Análisis Multivariado

Producto académico 01

Kevin Heberth Haquehua Apaza

14 de julio del 2025

Tabla de contenidos

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 $\mu_0 = (9{,}5)$, usando el nivel de significancia $\alpha = 0.01$

Solución

Para este caso se tiene una prueba de hipótesis del vector de medias $N_p(\mu, \Sigma)$ y con Σ desconocido, nos planteamos las siguientes hipótesis:

1) Formular las hipótesis

- $H_0 : \mu = \binom{9}{5}$
- $H_a : \mu \neq \binom{9}{5}$

2) Establecer el nivel de significancia

$$\alpha = 0.01$$

```
alpha <- 0.01 ; alpha

## [1] 0.01
```

3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar H_0

$$\frac{(n-p)d^{-1}S^{-1}d}{p} > F_{(p,n-p,\alpha)}$$

Donde $d = \bar{x} - \mu_0$ y S 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 $F_c^2 < F_t^2$, no se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.01$ el vector de medias muestrales es igual al vector $\mu_0 = (9,5)$

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 $\mu_0 = (7{,}11)$, usando el nivel de significancia $\alpha = 0.03$

Solución

Para este caso se tiene una prueba de hipótesis del vector de medias $N_p(\mu, \Sigma)$ y con Σ desconocido, nos planteamos las siguientes hipótesis:

1) Formular las hipótesis

- $H_0: \mu = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$
- $H_a: \mu \neq \begin{pmatrix} 7 \\ 11 \end{pmatrix}$

2) Establecer el nivel de significancia

$$\alpha = 0.03$$

```
alpha <- 0.03 ; alpha
```

```
## [1] 0.03
```

3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar H_0

$$\frac{(n-p)d^{-1}S^{-1}d}{p} > F_{(p,n-p,\alpha)}$$

Donde $d = \bar{x} - \mu_0$ y S 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 $F_c^2 < F_t^2$, no se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.03$ el vector de medias muestrales es igual al vector $\mu_0 = (7{,}11)$

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 $\mu_0 = (4{,}50{,}10)$, usando el nivel de significancia $\alpha = 0.03$

Solución

Para este caso se tiene una prueba de hipótesis del vector de medias $N_p(\mu, \Sigma)$ y con Σ desconocido, nos planteamos las siguientes hipótesis:

1) Formular las hipótesis

- $H_0: \mu = \begin{pmatrix} 4 \\ 50 \\ 10 \end{pmatrix}$
- $H_a: \mu \neq \begin{pmatrix} 4 \\ 50 \\ 10 \end{pmatrix}$

2) Establecer el nivel de significancia

$$\alpha = 0.03$$

```
alpha <- 0.03 ; alpha

## [1] 0.03
```

3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar H_0

$$\frac{(n-p)d^{-1}S^{-1}d}{p} > F_{(p,n-p,\alpha)}$$

Donde $d = \bar{x} - \mu_0$ y S 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 $F_c^2 < F_t^2$, no se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.03$ el vector de medias muestrales de la transpiración de las 20 mujeres saludables es igual al vector $\mu_0 = (4{,}50{,}10)$

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
```

```r
#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

```r
#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
```

```r
#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 $\alpha = 0.06$

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

- $H_0: \bar{\mu_m} - \bar{\mu_h} = 0$
- $H_a: \bar{\mu_m} - \bar{\mu_h} \neq 0$

2) Establecer el nivel de significancia

$$\alpha = 0.06$$

```
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 T^2

```
# 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 $F_c^2 > F_t^2$, se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.06$ 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.

a) Hacer un Q-Q-plot con los datos univariados y además testear su normalidad.
b) Una transformación de Box y Cox que mejora la normalidad de los datos para la puerta cerrada se obtiene $\lambda = 0.25$. Aplicar la transformación a ambas variables $y_{ij} = x_{ij}^{1/4}, j = 1,2$ y comprobarlo a través de nuevos Q-Q-plots y pruebas de normalidad.
c) Hallar $\bar{y}, S, S^{-1}$ para los datos transformados.
d) Comprobar que los datos transformados efectivamente siguen una distribución $N_2(\mu, \Sigma)$, hallar la elipse de confianza de nivel simultáneo 0.95 , dar sus direcciones principales, la longitud de sus ejes y hacer un gráfico aproximado.
e) Testear $H_0: \mu = (0.562,0.589)$ versus $H_1: \mu \neq (0.562,0.589)$ con nivel 0.05
f) Testear $H_0: \mu = (0.55,0.60)$ versus $H_1: \mu \neq (0.55,0.60)$ con nivel 0.05
g) Testear $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ 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, 0.0
5,
                                     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
```

