Análisis Multivariado

Producto académico 01

Kevin Heberth Haquehua Apaza

14 de julio del 2025

# 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

Aca va la solución

## 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

Aca va la solución

## 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

Aca va la solución

## 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