2 8 9  
## 3 6 9  
## 4 8 10

Pruebe si el vector de medias muestrales igual al vector , usando el nivel de significancia

### Solución

Para este caso se tiene una prueba de hipótesis del vector de medias y con desconocido, nos planteamos las siguientes hipótesis:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.03 ; alpha

## [1] 0.03

#### 3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar

Donde y es la matriz de covarianza de la muestra, hallemos los resultados en R

#Covarianza de la muestra  
S <- cov(X\_2) ; S

## x1 x2  
## x1 8.000000 -3.333333  
## x2 -3.333333 2.000000

#Vector de medias muestrales  
xbar <- colMeans(X\_2) ; xbar

## x1 x2   
## 6 10

#Valor mu\_0  
mu0 <- matrix(c(7,11),2,1) ; mu0

## [,1]  
## [1,] 7  
## [2,] 11

#Vector de diferencia d  
d <- xbar - mu0 ; d

## [,1]  
## [1,] -1  
## [2,] -1

#Cantidad de datos en la muestra  
n=length(X\_2$x1) ; n

## [1] 4

#Cantidad de variables  
p=dim(X\_2)[2] ; p

## [1] 2

#Hallar el valor F calculado  
fis\_c <- ((n - p) \* t(d) %\*% solve(S) %\*% d )/p  
fis\_c

## [,1]  
## [1,] 3.409091

#### 4) Hallar el valor de la región crítica

fis\_q <- qf(1-alpha, df1 = p, df2 = n - p)  
fis\_q

## [1] 32.33333

De la misma forma con el pvalor

pf(fis\_c, df1 = p, df2 = n-p, lower.tail = F)

## [,1]  
## [1,] 0.2268041

#### 5) Conclusión

Como , no se rechaza por lo tanto, a un nivel de significancia el vector de medias muestrales es igual al vector

## Ejercicio 3:

Se analizó la transpiración de 20 mujeres saludables: Se midieron tres componentes

* x1: tasa de sudoración
* x2: contenido de sodio
* x3: contenido de potasio.

El cual el resultado es el siguiente

x1 <- c(3.7, 5.7, 3.8, 3.2, 3.1, 4.6, 2.4, 7.2, 6.7, 5.4,  
 3.9, 4.5, 3.5, 4.5, 1.5, 8.5, 4.5, 6.5, 4.1, 5.5)  
x2 <- c(48.5, 65.1, 47.2, 53.2, 55.5, 36.1, 24.8, 33.1, 47.4, 54.1,  
 36.9, 58.8, 27.8, 40.2, 13.5, 56.4, 71.6, 52.8, 44.1, 40.9)  
x3 <- c(9.3, 8.0, 10.9, 12.0, 9.7, 7.9, 14.0, 7.6, 8.5, 11.3,  
 12.7, 12.3, 9.8, 8.4, 10.1, 7.1, 8.2, 10.9, 11.2, 9.4)  
X3 <- data.frame(x1 = x1,  
 x2 = x2,  
 x3 = x3)  
X3

## x1 x2 x3  
## 1 3.7 48.5 9.3  
## 2 5.7 65.1 8.0  
## 3 3.8 47.2 10.9  
## 4 3.2 53.2 12.0  
## 5 3.1 55.5 9.7  
## 6 4.6 36.1 7.9  
## 7 2.4 24.8 14.0  
## 8 7.2 33.1 7.6  
## 9 6.7 47.4 8.5  
## 10 5.4 54.1 11.3  
## 11 3.9 36.9 12.7  
## 12 4.5 58.8 12.3  
## 13 3.5 27.8 9.8  
## 14 4.5 40.2 8.4  
## 15 1.5 13.5 10.1  
## 16 8.5 56.4 7.1  
## 17 4.5 71.6 8.2  
## 18 6.5 52.8 10.9  
## 19 4.1 44.1 11.2  
## 20 5.5 40.9 9.4

Pruebe si el vector de medias muestrales igual al vector , usando el nivel de significancia

### Solución

Para este caso se tiene una prueba de hipótesis del vector de medias y con desconocido, nos planteamos las siguientes hipótesis:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.03 ; alpha

## [1] 0.03

#### 3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar

Donde y es la matriz de covarianza de la muestra, hallemos los resultados en R

#Covarianza de la muestra  
S <- cov(X3) ; S

## x1 x2 x3  
## x1 2.879368 10.0100 -1.809053  
## x2 10.010000 199.7884 -5.640000  
## x3 -1.809053 -5.6400 3.627658

#Vector de medias muestrales  
xbar <- colMeans(X3) ; xbar

## x1 x2 x3   
## 4.640 45.400 9.965

#Valor mu\_0  
mu0 <- matrix(c(4,50,10),3,1) ; mu0

## [,1]  
## [1,] 4  
## [2,] 50  
## [3,] 10

#Vector de diferencia d  
d <- xbar - mu0 ; d

## [,1]  
## [1,] 0.640  
## [2,] -4.600  
## [3,] -0.035

#Cantidad de datos en la muestra  
n=length(X3$x1) ; n

## [1] 20

#Cantidad de variables  
p=dim(X3)[2] ; p

## [1] 3

#Hallar el valor F calculado  
fis\_c <- ((n - p) \* t(d) %\*% solve(S) %\*% d )/p  
fis\_c

## [,1]  
## [1,] 2.759319

#### 4) Hallar el valor de la región crítica

fis\_q <- qf(1-alpha, df1 = p, df2 = n - p)  
fis\_q

## [1] 3.791176

De la misma forma con el pvalor

pf(fis\_c, df1 = p, df2 = n-p, lower.tail = F)

## [,1]  
## [1,] 0.07411877

#### 5) Conclusión

Como , no se rechaza por lo tanto, a un nivel de significancia el vector de medias muestrales de la transpiración de las 20 mujeres saludables es igual al vector

