Centro de Investigación en Matemáticas (CIMAT). Unidad Monterrey

Estadítica Multivariada
Tarea 3

Marcelo Alberto Sanchez Zaragoza

20 de marzo de 2021

Problema 1

La matriz de datos para una muestra aleatoria de tamaño $n = 3$ de una población bivariada está dada por:

$$X = \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{pmatrix}$$

a) Verifica que T^2 permanece sin cambios sin cada observación $x_j, j = 1, 2, 3$ es reemplazada por Cx_j, donde

$$C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Note que las observaciones

$$Cx_j = \begin{pmatrix} x_{j1} - x_{j2} \\ x_{j1} + x_{j2} \end{pmatrix}$$

Producen la matriz

$$\begin{pmatrix} (6-9) & (10-6) & (8-3) \\ (6+9) & (10+6) & (8+3) \end{pmatrix}^t$$

Solución

Inciso a)

Sabemos como se define T^2 y solo debemos encontrar los valores nuevos con el cambio solicitado y revisar que no hay cambio. Primero encontramos los nuevos registros, en este caso van a ser: Cx_1, Cx_2 y Cx_3.

Expresando las respectivas multiplicaciones:

$$Cx_1 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 9 \end{pmatrix} = \begin{pmatrix} -3 \\ 15 \end{pmatrix}$$

$$Cx_2 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 16 \end{pmatrix}$$

$$Cx_3 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix}$$

Finalmente nuestra nueva matriz es:

$$C_1 = \begin{pmatrix} -3 & 15 \\ 4 & 16 \\ 5 & 11 \end{pmatrix}$$

Ahora vamos a analizar el valor que resulta para T^2, en este caso vamos a tomar un vector $\mu_0 = (9,5)^t$, donde al realizar el cambio respectivo tenemos que es igual a $C\mu_0 = (4,14)^t$. Mostrando nuestras matrices involucradas tenemos:

$$\bar{x}_c = \begin{pmatrix} 2 \\ 14 \end{pmatrix} \quad S_c = \begin{pmatrix} 19 & -5 \\ -5 & 7 \end{pmatrix} \quad S_c^{-1} = \begin{pmatrix} 7/108 & 5/108 \\ 5/108 & 19/108 \end{pmatrix}$$

La expresión que tenemos para encontrar T^2 es la siguiente:

$$T^2 = 3 \begin{pmatrix} 2-4 & 14-14 \end{pmatrix} \begin{pmatrix} 7/108 & 5/108 \\ 5/108 & 19/108 \end{pmatrix} \begin{pmatrix} 2-4 \\ 14-14 \end{pmatrix} = 3\frac{7}{27} = \frac{7}{9}$$

Nuestro valor con los cambios solicitados dio un resultado de $T^2 = \frac{7}{9}$ y sin el cambio da como resultado:

$$T^2 = 3 \begin{pmatrix} 8-9 & 6-5 \end{pmatrix} \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix} \begin{pmatrix} 8-9 \\ 6-5 \end{pmatrix} = 3\frac{7}{27} = \frac{7}{9}$$

Finalmente realizando la misma operación con los valores originales vemos que el valor de T^2 no cambio.

Problema 2

Dadas las siguiente muestra de observaciones bivariadas:

$$X = \begin{pmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{pmatrix}$$

a) Evalua T^2,para probar $H_0 : \mu^t = [7, 11]$.

b) Especifica la distribución de T^2.

c) Usando a) y b) prueba H_0 en $\alpha = 0{,}5$. ¿Que conclusión tiene?

d) Evalue T^2 utilizando la relación que tiene con la lambda de Wilks.

e) Evalue Λ y la lambda de Wilks.

Solución

Inciso a)

Para desarrollar este primer inciso vamos a encontrar las matrices que corresponden a $\bar{x}$ y S. Las cuales son:

$$\bar{x} = \begin{pmatrix} 6 \\ 10 \end{pmatrix} \quad S = \begin{pmatrix} 8 & -10/3 \\ -10/3 & 2 \end{pmatrix} \quad S^{-1} = \begin{pmatrix} 9/22 & 15/22 \\ 15/22 & 18/11 \end{pmatrix}$$

Ahora vamos a encontrar el valor de T^2:

$$T^2 = 4 \begin{pmatrix} 6-7 & 10-11 \end{pmatrix} \begin{pmatrix} 9/22 & 15/22 \\ 15/22 & 18/11 \end{pmatrix} \begin{pmatrix} 6-7 \\ 10-11 \end{pmatrix} = 4\frac{75}{22} = \frac{150}{11}$$

El valor de $T^2 = \frac{150}{11}$.

