

Inferencia Vector de Medias

DIEGO ALEJANDRO PORTELA CASSAB.^a, LIZ S. POLANÍA GIRALDO^b, YENNY MABEL ROJAS M.^c



Punto 1

- a. A partir del conjunto de datos College, que muestra las siguientes tres variables: X_1 : 'Ciencias sociales e historia', X_2 : 'Competencias orales' y X_3 : 'Ciencias' donde se reportan los puntajes de cada una de estas tres materias para una muestra de $n = 87$. Se procede a probar el siguiente sistema de hipótesis, con $\alpha = 0.05$:

$$\begin{cases} H_0 & : \mu = (500, 50, 30) \\ H_a & : \mu \neq (500, 50, 30) \end{cases}$$

En primer lugar se calculan el vector de medias \bar{X} y la matriz de varianzas y covarianzas muestrales S y se presentan a continuación:

$$\bar{X}^t = (526.59 \quad 54.69 \quad 25.13)$$

$$S = \begin{pmatrix} 5808.06 & 597.83 & 222.03 \\ 597.83 & 126.05 & 23.39 \\ 222.03 & 23.39 & 23.11 \end{pmatrix}$$

Antes de proceder a probar la hipótesis se realiza una prueba de normalidad multivariada con el fin de establecer la teoría con la cual se va a trabajar.

Con la ayuda del software R se realiza la prueba de normalidad multivariada de Shapiro-Wilk, el cual con un $p_{valor} = 0.06082$, que a un nivel de significancia del 10 % permite concluir que los datos con los que estamos trabajando provienen de una distribución normal multivariada; a partir de este resultado se puede hacer una prueba de hipótesis para la inferencia del vector de medias basados en normalidad.

Se calcula el estadístico de prueba T^2 :

$$T^2 = n(\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0) = 223.31$$

Y el valor crítico viene dado por:

$$\frac{(n-1)p}{n-p} F_{p, n-p, \alpha} = \frac{(87-1)3}{87-3} F_{3, 84, 0.05} = 10.28544$$

Ahora, al comparar el valor del estadístico de prueba y el valor crítico, se observa que $T^2 = 223.31 < 10.28544$, por lo cual, se rechaza H_0 a un nivel de significancia del 5 %, en otras palabras, a partir de la información que otorga la muestra se puede concluir que a un nivel de significancia del 5 % la media poblacional es diferente al vector $\mu = (500, 50, 30)$.

^aCódigo: 25461226. E-mail: daportelac@unal.edu.co

^bCódigo: 2514949. E-mail: lspolaniag@unal.edu.co

^cCódigo: 25141114. E-mail: ymrojas@unal.edu.co

- b. Determine la magnitud y dirección de los ejes del elipsoide de confianza al 95 %.

En primer lugar, se calcula la matriz inversa de S :

$$S^{-1} = \begin{pmatrix} 0.00043 & -0.00157 & -0.00256 \\ -0.00157 & 0.01550 & -0.00057 \\ -0.00256 & -0.00057 & 0.06840 \end{pmatrix}$$

Los Valores y Vectores propios de S vienen dados por:

$$\begin{aligned} \lambda_1 &= 5878.7916 & e_1^T &= (0.99390539, 0.10344339, 0.03809906) \\ \lambda_2 &= 63.83510 & e_2^T &= (0.1037315, -0.9945892, -0.00566024) \\ \lambda_3 &= 14.59806 & e_3^T &= (-0.0373074, -0.0095778, 0.99925794) \end{aligned}$$

La magnitud y dirección de los ejes del elipsoide vienen dados por:

$$\begin{aligned} \sqrt{\lambda_1} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p,\alpha}} &= 26.3631 \\ \sqrt{\lambda_2} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p,\alpha}} &= 2.7471 \\ \sqrt{\lambda_3} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p,\alpha}} &= 1.3137 \end{aligned}$$

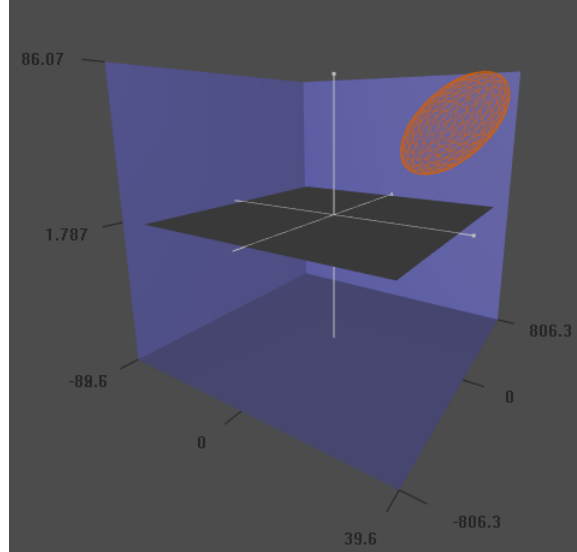
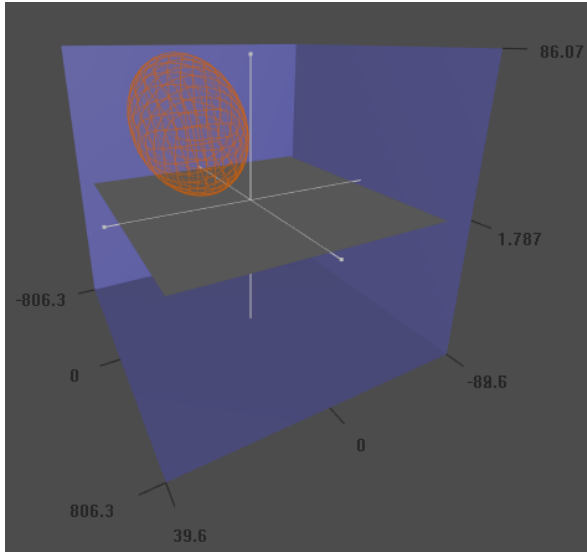