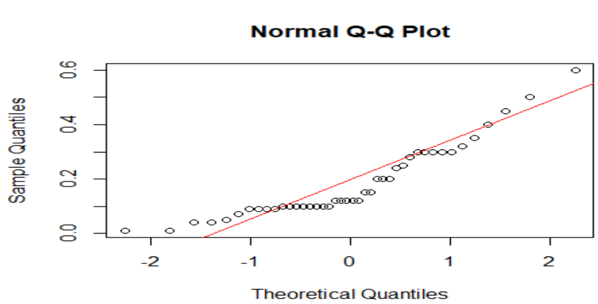
Solución

a) **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

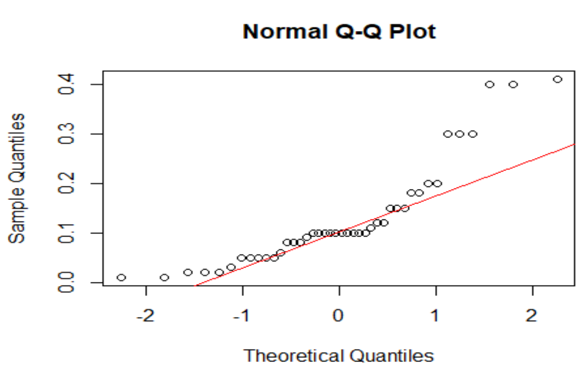
```
#Libreria a utilizar
library(nortest)

# 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.

b) **Una transformación de Box y Cox que mejora la normalidad de los datos para la puerta cerrada se obtiene $\lambda = 0.25$. Aplicar la transformación a ambas variables $y_{ij} = x_{ij}^{1/4}, j = 1,2$ y comprobarlo a través de nuevos Q-Q-plots y pruebas de normalidad.**

Realizar la transformación

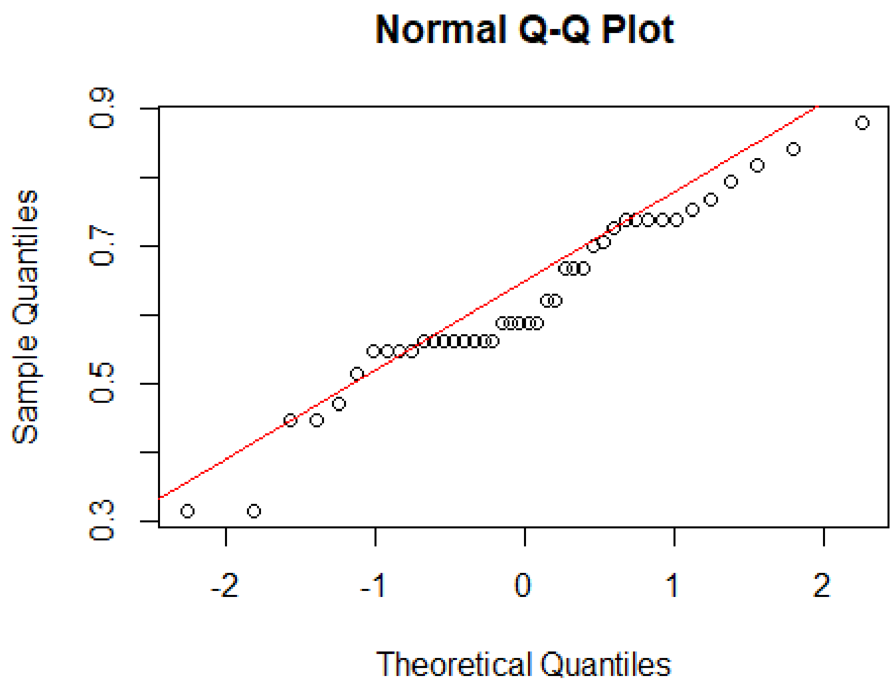
```
DatoHornos$AbiertoBoxCox <- DatoHornos$Abierto^(1/4)
DatoHornos$CerradoBoxCox <- DatoHornos$Cerrado^(1/4)
```

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

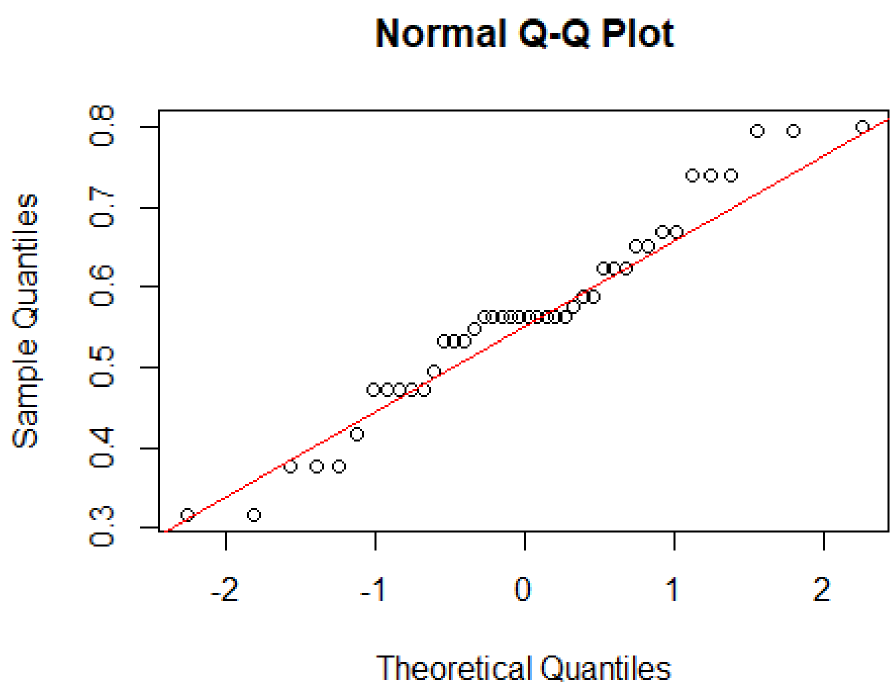
```
# Shapiro Wilk
shapiro.test(DatoHornos$AbiertoBoxCox)