Inciso b)
La distribución que va a tener nuestra T^2 la vamos a poder obtener de la siguiente relación que existe: $T \sim \frac{(n-1)p}{n-p} F_{p,n-p}$, en este caso tenemos que $p = 2$ y $n = 2$, finalmente tenemos que $T \sim 3F_{2,2}$.

Inciso c)
Nos proporcionan un $\alpha = 0{,}5$ para analizar nuestra hipotesis H_0. Dado que ya encontramos el valor de T^2 vamos a buscar nuestro valor critico, el valor de $3F_{2,2} = 3(19) = 57$. Entonces nuestra regla de decisión es: no rechazar H_0 si $T^2 \leq 57$, de otra manera se rechaza H_0. Dado que nuestro valor de T^2 es 13.64 no rechazamos H_0 con un valor de $\alpha = 0{,}05$.

Inciso d)
Nos menciona la relación que existe entre la lambda de Wilks y T^2 la cual es la siguiente:

$$T^2 = \frac{n-1|\widehat{\Sigma}_0|}{|\widehat{\Sigma}|} - (n-1)$$

Buscando cada una de las matrices tenemos:

$$\widehat{\Sigma} = \begin{pmatrix} 8 & -10/3 \\ -10/3 & 2 \end{pmatrix} \quad \widehat{\Sigma}_0 = \begin{pmatrix} 28/3 & -2 \\ -2 & 10/3 \end{pmatrix}$$

Donde finalmente al evaluar tenemos lo siguiente:

$$T^2 = \frac{n-1|\widehat{\Sigma}_0|}{|\widehat{\Sigma}|} - (n-1) = \frac{3(\frac{244}{9})}{\frac{44}{9}} - 3 = \frac{150}{11}$$

Encontramos que es el mismo valor de T^2 que encontramos en incisos anteriores.

Inciso e)
Ahora nos piden evaluar Λ y la lambda de Wilks, entonces ocuparemos la siguiente relación:

$$\Lambda = \frac{|\widehat{\Sigma}|}{|\widehat{\Sigma}_0|}^{\frac{n}{2}} = \frac{44^2}{244^2} = 0{,}03251$$

Para nuestra lambda de Wilks, tenemos:

$$\Lambda^{\frac{1}{2}} = \frac{44}{244} = 0{,}18032$$

Problema 3

El departamento de control de calidad de un fabricante de hornos de microondas es requerido por el gobierno federal para monitorear la cantidad de radicación emitida por los hornos que fabrican. Se realizaron mediciones de la radiación emitida por 42 hornos seleccionados al azar con las puertas cerradas y abiertas.

a) Construte una elipse de confianza del 95 % para μ, considerando la transformación de las variables.

b) Prueba si $\mu^t = (0{,}562, 0{,}589)$ está en la región de confianza.

c) Calcula los valores y vectores propios de S y obten la gráfica del elipsoide de confianza.

d) Realiza una prueba para la hipótesis $H_0 : \mu^t = (0{,}55, 0{,}60)$ en un nivel de significancia de $\alpha = 0{,}5$.

Solución

Inciso a)
Al proporcionarnos el nuevo cambio de variable encontramos que nuestras matrices de $\bar{x}$ y S, son las siguientes:

$$\bar{x} = \begin{pmatrix} 0{,}564258 \\ 0{,}602981 \end{pmatrix} \quad S = \begin{pmatrix} 0{,}014350 & 0{,}011715 \\ 0{,}011715 & 0{,}014545 \end{pmatrix} \quad S^{-1} = \begin{pmatrix} 203{,}018 & -163{,}391 \\ -163{,}391 & 200{,}228 \end{pmatrix}$$

Una vez que tenemos nuestras matrices podemos escribir nuestra región que nos solicitan la cual estara dada por la siguiente expresión:

$$n(\bar{x} - \mu_0)^t S^{-1}(\bar{x} - \mu_0) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

Al sustituir tenemos:

$$42 \begin{pmatrix} 0{,}564 - \mu_1 & 0{,}603 - \mu_2 \end{pmatrix} \begin{pmatrix} 203{,}018 & -163{,}391 \\ -163{,}391 & 200{,}228 \end{pmatrix} \begin{pmatrix} 0{,}564 - \mu_1 \\ 0{,}603 - \mu_2 \end{pmatrix} \leq \frac{2(41)}{40} F_{2,40}(0{,}05)$$