## Ejercicio 4:

Tenemos una muestra de 15 mujeres y 12 hombres. Se presenta la media de los valores de las diferentes variables medidas.

nm <- 15 #Cantidad de mujeres  
nh <- 12 #Cantidad de hombres  
p <- 7 #Cantidad de variables  
#Medias muestrales de mujeres  
xbarm=matrix(c(168.78,63.89,38.98,73.46,45.85,57.24,43.09),7,1) ; xbarm

## [,1]  
## [1,] 168.78  
## [2,] 63.89  
## [3,] 38.98  
## [4,] 73.46  
## [5,] 45.85  
## [6,] 57.24  
## [7,] 43.09

#Medias muestrales de hombres  
xbarh=matrix(c(177.58,74.25,41.67,77.75,49.00,58.00,45.62),7,1) ; xbarh

## [,1]  
## [1,] 177.58  
## [2,] 74.25  
## [3,] 41.67  
## [4,] 77.75  
## [5,] 49.00  
## [6,] 58.00  
## [7,] 45.62

Y las matrices de covarianzas

#Matriz de covarianzas de mujeres  
Sm=matrix(c(37.64,22.10,6.38,15.65,9.49,2.75,9.02,  
 22.10,80.40,7.36,12.94,14.39,7.20,9.31,  
 6.38,7.36,1.92,3.06,1.49,0.76,1.98,  
 15.65,12.94,3.06,7.41,3.99,1.17,4.53,  
 9.49,14.39,1.49,3.99,9.42,2.559,1.12,  
 2.75,7.2,0.76,1.17,2.559,2.94,0.95,  
 9.02,9.31,1.98,4.53,1.12,0.95,3.78),7,7) ; Sm

## [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
## [1,] 37.64 22.10 6.38 15.65 9.490 2.750 9.02  
## [2,] 22.10 80.40 7.36 12.94 14.390 7.200 9.31  
## [3,] 6.38 7.36 1.92 3.06 1.490 0.760 1.98  
## [4,] 15.65 12.94 3.06 7.41 3.990 1.170 4.53  
## [5,] 9.49 14.39 1.49 3.99 9.420 2.559 1.12  
## [6,] 2.75 7.20 0.76 1.17 2.559 2.940 0.95  
## [7,] 9.02 9.31 1.98 4.53 1.120 0.950 3.78

#Matriz de covarianzas de hombres  
Sh=matrix(c(45.53,48.84,9.48,14.34,14.86,9.45,8.92,  
 48.84,74.20,9.63,19.34,19.77,9.90,5.23,  
 9.48,9.63,2.79,2.09,3.23,1.86,2.31,  
 14.34,19.34,2.09,12.57,6.18,2.36,1.21,  
 14.86,19.77,3.23,6.18,6.77,3.02,1.84,  
 9.45,9.90,1.86,2.36,3.02,3.13,2.63,  
 8.92,5.23,2.31,1.21,1.84,2.63,6.14),7,7) ; Sh

## [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
## [1,] 45.53 48.84 9.48 14.34 14.86 9.45 8.92  
## [2,] 48.84 74.20 9.63 19.34 19.77 9.90 5.23  
## [3,] 9.48 9.63 2.79 2.09 3.23 1.86 2.31  
## [4,] 14.34 19.34 2.09 12.57 6.18 2.36 1.21  
## [5,] 14.86 19.77 3.23 6.18 6.77 3.02 1.84  
## [6,] 9.45 9.90 1.86 2.36 3.02 3.13 2.63  
## [7,] 8.92 5.23 2.31 1.21 1.84 2.63 6.14

Pruebe si existen diferencias detectables entre las dos muestras, con un nivel de significancia

### Solución

Para este caso se tiene una prueba de hipótesis en la que se desea saber si hay diferencia de medias:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.06 ; alpha

## [1] 0.06

#### 3) Hallar el estadístico de prueba

# calculo de la matriz de var y cov conjunta  
Sp= ((nm-1)\*Sm + (nh-1)\*Sh)/(nm+nh-2)  
Sp

## [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
## [1,] 41.1116 33.8656 7.7440 15.0736 11.85280 5.69800 8.9760  
## [2,] 33.8656 77.6720 8.3588 15.7560 16.75720 8.38800 7.5148  
## [3,] 7.7440 8.3588 2.3028 2.6332 2.25560 1.24400 2.1252  
## [4,] 15.0736 15.7560 2.6332 9.6804 4.95360 1.69360 3.0692  
## [5,] 11.8528 16.7572 2.2556 4.9536 8.25400 2.76184 1.4368  
## [6,] 5.6980 8.3880 1.2440 1.6936 2.76184 3.02360 1.6892  
## [7,] 8.9760 7.5148 2.1252 3.0692 1.43680 1.68920 4.8184

Hallar el valor

# estadistico de prueba  
T2=((nm\*nh)/(nm+nh))\* (t(xbarm-xbarh))%\*%solve(Sp)%\*%(xbarm-xbarh)  
T2

## [,1]  
## [1,] 26.08934

n <- nm + nh ; n

## [1] 27

Hallar el F calculado

Fc=((n-p-1)/(p\*(n-2)))\*T2  
Fc

## [,1]  
## [1,] 2.832557

#### 4) Hallar el valor de la región crítica

Ft <- qf(1-alpha, df1 = p, df2 = n - p - 1)  
Ft

## [1] 2.414015

#### 5) Conclusión

Como , se rechaza por lo tanto, a un nivel de significancia se tienen diferencias detectables entre la muestra de hombres y mujeres

## Ejercicio 5:

En un país, el gobierno federal exige que el Departamento de Control de Calidad de toda fábrica de hornos microondas monitoree la cantidad de radiación emitida cuando las puertas del horno están cerradas y cuando éstas están abiertas. Se observaron las radiaciones emitidas por 42 hornos elegidos al azar. Los datos aparecen en la tabla, con la puerta abierta y con la puerta cerrada.