##
## 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.

c) **Hallar $\bar{y}, S, S^{-1}$ 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 (S^{-1})

```
Sinv <- solve(S) ; Sinv

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

d) **Comprobar que los datos transformados efectivamente siguen una distribución $N_2(\mu, \Sigma)$, 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

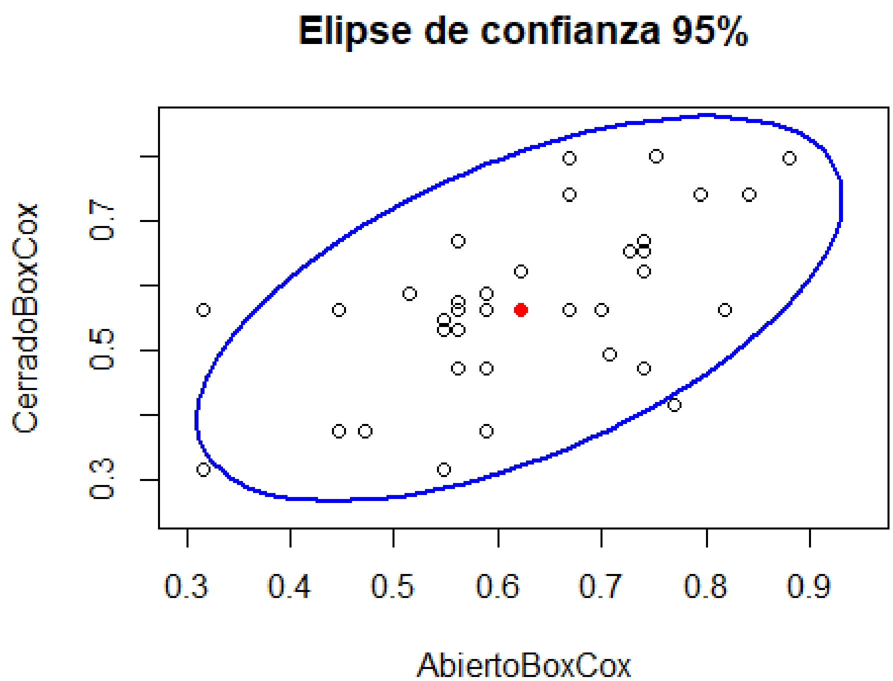
```
Roy =mvn(Y, mvn_test = "royston")
Roy$multivariate_normality