Buscando el valor de $F_{2,40}(0{,}05) = 3{,}232$ tenemos finalmente:

$$42(203{,}018)(0{,}564 - \mu_1)^2 + 42(200{,}228)(0{,}603 - \mu_2)^2 - 84(163{,}391)(0{,}564 - \mu_1)(0{,}603 - \mu_2) \leq 6{,}6256$$

$$8526{,}756(0{,}564 - \mu_1)^2 + 8409{,}576(0{,}603 - \mu_2)^2 - 13724{,}84(0{,}564 - \mu_1)(0{,}603 - \mu_2) \leq 6{,}6256$$

Inciso b)
Como nos proporcionan $\mu^t = (0{,}562, 0{,}589)$ solo debemos sustituir en la

anterior expresión para poder saber si se encuentra en la región de confianza.

$$8526{,}756(0{,}564-0{,}562)^2+8409{,}576(0{,}603-0{,}589)^2-$$
$$13724{,}84(0{,}564-0{,}562)(0{,}603-0{,}589)=1{,}2980\leq 6{,}6256$$

Por lo que podemos observar que se encuentra en la región.

Inciso c)
Al buscar los valores y vectores propios de la matriz S, nos resulto: $\lambda_1 = 0{,}026163$ y $\lambda_2 = 0{,}0027319$ con vectores propios $e_1 = \begin{pmatrix} 0{,}70457 \\ 0{,}71004 \end{pmatrix}$ y $e_2 = \begin{pmatrix} -0{,}71004 \\ 0{,}70457 \end{pmatrix}$ respectivamente.

Los vectores propios y valores propios nos ayudaran a obtener nuestra gráfica. Sabemos que el centro se encuentra en $\bar{x}$ y el tamaño de los ejes los encontraremos con lo siguiente:

$$\sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}=\sqrt{0{,}026}\sqrt{\frac{2(41)}{42(40)}(3{,}232)}=0{,}64$$

luego para nuestro segundo valor propio

$$\sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}=\sqrt{0{,}0027}\sqrt{\frac{2(41)}{42(40)}(3{,}232)}=0{,}018$$

En la figura 0.1 se observa la elipse que nos pide el inciso.

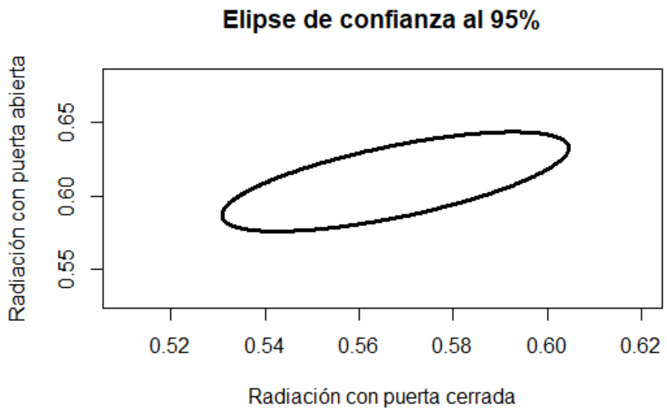

Figura 0.1: Elipse de confianza al 95 % con los datos.

Inciso d)
Para este inciso vamos a evaluar $\mu_a = (0{,}55, 0{,}60)^t$ con un nivel de significancia de $\alpha = 0{,}05$.

$$8526{,}756(0{,}564 - 0{,}55)^2 + 8409{,}576(0{,}603 - 0{,}60)^2 -$$
$$13724{,}84(0{,}564 - 0{,}55)(0{,}603 - 0{,}60) = 1{,}17048 \leq 6{,}6256$$

Se observa que se cumple lo anterior por lo que no rechazamos H_0.

Problema 4

Sabemos que T^2 es igual al t-valor cuadrado univariado más grande construido a partir de la combinación lineal $a^t x_j$, con $a = S^{-1}(\bar{x} - \mu_0)$.

a) Usando los resultados del ejercicio anterior y la misma hipotesis nula H_0 del inciso d), evalua a para los datos transformados de radiaciones de los hornos.

b) Verifica que el valor de t^2 calculado con esta a es igual a la T^2 del ejercicio anterior.