1. Hacer un Q-Q-plot con los datos univariados y además testear su normalidad.
2. Una transformación de Box y Cox que mejora la normalidad de los datos para la puerta cerrada se obtiene . Aplicar la transformación a ambas variables y comprobarlo a través de nuevos Q-Q-plots y pruebas de normalidad.
3. Hallar para los datos transformados.
4. Comprobar que los datos transformados efectivamente siguen una distribución , hallar la elipse de confianza de nivel simultáneo 0.95 , dar sus direcciones principales, la longitud de sus ejes y hacer un gráfico aproximado.
5. Testear versus con nivel 0.05
6. Testear versus con nivel 0.05
7. Testear versus con nivel 0.05

DatoHornos <- data.frame(Horno = 1:42,  
 Abierto = c(0.3, 0.09, 0.3, 0.1, 0.1, 0.12, 0.09, 0.1,  
 0.24, 0.1, 0.07, 0.05, 0.04, 0.45, 0.12, 0.2,  
 0.04, 0.01, 0.01, 0.6, 0.12, 0.1, 0.35, 0.3,  
 0.15, 0.3, 0.15, 0.09, 0.09, 0.28, 0.1, 0.1, 0.1, 0.5,  
 0.12, 0.25, 0.3, 0.4, 0.2, 0.32, 0.2, 0.12),  
 Cerrado = c(0.15, 0.09, 0.18, 0.1, 0.05, 0.12, 0.08, 0.05,  
 0.1, 0.08, 0.12, 0.02, 0.1, 0.1, 0.1, 0.1,  
 0.02, 0.1, 0.01, 0.4, 0.1, 0.05, 0.03, 0.05,  
 0.1, 0.15, 0.15, 0.01, 0.08, 0.18, 0.1, 0.2,  
 0.11, 0.3, 0.02, 0.06, 0.2, 0.3, 0.4, 0.41, 0.3, 0.05))  
DatoHornos

## Horno Abierto Cerrado  
## 1 1 0.30 0.15  
## 2 2 0.09 0.09  
## 3 3 0.30 0.18  
## 4 4 0.10 0.10  
## 5 5 0.10 0.05  
## 6 6 0.12 0.12  
## 7 7 0.09 0.08  
## 8 8 0.10 0.05  
## 9 9 0.24 0.10  
## 10 10 0.10 0.08  
## 11 11 0.07 0.12  
## 12 12 0.05 0.02  
## 13 13 0.04 0.10  
## 14 14 0.45 0.10  
## 15 15 0.12 0.10  
## 16 16 0.20 0.10  
## 17 17 0.04 0.02  
## 18 18 0.01 0.10  
## 19 19 0.01 0.01  
## 20 20 0.60 0.40  
## 21 21 0.12 0.10  
## 22 22 0.10 0.05  
## 23 23 0.35 0.03  
## 24 24 0.30 0.05  
## 25 25 0.15 0.10  
## 26 26 0.30 0.15  
## 27 27 0.15 0.15  
## 28 28 0.09 0.01  
## 29 29 0.09 0.08  
## 30 30 0.28 0.18  
## 31 31 0.10 0.10  
## 32 32 0.10 0.20  
## 33 33 0.10 0.11  
## 34 34 0.50 0.30  
## 35 35 0.12 0.02  
## 36 36 0.25 0.06  
## 37 37 0.30 0.20  
## 38 38 0.40 0.30  
## 39 39 0.20 0.40  
## 40 40 0.32 0.41  
## 41 41 0.20 0.30  
## 42 42 0.12 0.05

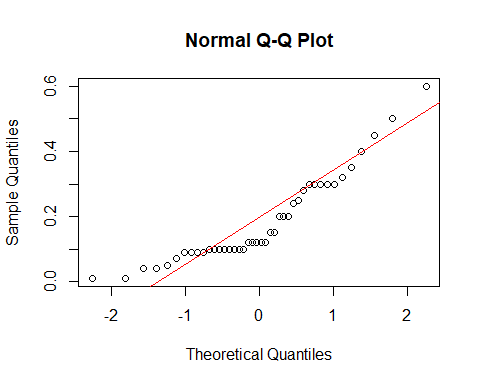
### Solución

1. **Hacer un Q-Q-plot con los datos univariados y además testear su normalidad.**

#### Para el horno abierto

Realizar el gráfico qqplot

qqnorm(DatoHornos$Abierto)  
qqline(DatoHornos$Abierto, col = "red")



Realizar las pruebas de normalidad

#Libreria a utilizar  
library(nortest)

# Shapiro Wilk  
shapiro.test(DatoHornos$Abierto)

##   
## Shapiro-Wilk normality test  
##   
## data: DatoHornos$Abierto  
## W = 0.8795, p-value = 0.0003706

# Anderson darling  
nortest::ad.test(DatoHornos$Abierto)

##   
## Anderson-Darling normality test  
##   
## data: DatoHornos$Abierto  
## A = 1.9007, p-value = 6.157e-05

# cramer vonn mises  
nortest::cvm.test(DatoHornos$Abierto)

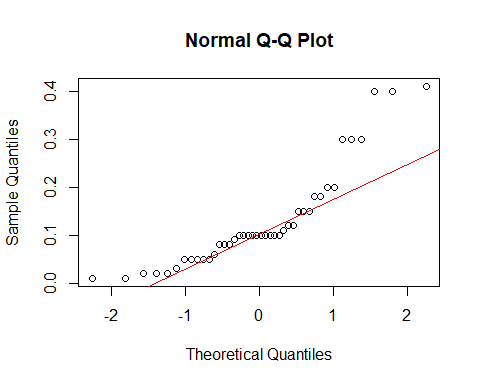
##   
## Cramer-von Mises normality test  
##   
## data: DatoHornos$Abierto  
## W = 0.35168, p-value = 7.873e-05

En las tres hipótesis y por el gráfico QQPlot se observa que los datos no siguen una distribución normal.