##     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")
```



e) **Testear $H_0: \mu = (0.562, 0.589)$ versus $H_1: \mu \neq (0.562, 0.589)$ con nivel 0.05**

Para este caso se tiene una prueba de hipótesis del vector de medias $N_p(\mu, \Sigma)$ y con Σ desconocido, nos planteamos las siguientes hipótesis:

1) Formular las hipótesis

- $H_0: \mu = \begin{pmatrix} 0.562 \\ 0.589 \end{pmatrix}$
- $H_a: \mu \neq \begin{pmatrix} 0.562 \\ 0.589 \end{pmatrix}$

2) Establecer el nivel de significancia

$$\alpha = 0.05$$

```
alpha <- 0.05 ; alpha

## [1] 0.05
```

3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar H_0

$$\frac{(n-p)d^{-1}S^{-1}d}{p} > F_{(p,n-p,\alpha)}$$

Donde $d = \bar{x} - \mu_0$ y S 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 $F_c^2 > F_t^2$, se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.05$ el vector de medias muestrales de los hornos es diferente al vector $\mu_0 = (0.562, 0.589)$

f) **Testear $H_0: \mu = (0.55, 0.60)$ versus $H_1: \mu \neq (0.55, 0.60)$ con nivel 0.05**

Para este caso se tiene una prueba de hipótesis del vector de medias $N_p(\mu, \Sigma)$ y con Σ desconocido, nos planteamos las siguientes hipótesis:

1) Formular las hipótesis

- $H_0: \mu = \begin{pmatrix} 0.55 \\ 0.60 \end{pmatrix}$
- $H_a: \mu \neq \begin{pmatrix} 0.55 \\ 0.60 \end{pmatrix}$

2) Establecer el nivel de significancia

$$\alpha = 0.05$$

```
alpha <- 0.05 ; alpha

## [1] 0.05
```

3) Hallar el estadístico de prueba

Con la varianza desconocida se utiliza la siguiente fórmula para rechazar H_0

$$\frac{(n-p)d^{-1}S^{-1}d}{p} > F_{(p,n-p,\alpha)}$$

Donde $d = \bar{x} - \mu_0$ y S 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 $F_c^2 > F_t^2$, se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.05$ el vector de medias muestrales de los hornos es diferente al vector $\mu_0 = (0.55,0.60)$

g) **Testear $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ 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

- $H_0: \mu_{\text{Abierto}} = \mu_{\text{Cerrado}}$
- $H_a: \mu_{\text{Abierto}} \neq \mu_{\text{Cerrado}}$

2) Establecer el nivel de significancia

$$\alpha = 0.05$$

```
alpha <- 0.05 ; alpha

## [1] 0.05
```

3) Hallar el estadístico de prueba

Prueba T para datos pareados

$$T_{exp} = \frac{\bar{d} - \mu_d}{\frac{\hat{S}_d}{\sqrt{n}}} \sim t_{n-1} = t_{16}$$

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
```

$$\bar{d} = \frac{0.12 + 0 + 0.09 + \cdots + (-0.07) + 0.12}{42}$$

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 $\hat{S}_d$

$$\hat{S}_d = \sqrt{\frac{(0.12 - 0.0575)^2 + \cdots + (0.12 - 0.0575)^2}{42 - 1}}$$

De la misma forma hallemos el resultado en R

```
ddesv <- sd(Y$Diferencia) ; ddesv

## [1] 0.1143836
```