Solución

Inciso a)

Nos mencionan que debemos realizar una combinación lineal de la siguiente forma: $a^t x_j$, con $a = S^{-1}(\bar{x} - \mu_0)$. De lo anterior encontramos que valor de la matriz a es:

$$a = \begin{pmatrix} 203{,}018 & -163{,}391 \\ -163{,}391 & 200{,}228 \end{pmatrix} \begin{pmatrix} 0{,}564 - \mu_1 \\ 0{,}603 - \mu_2 \end{pmatrix}$$

Partiendo de nuestra hipotesis nula que nos proporciona el inciso d) del ejercicio 3, tenemos que $H_0 : \mu^t = (0{,}55, 0{,}60)$, por lo que solo debemos sustituir en los valores anteriores y ya tendremos nuestro vector a.

$$a = \begin{pmatrix} 2{,}352079 \\ -1{,}686634 \end{pmatrix}$$

Ahora debemos encontrar el valor de t^2, lo encontramos con lo siguiente:

$$t^2 = \frac{n(a^t(\bar{x} - \mu))^2}{a^t S a} \leq t^2_{n-1}(\alpha/2)$$

Al realizar dichas operaciones tenemos lo siguiente:

$$t^2 = \frac{42(\begin{pmatrix} 2{,}352079 & -1{,}6866 \end{pmatrix} \begin{pmatrix} 0{,}14 \\ 0{,}003 \end{pmatrix})^2}{\begin{pmatrix} 2{,}352079 & -1{,}6866 \end{pmatrix} \begin{pmatrix} 0{,}014350 & 0{,}011715 \\ 0{,}011715 & 0{,}014545 \end{pmatrix} \begin{pmatrix} 2{,}352079 \\ -1{,}6866 \end{pmatrix}} = 1{,}31$$

Inciso b)

Observamos que los valores en ambos casos son casi iguales, solo hay una ligera diferencia en algunos decimales y puede ser resultado de que tantos decimales se tomaron para el calculo de cada uno.

Problema 5

Los datos en el archivo representan las longitudes en centímetros de siete osos hembras a los 2,3,4 y 5 años de edad.

a) Obtener los intervalos de confianza simultaneos T^2 del 95 % para las cuatro medias poblacionales de la longitud por año.

b) Respecto al inciso a), obtener los intervalos de confianza simultaneos de T^2 del 95 % para los tres aumentos anuales sucesivos en la longitud media.

c) Obtener la elipse de confianza T^2 del 95 % para el aumento medio de la longitud de 2 a 3 años y el aumento medio de la longitud de 4 a 5 años.

d) Construir los intervalos de confianza de 95 % de Bonferroni para el conjunto formulado por las cuatro longitudes medias y los tres aumentos anuales sucesivos en la longitud media, compara los resultados con los obtenidos en a) y b).

Solución

Inciso a)

Para poder encontrar los 4 intervalos de confianza de cada media poblacional, nos vamos a apoyar de lo siguiente:

$$\left(a^t\bar{X} - \sqrt{\tfrac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)a^tSa}, a^t\bar{X} + \sqrt{\tfrac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)a^tSa}\right)$$

Donde vamos a ir variando la forma del vector a, notese que las elecciones sucesivas de $a^t = [100..,0], ..., a^t = [000...,1]$.

Para encontrar cada uno de los intervalos vamos a basarnos en:

$$\bar{x}_i - \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}}$$

Donde $i = 1, ..., 4$. Las matrices que vamos a ocupar son las siguientes:

$$\bar{x} = \begin{pmatrix} 143{,}28571 \\ 159{,}28571 \\ 173{,}14285 \\ 177{,}14285 \end{pmatrix} \quad S = \begin{pmatrix} 15{,}238095 & -8{,}928571 & -6{,}380952 & -4{,}547619 \\ -8{,}928571 & 99{,}904762 & -6{,}380952 & -54{,}214286 \\ -6{,}380952 & -6{,}380952 & 15{,}809524 & 1{,}476190 \\ -4{,}547619 & -54{,}214286 & 1{,}476190 & 45{,}47619 \end{pmatrix}$$

Y buscando el valor de $\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha) = \frac{24(9{,}117)}{3}$, con $\alpha = 0{,}05$, $p = 4$ y $n = 7$.
Lo que solo nos restaria es realizar las operaciones correpondientes en cada intervalo y proporcionarlo.
Cada intervalo de confianza del 95 % queda de la siguiente forma:

a) Para μ_1 es $(130{,}68522, 155{,}8861)$

b) Para μ_2 es $(127{,}0219, 191{,}5494)$

c) Para μ_3 es $(160{,}3082, 185{,}9774)$

d) Para μ_4 es $(155{,}375, 198{,}9106)$

Inciso b)
Para este inciso vamos a tener que cambiar nuestra $a_1 = [-1, 1, 0, 0], a_2 = [0, -1, 1, 0]$ y $a_3 = [0, 0 - 1, 1]$ ya que nos piden el intervalo de confianza para los tres aumentos anuales.
Buscamos nuevamente el valor de $\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha) = \frac{24(9{,}117)}{3}$, con $\alpha = 0{,}05$, $p = 4$ y $n = 7$. Una vez definido nuestros valores solo nos resta sustuir lo anterior en la siguiente expresión:

$$\left(a^t\bar{X} - \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)a^tSa}, a^t\bar{X} + \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)a^tSa}\right)$$