#### Para el horno cerrado

Realizar el gráfico qqplot

qqnorm(DatoHornos$Cerrado)  
qqline(DatoHornos$Cerrado, col = "red")



Realizar las pruebas de normalidad

# Shapiro Wilk  
shapiro.test(DatoHornos$Cerrado)

##   
## Shapiro-Wilk normality test  
##   
## data: DatoHornos$Cerrado  
## W = 0.8264, p-value = 1.703e-05

# Anderson darling  
nortest::ad.test(DatoHornos$Cerrado)

##   
## Anderson-Darling normality test  
##   
## data: DatoHornos$Cerrado  
## A = 2.5997, p-value = 1.121e-06

# cramer vonn mises  
nortest::cvm.test(DatoHornos$Cerrado)

##   
## Cramer-von Mises normality test  
##   
## data: DatoHornos$Cerrado  
## W = 0.45327, p-value = 6.824e-06

En las tres hipótesis y por el gráfico QQPlot se observa que los datos no siguen una distribución normal.

1. **Una transformación de Box y Cox que mejora la normalidad de los datos para la puerta cerrada se obtiene . Aplicar la transformación a ambas variables y comprobarlo a través de nuevos Q-Q-plots y pruebas de normalidad.**

Realizar la transformación

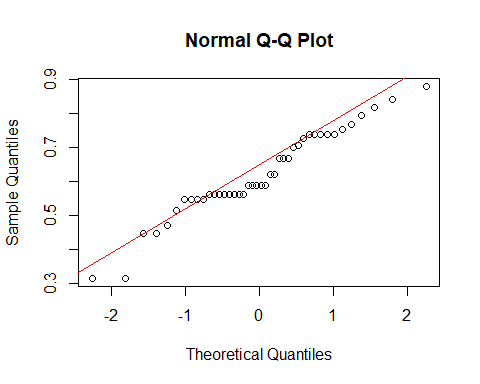
DatoHornos$AbiertoBoxCox <- DatoHornos$Abierto^(1/4)  
DatoHornos$CerradoBoxCox <- DatoHornos$Cerrado^(1/4)

Ver la normalidad en ambos grupos

#### Para el horno abierto

Realizar el gráfico qqplot

qqnorm(DatoHornos$AbiertoBoxCox)  
qqline(DatoHornos$AbiertoBoxCox, col = "red")



Realizar las pruebas de normalidad

# Shapiro Wilk  
shapiro.test(DatoHornos$AbiertoBoxCox)

##   
## Shapiro-Wilk normality test  
##   
## data: DatoHornos$AbiertoBoxCox  
## W = 0.9571, p-value = 0.1162

# Anderson darling  
nortest::ad.test(DatoHornos$AbiertoBoxCox)

##   
## Anderson-Darling normality test  
##   
## data: DatoHornos$AbiertoBoxCox  
## A = 0.79063, p-value = 0.03709

# cramer vonn mises  
nortest::cvm.test(DatoHornos$AbiertoBoxCox)

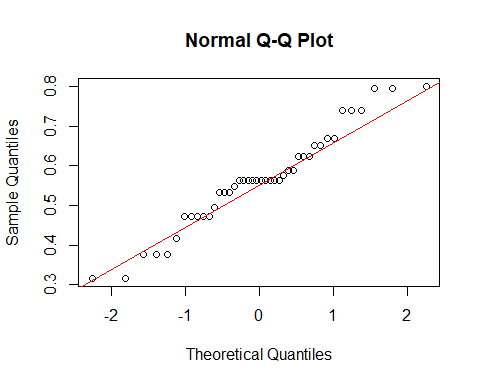
##   
## Cramer-von Mises normality test  
##   
## data: DatoHornos$AbiertoBoxCox  
## W = 0.14704, p-value = 0.02473

Se observa que solo en la prueba de Shapiro Wilk los datos del horno abierto transformado sigue una distribución normal. Se podría tomar como una distribución normal pero no de forma tan certera.

#### Para el horno cerrado

Realizar el gráfico qqplot

qqnorm(DatoHornos$CerradoBoxCox)  
qqline(DatoHornos$CerradoBoxCox, col = "red")



Realizar las pruebas de normalidad

# Shapiro Wilk  
shapiro.test(DatoHornos$CerradoBoxCox)

##   
## Shapiro-Wilk normality test  
##   
## data: DatoHornos$CerradoBoxCox  
## W = 0.95991, p-value = 0.1466

# Anderson darling  
nortest::ad.test(DatoHornos$CerradoBoxCox)

##   
## Anderson-Darling normality test  
##   
## data: DatoHornos$CerradoBoxCox  
## A = 0.66526, p-value = 0.07666

# cramer vonn mises  
nortest::cvm.test(DatoHornos$CerradoBoxCox)

##   
## Cramer-von Mises normality test  
##   
## data: DatoHornos$CerradoBoxCox  
## W = 0.12336, p-value = 0.05165

En las tres hipótesis y por el gráfico QQPlot se observa que los datos siguen una distribución normal.