Y hallemos el valor calculado

$$T_{exp} = \frac{0.0575 - 0}{\frac{0.1144}{\sqrt{42}}}$$

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 $t_c > t_t$, se rechaza H_0 por lo tanto, a un nivel de significancia $\alpha = 0.05$ 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
```

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

Solución

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

Hallemos los valores y resolvamos por la estandarización

$$\bar{x}_{dureza} = 65, \mu_{dureza} = 60, \sigma_{dureza} = \sqrt{4} = 2$$

Reemplazemos el valor en la distribución normal estándar

$$\begin{aligned} Z &= \frac{\bar{x}_{dureza} - \mu_{dureza}}{\sigma_{dureza}} \\ Z &= \frac{65 - 60}{2} \\ Z &= 2.5 \end{aligned}$$

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%

b) **¿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

- $\mu_3 = 60$
- $\mu_{12} = \begin{pmatrix} 25 \\ 200 \end{pmatrix}$
- $(\Sigma_{31} \quad \Sigma_{32}) = (0.5 \quad 2)$
- $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 1.2 & 3 \\ 3 & 25 \end{pmatrix}$
- $\Sigma_{33} = 4$
- El vector condicional $x_{12} = \begin{pmatrix} 24 \\ 190 \end{pmatrix}$

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

$$X_3 | X_1 = 24, X_2 = 190 \sim N(\mu = 59.26, \sigma^2 = 3.76)$$

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

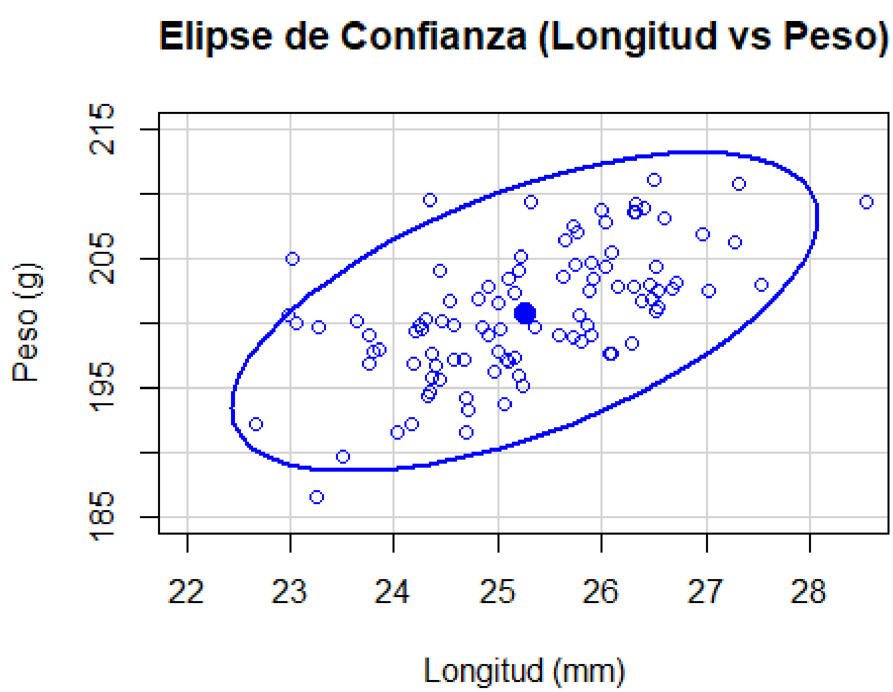
```
# Librerias a utilizar
library(MASS)
library(car)

## 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  23.06632 199.9470 60.79119
## 2  22.67176 192.1982 59.00015
## 3  26.07864 197.6141 57.11098
## 4  24.67349 197.2150 59.74945
## 5  26.32128 209.2128 64.39751
## 6  24.34872 194.6740 60.19679
## 7  25.99904 208.6780 60.60356
## 8  24.35440 197.6006 59.06215
## 9  25.19605 204.0093 62.12137
## 10 27.01923 202.5452 60.09876
## 11 24.69673 191.4768 58.13940
## 12 24.46738 200.0689 59.22364
## 13 24.80220 201.8237 61.05216
## 14 26.54900 201.2097 64.05309
## 15 23.85327 197.9570 57.97113
## 16 23.75322 196.8589 62.95476
## 17 25.34436 199.6861 60.66779
## 18 24.44294 195.5376 58.30402
## 19 24.29584 200.2394 60.93935
## 20 25.16456 197.3987 59.64814
## 21 27.52347 202.8793 61.52805
## 22 26.52368 200.8732 62.98699
## 23 25.23526 195.2073 58.83063
## 24 25.07241 197.0996 61.99119
## 25 26.28854 198.4972 61.19428
## 26 26.46684 202.9985 59.48476
## 27 27.31096 210.7438 61.40574
## 28 25.20766 205.0648 61.70425
## 29 26.08650 205.3785 59.72006
## 30 26.03425 207.8510 59.73118
## 31 24.02212 191.6153 59.98797
## 32 23.02045 205.0312 55.88903
## 33 25.15969 202.3128 60.39030
## 34 24.26827 199.5476 60.52191
## 35 24.68928 194.2122 57.22364
```