Cada intervalo de confianza del 95 % queda de la siguiente forma:

a) Para μ_1 es $(-21{,}24, 53{,}24)$

b) Para μ_2 es $(-22{,}70, 50{,}42)$

c) Para μ_3 es $(-20{,}59, 28{,}69)$

Inciso c)
Para encontrar la región de confianza de μ_{2-3} y μ_{4-5} vamos a ocupar lo siguiente:

$$\begin{pmatrix} 16-\mu_1 & 4-\mu_2 \end{pmatrix} \begin{pmatrix} 0{,}011024 & 0{,}009386 \\ 0{,}009386 & 0{,}025135 \end{pmatrix} \begin{pmatrix} 16-\mu_1 \\ 4-\mu_2 \end{pmatrix} \leq \frac{72{,}96}{7} = 10{,}42$$

Así para encontrar nuestros ejes de la elipse que nos propocionen el 95 %, confianza tenemos:

- Eje Mayor $= \sqrt{157{,}8}\sqrt{72{,}96} \begin{pmatrix} 0{,}895 \\ -0{,}447 \end{pmatrix}$
- Eje Menor $= \sqrt{33{,}53}\sqrt{72{,}96} \begin{pmatrix} 0{,}447 \\ 0{,}895 \end{pmatrix}$

Donde finalmente tenemos la siguiente elipse, se ilustra en la figura 0.2:

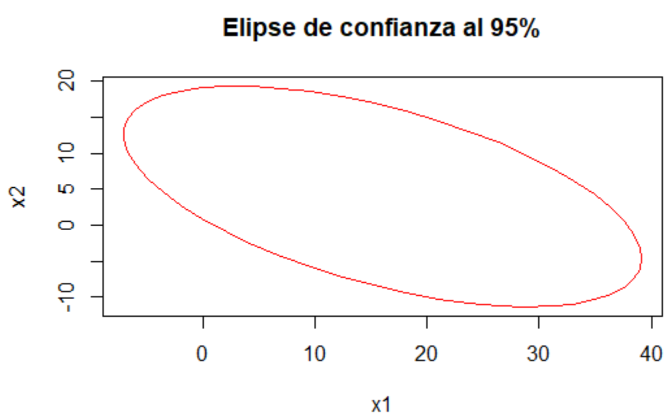

Figura 0.2: Elipse de confianza al 95 % con los datos.

Inciso d)
En este inciso nos piden encontrar los intervalos de confianza del 95 % de Bonferroni, en este caso vamos a ocupar lo siguiente:

$$\bar{x}_i - t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}}$$

Donde para cada caso vamos a ir variando nuestras matrices que ocupemos, vamos a encontrar los intervalos para el inciso a), donde $n = 7$ y $p = 4$. Para este primer conjunto de datos encontramos que $t_{n-1}(\alpha/2p) = t_6(1/160) \approx 3{,}7074$. Una vez que ya tenemos los datos procedemos a sustituir en la expresión que mostramos. Cada intervalo de confianza del 95 % queda de la siguiente forma:

a) Para μ_1 es $(137{,}81573, 148{,}75569)$

b) Para μ_2 es $(145{,}27973, 173{,}29169)$

c) Para μ_3 es $(167{,}571251, 178{,}71444)$

d) Para μ_4 es $(167{,}69327, 186{,}592429)$

Para el segundo conjunto de intervalos tenemos:

a) Para μ_1 es $(-1{,}43, 33{,}43)$

b) Para μ_2 es $(-3{,}25, 20{,}97)$

c) Para μ_3 es $(-7{,}55, 15{,}55)$

Al analizar ambos conjuntos de intervalos con los anteriores en muchos de ellos se logra ver que cambio el intervalo, en algunos casos se ve una diferencia pequeña mientras que en otros si hay mucha diferencia. Buscando una posible razón de que sean distintos puede que los intervalos de confianza ocupando T^2 nos intervalos demasiado anchos debido a que solo se aplican a las medias de cada año.