1. **Hallar para los datos transformados.**

Primeramente hallemos el vector de medias muestrales

# Extraer los datos  
Y <- DatoHornos[,c(4,5)] ; Y

## AbiertoBoxCox CerradoBoxCox  
## 1 0.7400828 0.6223330  
## 2 0.5477226 0.5477226  
## 3 0.7400828 0.6513556  
## 4 0.5623413 0.5623413  
## 5 0.5623413 0.4728708  
## 6 0.5885662 0.5885662  
## 7 0.5477226 0.5318296  
## 8 0.5623413 0.4728708  
## 9 0.6999271 0.5623413  
## 10 0.5623413 0.5318296  
## 11 0.5143687 0.5885662  
## 12 0.4728708 0.3760603  
## 13 0.4472136 0.5623413  
## 14 0.8190363 0.5623413  
## 15 0.5885662 0.5623413  
## 16 0.6687403 0.5623413  
## 17 0.4472136 0.3760603  
## 18 0.3162278 0.5623413  
## 19 0.3162278 0.3162278  
## 20 0.8801117 0.7952707  
## 21 0.5885662 0.5623413  
## 22 0.5623413 0.4728708  
## 23 0.7691606 0.4161791  
## 24 0.7400828 0.4728708  
## 25 0.6223330 0.5623413  
## 26 0.7400828 0.6223330  
## 27 0.6223330 0.6223330  
## 28 0.5477226 0.3162278  
## 29 0.5477226 0.5318296  
## 30 0.7274272 0.6513556  
## 31 0.5623413 0.5623413  
## 32 0.5623413 0.6687403  
## 33 0.5623413 0.5759014  
## 34 0.8408964 0.7400828  
## 35 0.5885662 0.3760603  
## 36 0.7071068 0.4949232  
## 37 0.7400828 0.6687403  
## 38 0.7952707 0.7400828  
## 39 0.6687403 0.7952707  
## 40 0.7521206 0.8001952  
## 41 0.6687403 0.7400828  
## 42 0.5885662 0.4728708

# Hallar el vector de medias  
ybar <- colMeans(Y) ; ybar

## AbiertoBoxCox CerradoBoxCox   
## 0.6211651 0.5636649

Ahora hallar la matriz de varianza, covarianza muestral (S)

S <- cov(Y) ; S

## AbiertoBoxCox CerradoBoxCox  
## AbiertoBoxCox 0.016114466 0.008891283  
## CerradoBoxCox 0.008891283 0.014751703

Y la matriz inversa de varianza covarianza

Sinv <- solve(S) ; Sinv

## AbiertoBoxCox CerradoBoxCox  
## AbiertoBoxCox 92.97629 -56.03953  
## CerradoBoxCox -56.03953 101.56545

1. **Comprobar que los datos transformados efectivamente siguen una distribución , hallar la elipse de confianza de nivel simultáneo 0.95 , dar sus direcciones principales, la longitud de sus ejes y hacer un gráfico aproximado.**

Veamos si los datos siguen una distribución multivariada con los test de Mardia, Henze-Zirkler y Royston

#Libreria a utilizar  
library(MVN)

Mardia = mvn(Y, mvn\_test = "mardia")  
Mardia$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Mardia Skewness 1.153 0.886 asymptotic ✓ Normal  
## 2 Mardia Kurtosis 0.312 0.755 asymptotic ✓ Normal

La prueba de Mardia indica que se tiene una distribución normal bivariada, sigamos con el test de Henze-Zirkler

HZ =mvn(Y, mvn\_test = "hz")   
HZ$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Henze-Zirkler 0.545 0.322 asymptotic ✓ Normal

El test de Henz también indica que es normal bivariada, ahora el test de Royston

Roy =mvn(Y, mvn\_test = "royston")   
Roy$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Royston 4.624 0.101 asymptotic ✓ Normal

En los tres se indica que se tiene una distribución normal, por lo tanto los datos siguen una distribución normal bivariada. Ahora hallemos la elipse de nivel simultáneo al 95% de confianza

*NOTA: Esta parte me hice ayudar de ChatGPT ya que todavía no se avanzó esa parte*

#Libreria a utilizar  
library(ellipse)

##   
## Attaching package: 'ellipse'

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

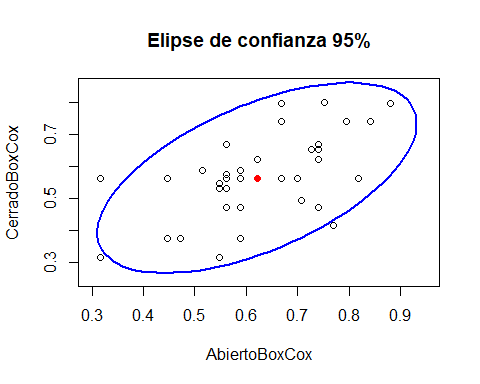
# Valores propios y vectores propios  
eig <- eigen(S)  
eig$values # Longitudes cuadradas de los ejes principales

## [1] 0.024350438 0.006515731

eig$vectors # Direcciones principales

## [,1] [,2]  
## [1,] -0.7336248 0.6795547  
## [2,] -0.6795547 -0.7336248

# Longitudes de los semiejes (ajustados al nivel de confianza)  
# Para 2 variables, F quantile: qf(0.95, 2, n-2) \* 2\*(n-1)/(n\*(n-2))  
n <- nrow(Y)  
f\_quantile <- qf(0.95, 2, n - 2)  
radius <- sqrt(2 \* (n - 1) \* f\_quantile / (n \* (n - 2)))  
  
semiaxes <- sqrt(eig$values) \* radius  
  
# Gráfico de elipse  
plot(Y, main = "Elipse de confianza 95%", xlim = c(0.3, 0.95), ylim = c(0.25, 0.85))  
lines(ellipse(S, centre = ybar, level = 0.95), col = "blue", lwd = 2)  
points(ybar[1], ybar[2], pch = 19, col = "red")



1. **Testear versus con nivel 0.05**

Para este caso se tiene una prueba de hipótesis del vector de medias y con desconocido, nos planteamos las siguientes hipótesis:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.05 ; alpha

## [1] 0.05

#### 3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar

Donde y es la matriz de covarianza de la muestra, hallemos los resultados en R

#Valor mu\_0  
mu0 <- matrix(c(0.562,0.589),2,1) ; mu0

## [,1]  
## [1,] 0.562  
## [2,] 0.589

#Vector de diferencia d  
d <- ybar - mu0 ; d

## [,1]  
## [1,] 0.05916505  
## [2,] -0.02533507

#Cantidad de variables  
p=dim(Y)[2] ; p

## [1] 2

#Hallar el valor F calculado  
fis\_c <- ((n - p) \* t(d) %\*% solve(S) %\*% d )/p  
fis\_c