```
## 36  24.91431 199.0957 59.18539
## 37  23.27031 199.7269 62.81487
## 38  24.57796 197.1502 61.66952
## 39  25.91082 203.3868 62.00529
## 40  24.24392 199.7989 59.54657
## 41  26.30204 208.5763 61.08699
## 42  24.96587 196.2637 59.59704
## 43  25.79929 198.6111 60.86925
## 44  26.67935 202.6548 61.02142
## 45  24.89851 202.8030 60.23110
## 46  25.06189 193.8000 57.53307
## 47  25.00575 197.7704 61.72183
## 48  25.10330 203.3902 59.53588
## 49  24.17938 196.9222 58.04396
## 50  25.88866 204.6749 61.32741
## 51  26.33351 208.5522 58.19017
## 52  24.44288 204.0888 57.89392
## 53  25.58359 199.0644 62.18552
## 54  25.85045 199.7824 54.77701
## 55  23.24322 186.6196 61.86472
## 56  25.78243 200.5411 61.22890
## 57  25.31109 209.3556 58.89433
## 58  26.39150 201.6509 58.89731
## 59  25.72597 207.5104 63.38685
## 60  26.15689 202.7572 60.32175
## 61  26.48558 201.8065 63.60417
## 62  26.09412 197.7187 61.78138
## 63  26.52157 204.3361 61.73196
## 64  25.90618 198.9722 60.35676
## 65  24.99854 201.5206 59.12592
## 66  25.09767 196.9895 61.34349
## 67  24.72013 193.2717 60.59456
## 68  27.28686 206.2578 60.25581
## 69  26.40878 208.8684 61.34382
## 70  24.32343 194.3830 58.05322
## 71  24.23905 199.8856 58.11660
## 72  26.59825 208.1387 59.86189
## 73  26.70476 203.0860 56.97030
## 74  23.78981 197.7426 57.75362
## 75  28.54689 209.3637 59.37380
## 76  24.84052 199.7307 59.79189
## 77  26.30448 202.7688 64.82254
## 78  26.54292 202.5198 60.20508
## 79  26.96830 206.8160 57.81476
## 80  23.64774 200.2143 60.88942
## 81  25.87329 202.5537 58.74651
## 82  25.19395 195.8835 59.23386
## 83  23.76515 199.0851 57.72971
## 84  22.98275 200.5881 61.15700
## 85  25.62319 203.5459 62.39030
## 86  25.75547 207.0208 60.44322
## 87  26.04158 204.3709 62.06200
## 88  24.15940 192.2433 59.43886
```

```
## 89  24.58445 199.8859 55.71815
## 90  24.39211 196.7236 59.59988
## 91  24.33376 209.5829 57.23955
## 92  25.65433 206.4100 59.92951
## 93  24.53601 201.7555 61.22785
## 94  25.01918 199.4589 57.21966
## 95  25.72737 198.9069 62.29197
## 96  23.50413 189.7457 59.82126
## 97  24.20451 199.3035 60.18448
## 98  24.35573 195.7701 59.37596
## 99  26.49328 211.0813 59.29119
## 100 25.73311 204.5125 59.78697
```

Longitud vs Peso

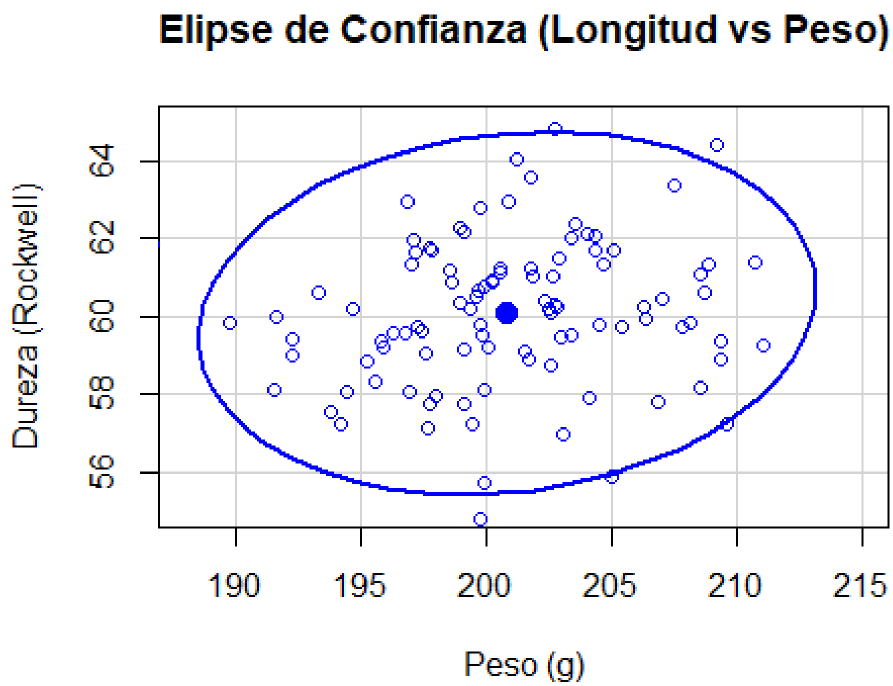
```
dataEllipse(simular$Longitud, simular$Peso,
            xlim = c(22,28.5), ylim = c(185,215),
            xlab = "Longitud (mm)", ylab = "Peso (g)",
            main = "Elipse de Confianza (Longitud vs Peso)",
            levels = 0.95,
            col = "blue",
            lwd = 2)
```



Peso vs Dureza