## [,1]  
## [1,] 11.17312

#### 4) Hallar el valor de la región crítica

fis\_q <- qf(1-alpha, df1 = p, df2 = n - p)  
fis\_q

## [1] 3.231727

De la misma forma con el pvalor

pf(fis\_c, df1 = p, df2 = n-p, lower.tail = F)

## [,1]  
## [1,] 0.0001396346

#### 5) Conclusión

Como , se rechaza por lo tanto, a un nivel de significancia el vector de medias muestrales de los hornos es diferente al vector

1. **Testear versus con nivel 0.05**

Para este caso se tiene una prueba de hipótesis del vector de medias y con desconocido, nos planteamos las siguientes hipótesis:

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.05 ; alpha

## [1] 0.05

#### 3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar

Donde y es la matriz de covarianza de la muestra, hallemos los resultados en R

#Valor mu\_0  
mu0 <- matrix(c(0.55,0.60),2,1) ; mu0

## [,1]  
## [1,] 0.55  
## [2,] 0.60

#Vector de diferencia d  
d <- ybar - mu0 ; d

## [,1]  
## [1,] 0.07116505  
## [2,] -0.03633507

#Cantidad de variables  
p=dim(Y)[2] ; p

## [1] 2

#Hallar el valor F calculado  
fis\_c <- ((n - p) \* t(d) %\*% solve(S) %\*% d )/p  
fis\_c

## [,1]  
## [1,] 17.89557

#### 4) Hallar el valor de la región crítica

fis\_q <- qf(1-alpha, df1 = p, df2 = n - p)  
fis\_q

## [1] 3.231727

De la misma forma con el pvalor

pf(fis\_c, df1 = p, df2 = n-p, lower.tail = F)

## [,1]  
## [1,] 2.810814e-06

#### 5) Conclusión

Como , se rechaza por lo tanto, a un nivel de significancia el vector de medias muestrales de los hornos es diferente al vector

1. **Testear versus con nivel 0.05**

En este caso se pide comparar si hay diferencia entre el horno abierto y el horno cerrado con los datos transformados, en el cual viendo que en ambos casos se tienen que siguen una distribución normal, se utilizaría una prueba t de comparaciones de medias, adicionalmente los registros son del mismo horno por lo que se tiene datos pareados. Realizemos la prueba de hipótesis

#### 1) Formular las hipótesis

#### 2) Establecer el nivel de significancia

alpha <- 0.05 ; alpha

## [1] 0.05

#### 3) Hallar el estadístico de prueba

Prueba T para datos pareados

Hallar las estimaciones insesgadas, primero hallemos el vector de diferencias

Y$Diferencia <- Y$AbiertoBoxCox - Y$CerradoBoxCox ; Y$Diferencia

## [1] 0.11774983 0.00000000 0.08872724 0.00000000 0.08947052 0.00000000  
## [7] 0.01589297 0.08947052 0.13758578 0.03051174 -0.07419752 0.09681050  
## [13] -0.11512773 0.25669493 0.02622487 0.10639898 0.07115329 -0.24611356  
## [19] 0.00000000 0.08484101 0.02622487 0.08947052 0.35298142 0.26721200  
## [25] 0.05999165 0.11774983 0.00000000 0.23149479 0.01589297 0.07607159  
## [31] 0.00000000 -0.10639898 -0.01356012 0.10081361 0.21250588 0.21218358  
## [37] 0.07134250 0.05518792 -0.12653042 -0.04807462 -0.07134250 0.11569539

Hallemos el valor en R

dbar <- mean(Y$Diferencia) ; dbar

## [1] 0.05750012

Y veamos si esta diferencia sigue una distribución normal

shapiro.test(Y$Diferencia)

##   
## Shapiro-Wilk normality test  
##   
## data: Y$Diferencia  
## W = 0.97132, p-value = 0.3653

El cual sigue una distribución normal

De la misma forma ahora hallemos

De la misma forma hallemos el resultado en R

ddesv <- sd(Y$Diferencia) ; ddesv

## [1] 0.1143836

Y hallemos el valor calculado

De la misma forma hallemos el resultado en R

tc <- (dbar - 0)/(ddesv/sqrt(n)) ; tc

## [1] 3.25784

#### 4) Hallar el valor de la región crítica

tq <- qt(1-alpha, df = n-1)  
tq

## [1] 1.682878

De la misma forma con el pvalor

pt(tc, df = n-1, lower.tail = F)

## [1] 0.001129325

#### 5) Conclusión

Como , se rechaza por lo tanto, a un nivel de significancia hay diferencias en la radiación de las puertas del horno, si se encuentran abiertos o cerrados.

## Ejercicio 6

En una planta automotriz, se monitorean tres características de una pieza crítica: longitud (mm), peso (g) y dureza (Rockwell). Estas características siguen una distribución normal trivariada con:

E\_X <- matrix(c(25, 200, 60),3,1) ; E\_X

## [,1]  
## [1,] 25  
## [2,] 200  
## [3,] 60

V <- matrix(c(1.2, 3, 0.5, 3, 25, 2, 0.5, 2, 4),3,3) ; V

## [,1] [,2] [,3]  
## [1,] 1.2 3 0.5  
## [2,] 3.0 25 2.0  
## [3,] 0.5 2 4.0

1. ¿Cuál es la probabilidad de que una pieza tenga dureza mayor a 65?
2. ¿Cuál es la densidad condicional de la dureza dado que la longitud fue 24 mm y el peso 190 g?
3. ¿Simule 100 observaciones y grafique las elipses de confianza bivariadas (longitud vs peso, peso vs dureza)?

### Solución

1. **¿Cuál es la probabilidad de que una pieza tenga dureza mayor a 65?**

Hallemos los casos para cada caso y resolver por la estandarización

Reemplazemos el valor en la distribución normal estándar

Ahora calcular la probabilidad de la dureza mayor a 65

pnorm(2.5, lower.tail = FALSE)

## [1] 0.006209665

El cual indica que la probabilidad de que una pieza tenga dureza mayor a 65 es aproximadamente del 0.62%

1. **¿Cuál es la densidad condicional de la dureza dado que la longitud fue 24 mm y el peso 190 g?**

Extraemos los datos de interes

* El vector condicional

Resolvamos en R

Sigma\_12\_12 <- V[c(1,2),c(1,2)] ; Sigma\_12\_12

## [,1] [,2]  
## [1,] 1.2 3  
## [2,] 3.0 25

Sigma\_3\_12 <- V[3,c(1,2), drop = FALSE] ; Sigma\_3\_12

## [,1] [,2]  
## [1,] 0.5 2

# Diferencia:  
x12\_resta\_mu12 <- matrix(c(24 - 25, 190 - 200), nrow=2) ; x12\_resta\_mu12

## [,1]  
## [1,] -1  
## [2,] -10

# Producto:  
mean\_cond <- 60 + Sigma\_3\_12 %\*% solve(Sigma\_12\_12) %\*% x12\_resta\_mu12 ; mean\_cond

## [,1]  
## [1,] 59.2619

# Varianza condicional:  
var\_cond <- 4 - Sigma\_3\_12 %\*% solve(Sigma\_12\_12) %\*% t(Sigma\_3\_12) ; var\_cond

## [,1]  
## [1,] 3.759524

El cual indica que la distribución condicional de la dureza sigue la siguiente distribución

1. **¿Simule 100 observaciones y grafique las elipses de confianza bivariadas (longitud vs peso, peso vs dureza)?**

# Librerias a utilizar  
library(MASS)  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

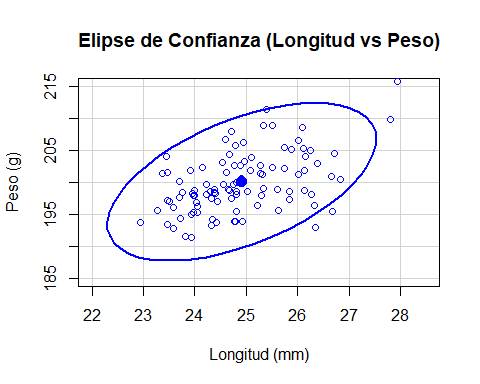
## The following object is masked from 'package:ellipse':  
##   
## ellipse

simular <- mvrnorm(n = 100, mu = as.vector(E\_X), Sigma = V)  
colnames(simular) <- c("Longitud", "Peso", "Dureza")  
simular <- as.data.frame(simular) ; simular

## Longitud Peso Dureza  
## 1 23.46894 201.6181 59.69791  
## 2 23.43931 203.9906 60.80211  
## 3 24.78916 193.8335 56.78280  
## 4 26.13321 198.7358 58.79835  
## 5 24.80160 198.5744 58.60447  
## 6 25.33710 199.0088 60.05361  
## 7 24.60242 206.6937 60.00470  
## 8 24.76755 193.8221 61.27182  
## 9 25.61774 195.5498 59.64168  
## 10 24.80134 195.5298 63.04036  
## 11 24.15029 202.3878 57.73484  
## 12 24.56285 199.6118 62.37413  
## 13 24.43187 197.0600 60.44450  
## 14 25.74539 205.5221 59.44184  
## 15 24.92537 193.8839 59.38368  
## 16 24.90056 200.6002 61.37534  
## 17 26.12683 201.9124 59.48598  
## 18 24.40974 198.2034 57.95150  
## 19 25.51683 208.9462 59.55282  
## 20 27.71856 217.5196 63.69399  
## 21 23.51264 197.0852 55.83859  
## 22 23.95930 198.2310 59.53617  
## 23 25.39573 211.3569 61.53991  
## 24 24.97258 203.3400 61.48726  
## 25 24.79661 205.7176 59.93189  
## 26 22.94392 193.7732 57.20584  
## 27 24.22930 199.6760 59.20831  
## 28 26.33319 196.4643 58.29612  
## 29 24.34505 194.2663 57.95887  
## 30 23.58284 192.8699 60.12748  
## 31 27.80770 209.7574 64.25940  
## 32 26.01220 201.2120 61.55859  
## 33 24.88963 202.6184 58.11731  
## 34 24.76331 202.5982 65.19385  
## 35 26.82170 200.4823 61.74456  
## 36 24.67177 198.7570 57.52994  
## 37 24.42610 193.7767 60.89647  
## 38 26.34892 192.9578 62.30046  
## 39 26.01118 206.6103 59.83830  
## 40 23.70087 200.0947 58.36781  
## 41 24.80545 200.0467 60.94961  
## 42 26.71801 204.5224 60.41420  
## 43 25.73916 202.1466 64.15386  
## 44 24.52856 203.1087 60.00117  
## 45 23.76770 198.3649 56.88286  
## 46 23.26778 195.5317 57.35588  
## 47 23.45693 193.4735 60.61590  
## 48 25.51432 202.3805 61.02164  
## 49 25.08422 201.8453 59.66891  
## 50 23.72014 194.3998 60.40220  
## 51 24.04245 195.2292 58.68580  
## 52 27.94052 215.8044 61.65018  
## 53 24.00123 197.9110 61.00465  
## 54 24.31693 197.4892 59.67276  
## 55 24.31266 193.2346 59.67918  
## 56 24.71434 207.9078 61.70072  
## 57 25.31639 201.2246 59.89263  
## 58 23.81921 191.5073 60.75334  
## 59 23.36677 201.4099 56.86953  
## 60 25.84811 198.5974 60.79827  
## 61 24.05260 196.2290 57.84750  
## 62 23.70846 197.7129 59.93964  
## 63 26.38206 203.0287 61.10242  
## 64 24.75088 199.6306 63.23453  
## 65 26.67388 195.4650 61.11391  
## 66 24.31044 198.6005 59.94980  
## 67 25.10004 203.8684 60.64952  
## 68 24.70944 197.4453 61.74955  
## 69 26.15611 204.0236 61.85622  
## 70 23.99485 198.6958 59.47356  
## 71 25.60062 198.9050 59.07702  
## 72 24.95041 206.2587 61.57782  
## 73 24.37865 198.4478 61.12765  
## 74 24.80165 198.1405 54.65727  
## 75 26.65795 200.8561 60.77849  
## 76 26.25810 198.0777 58.27953  
## 77 25.26811 202.7021 62.15142  
## 78 23.94121 191.4340 58.85571  
## 79 23.98046 195.2309 62.01879  
## 80 25.83260 197.2625 63.74006  
## 81 25.33378 208.8694 58.11663  
## 82 24.62104 201.4889 56.38087  
## 83 24.23065 198.0945 58.38474  
## 84 23.45869 197.1978 59.82100  
## 85 25.21681 196.3874 60.43052  
## 86 23.58700 196.0066 60.84325  
## 87 25.29529 198.0309 60.61428  
## 88 26.09120 208.5934 61.99426  
## 89 25.88397 205.0947 60.34379  
## 90 24.65258 198.9179 57.40703  
## 91 23.92649 195.0301 59.64832  
## 92 24.66406 204.3004 60.62142  
## 93 23.92198 201.8997 60.28801  
## 94 25.01384 198.5463 58.89244  
## 95 26.24051 204.9426 61.43068  
## 96 24.38665 198.9601 59.78010  
## 97 24.03137 196.9134 58.14965  
## 98 26.12031 205.2992 58.40943  
## 99 25.26804 201.4479 58.39695  
## 100 23.96501 197.9785 58.72092