```
dataEllipse(simular$Peso, simular$Dureza,
            xlim = c(188,215), ylim = c(55,65),
            xlab = "Peso (g)", ylab = "Dureza (Rockwell)",
            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
```

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

Solución

a) **¿Cuál es la distribución del promedio muestral si se toman 25 pacientes?**

La distribución de la muestra de los 25 pacientes tiene la siguiente distribución

$$\bar{X} \sim N\left(\mu = \begin{pmatrix} 90 \\ 130 \\ 72 \end{pmatrix}, \frac{\Sigma}{25} = \begin{pmatrix} 1 & 0.4 & 0.2 \\ 0.4 & 4 & 0.32 \\ 0.2 & 0.32 & 1.44 \end{pmatrix}\right)$$

b) **Hallar el intervalo simultáneo de confianza del 95% para los tres parámetros**

Tomemos el tamaño de muestra de los 25 pacientes y un nivel de significancia $\alpha = 0.05$

```
n <- 25
alpha <- 0.05
p <- dim(SIGMA)[2]

# Método Bonferroni
alpha_bonf <- alpha / (2 * p)
t_bonf <- qt(1 - alpha_bonf, df = n - 1)

#error estandar
sde <- sqrt(diag(SIGMA) / n)

# Límites de los intervalos
limite_inferior <- E_X - t_bonf * sde
limite_superior <- E_X + t_bonf * sde
```

```r
# Mostrar intervalos
intervalos <- data.frame(
  Variable = c("Glucosa", "Presion", "Frecuencia"),
  Inferior = round(limite_inferior, 2),
  Superior = round(limite_superior, 2)
)

print(intervalos)

##     Variable Inferior Superior
## 1    Glucosa    87.43    92.57
## 2    Presion   124.85   135.15
## 3 Frecuencia    68.91    75.09
```

Ejercicio 8

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

```r
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
```

a) ¿Cuál es la distribución de la puntuación total combinada?
b) Si se observa una velocidad de procesamiento de 100, ¿cuál es la distribución condicional del resto?

Solución

a) **¿Cuál es la distribución de la puntuación total combinada?**

Se tiene que

$$\bar{X} = (X_1, X_2, X_3) \sim N\left(\mu = \begin{pmatrix} 100 \\ 105 \\ 95 \end{pmatrix}, \Sigma = \begin{pmatrix} 15 & 10 & 5 \\ 10 & 20 & 8 \\ 5 & 8 & 10 \end{pmatrix}\right)$$

La puntuación total combinada estaría dado por

$$T = X_1 + X_2 + X_3 \sim N(\mu_T, \sigma_T^2)$$

Donde

- $\mu_T = 100 + 105 + 95 = 300$
- $\sigma_T^2 = \sum Var(X_i) + 2Cov(X_i, X_j)$

Hallando los cálculos en R

```
# Media total
mu_T <- sum(mu) ; mu_T

## [1] 300

# Varianza total combinada
sigma_T2 <- sum(Sigma) ; sigma_T2

## [1] 91
```

En el cual se tiene que la distribución de la puntuación total combinada se distribuye de la siguiente manera

$$T \sim N(300{,}91)$$

b) **Si se observa una velocidad de procesamiento de 100, ¿cuál es la distribución condicional del resto?**

Se pide hallar

$$(X_1, X_2)|X_3 = 100$$

Extraemos los datos de interes

- $(\mu_1, \mu_2) = (100{,}105)$
- $\mu_2 = 95$
- $(\Sigma_{31} \quad \Sigma_{32}) = (0.5 \quad 2)$
- $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 15 & 10 \\ 10 & 20 \end{pmatrix}$
- $\begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$
- $\Sigma_{33} = 10$
- El vector condicional $x_3 = 100$

Resolvamos en R