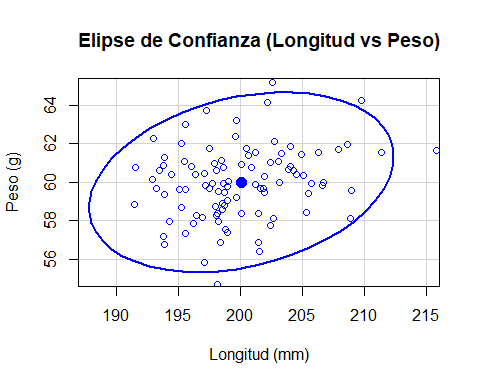
#### Longitud vs Peso

dataEllipse(simular$Longitud, simular$Peso,   
 xlim = c(22,28.5), ylim = c(185,215),  
 xlab = "Longitud (mm)", ylab = "Peso (g)",  
 main = "Elipse de Confianza (Longitud vs Peso)",  
 levels = 0.95,   
 col = "blue",   
 lwd = 2)



#### Peso vs Dureza

dataEllipse(simular$Peso, simular$Dureza,   
 xlim = c(188,215), ylim = c(55,65),  
 xlab = "Longitud (mm)", ylab = "Peso (g)",  
 main = "Elipse de Confianza (Longitud vs Peso)",  
 levels = 0.95,   
 col = "blue",   
 lwd = 2)



## Ejercicio 7

Se estudia la respuesta a un tratamiento en pacientes diabéticos midiendo: glucosa en sangre, presión arterial sistólica y frecuencia cardíaca. Se modelan como una variable aleatoria normal trivariada:

E\_X <- matrix(c(90, 130, 72),3,1) ; E\_X

## [,1]  
## [1,] 90  
## [2,] 130  
## [3,] 72

SIGMA <- matrix(c(25, 10, 5, 10, 100, 8, 5, 8, 36),3,3) ; SIGMA

## [,1] [,2] [,3]  
## [1,] 25 10 5  
## [2,] 10 100 8  
## [3,] 5 8 36

1. ¿Cuál es la distribución del promedio muestral si se toman 25 pacientes?
2. Hallar el intervalo simultáneo de confianza del 95% para los tres parámetros

### Solución

1. **¿Cuál es la distribución del promedio muestral si se toman 25 pacientes?**
2. **Hallar el intervalo simultáneo de confianza del 95% para los tres parámetros**

## Ejercicio 8

A 60 estudiantes se les aplican tres pruebas: memoria verbal, razonamiento lógico y velocidad de procesamiento. Se asume una distribución normal trivariada:

mu <- matrix(c(100, 105, 95)) ; mu

## [,1]  
## [1,] 100  
## [2,] 105  
## [3,] 95

Sigma <- matrix(c(15, 10, 5, 10, 20, 8, 5, 8, 10),3,3) ; Sigma

## [,1] [,2] [,3]  
## [1,] 15 10 5  
## [2,] 10 20 8  
## [3,] 5 8 10

1. ¿Cuál es la distribución de la puntuación total combinada?
2. Si se observa una velocidad de procesamiento de 100, ¿cuál es la distribución condicional del resto?

### Solución

Aca va la solución

## Ejercicio 9

Se recopilan datos semanales sobre temperatura , humedad relativa y velocidad del viento en una ciudad. Se modela:

E\_X <- matrix(c(22, 60, 12),3,1) ; E\_X

## [,1]  
## [1,] 22  
## [2,] 60  
## [3,] 12

Sigma <- matrix(c(4, -2, 0.5, -2, 9, 1.5, 0.5, 1.5, 1),3,3) ; Sigma

## [,1] [,2] [,3]  
## [1,] 4.0 -2.0 0.5  
## [2,] -2.0 9.0 1.5  
## [3,] 0.5 1.5 1.0

1. Hallar la distribución de una combinación lineal:
2. ¿Cuál es la probabilidad de que el índice climático sea > 27?

### Solución

Aca va la solución