```
x3 <- 100

mu_12 <- mu[1:2, , drop = FALSE] ; mu_12

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

mu_3 <- mu[3] ; mu_3
```

```
## [1] 95

Sigma_12_12 <- Sigma[c(1,2),c(1,2)] ; Sigma_12_12

##      [,1] [,2]
## [1,]   15   10
## [2,]   10   20

Sigma_12_3 <- Sigma[c(1,2),3, drop = FALSE] ; Sigma_12_3

##      [,1]
## [1,]    5
## [2,]    8

Sigma_33 <- Sigma[3,3, drop = FALSE] ; Sigma_33

##      [,1]
## [1,]   10

# Media condicional
mucond <- mu_12 + Sigma_12_3 %*% solve(Sigma_33) %*% (x3 - mu_3) ; mucond

##       [,1]
## [1,] 102.5
## [2,] 109.0

# Covarianza condicional
Sigmacond <- Sigma_12_12 - Sigma_12_3 %*% solve(Sigma_33) %*% t(Sigma_12_3) ; Sigmac
ond

##      [,1] [,2]
## [1,] 12.5  6.0
## [2,]  6.0 13.6
```

El cual indica que la distribución condicional de la memoria verba y razonamiento lógico dado la velocidad del procesamiento de 100 sigue la siguiente distribución

$$(X_1, X_2)|X_3 = 100 \sim N\left(\mu = \begin{pmatrix} 102.5 \\ 109 \end{pmatrix}, \Sigma = \begin{pmatrix} 12.5 & 6 \\ 6 & 13.6 \end{pmatrix}\right)$$

Ejercicio 9

Se recopilan datos semanales sobre temperatura ($°C$), humedad relativa ($\%$) y velocidad del viento (km/h) 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
```

a) Hallar la distribución de una combinación lineal:

$$\text{índice climático} = 0.5 \cdot \text{Temp} + 0.3 \cdot \text{Hum} - 0.2 \cdot \text{Viento}$$

b) ¿Cuál es la probabilidad de que el índice climático sea > 27?

Solución

a) **Hallar la distribución de una combinación lineal:**

Sea $X_1 = Temperatura, X_2 = Humedad \quad y \quad X_3 = Viento$ y el índice climático como Y se tiene la siguiente expresión

$$Y = 0.5X_1 + 0.3X_2 - 0.2X_3 = (0.5 \quad 0.3 \quad -0.2)\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = AX$$

Además

$$Si \quad X \sim N_3 \to Y = AX \sim N_3(\mu_Y = A\mu_X, \Sigma_Y = A\Sigma A^t)$$

Donde

- $\mu_Y = (0.5 \quad 0.3 \quad -0.2)\begin{pmatrix} 22 \\ 60 \\ 12 \end{pmatrix}$
- $\Sigma_Y = (0.5 \quad 0.3 \quad -0.2)\begin{pmatrix} 4 & -2 & 0.5 \\ -2 & 9 & 1.5 \\ 0.5 & 1.5 & 1 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.3 \\ -0.2 \end{pmatrix}$

Hallemos los resultados en R

```
(A = matrix(c(0.5,0.3,-0.2),1,3))

##      [,1] [,2] [,3]
## [1,]  0.5  0.3 -0.2

muY = A %*% E_X ; muY

##      [,1]
## [1,] 26.6

SY = A %*% Sigma %*% t(A) ; SY

##      [,1]
## [1,] 0.97

sdY = sqrt(SY) ; sdY
```

```
##       [,1]
## [1,] 0.9848858
```

En el que se tiene que el índice climático (Y) sigue una distribución normal expresado de la siguiente forma

$$Y \sim N(\mu = 26.6, \sigma^2 = 0.97)$$

b) **¿Cuál es la probabilidad de que el índice climático sea > 27?**

Hallemos los valores y resolvamos por la estandarización

$$\bar{y} = 27, \mu = 26.6, \sigma = \sqrt{0.97} = 0.98$$

Reemplazemos el valor en la distribución normal estándar

$$\begin{aligned} Z &= \frac{\bar{y} - \mu}{\sigma} \\ Z &= \frac{27 - 26.6}{0.98} \end{aligned}$$

Hallemos el valor en R

```
Z = (27 - muY)/sdY ; Z

##       [,1]
## [1,] 0.4061385
```

Ahora calcular la probabilidad de la dureza mayor a 27

```
pnorm(Z, lower.tail = FALSE)

##       [,1]
## [1,] 0.3423204
```

El cual indica que la probabilidad de que el índice climático sea mayor a 27 es aproximadamente del 34.23%