TABLA 1: Elipsoide formado por las 3 variables visto desde dos perspectiva diferentes

- c. Construya los intervalos de confianza al 95 %. Explique los resultados.

Se calculan los intervalos de confianza simultáneos al 95 % para μ_1, μ_2 y μ_3 . Se tiene entonces:

$$\frac{p(n-1)}{n-p} F_{p,n-p,\alpha} = \frac{3(86)}{87-3} F_{3,84,0.05} = 10.28$$

Ahora, se tienen los intervalos:

$$\begin{aligned}
526.59 - \sqrt{10.28} \sqrt{\frac{5808.06}{87}} &\leq \mu_1 \leq 526.59 + \sqrt{10.28} \sqrt{\frac{5808.06}{87}} \\
500.38 &\leq \mu_1 \leq 552.79 \\
54.69 - \sqrt{10.28} \sqrt{\frac{126.05}{87}} &\leq \mu_2 \leq 54.69 + \sqrt{10.28} \sqrt{\frac{126.05}{87}} \\
50.83 &\leq \mu_2 \leq 58.55 \\
25.13 - \sqrt{10.28} \sqrt{\frac{23.11}{87}} &\leq \mu_3 \leq 25.13 + \sqrt{10.28} \sqrt{\frac{23.11}{87}} \\
23.47 &\leq \mu_3 \leq 26.78
\end{aligned}$$

Component-Wise

Los intervalos vienen dados de la forma:

$$\begin{aligned}
\bar{x}_1 - t_{n-1, \alpha/2} \sqrt{\frac{S_{11}}{n}} &\leq \mu_1 \leq \bar{x}_1 + t_{n-1, \alpha/2} \sqrt{\frac{S_{11}}{n}} \\
&\vdots \\
\bar{x}_p - t_{n-1, \alpha/2} \sqrt{\frac{S_{pp}}{n}} &\leq \mu_p \leq \bar{x}_p + t_{n-1, \alpha/2} \sqrt{\frac{S_{pp}}{n}}
\end{aligned}$$

En este caso, se tiene entonces que los intervalos se contruyeron y se obtuvo el siguiente resultado:

$$\begin{aligned}
526.59 - 1.99 \sqrt{\frac{5808.06}{87}} &\leq \mu_1 \leq 525.29 + 1.99 \sqrt{\frac{5808.06}{87}} \\
510.34 &\leq \mu_1 \leq 542.8289 \\
54.69 - 1.99 \sqrt{\frac{126.05}{87}} &\leq \mu_2 \leq 54.69 + 1.99 \sqrt{\frac{126.05}{87}} \\
52.30 &\leq \mu_2 \leq 57.08 \\
25.13 - 1.99 \sqrt{\frac{23.11}{87}} &\leq \mu_3 \leq 25.13 + 1.99 \sqrt{\frac{23.11}{87}} \\
24.10 &\leq \mu_3 \leq 26.15
\end{aligned}$$

Ahora

$$P(\text{Todos los t-intervalos contengan los } \mu'_i s) = \frac{t_{n-1, \alpha/2}}{\sqrt{\frac{p(n-1)}{n-p} F_{p, n-p, \alpha}}} = \frac{1.98}{\sqrt{\frac{3(86)}{84} 2.71}} = 0.4521$$

Mediante el método que se acaba de ilustrar, la probabilidad que todos los parámetros se encuentren en los intervalos calculados es de 45.21 %

BonFerroni

Los intervalos vienen dados de la forma:

$$\begin{aligned}
\bar{x}_1 - t_{n-1, \alpha/2p} \sqrt{\frac{S_{11}}{n}} &\leq \mu_1 \leq \bar{x}_1 + t_{n-1, \alpha/2p} \sqrt{\frac{S_{11}}{n}} \\
&\vdots \\
\bar{x}_p - t_{n-1, \alpha/2p} \sqrt{\frac{S_{pp}}{n}} &\leq \mu_p \leq \bar{x}_p + t_{n-1, \alpha/2p} \sqrt{\frac{S_{pp}}{n}}
\end{aligned}$$

En este caso, se tiene entonces que los intervalos se contruyeron y se obtuvo el siguiente resultado:

$$\begin{aligned}
 526.59 - 2.44\sqrt{\frac{5808.06}{87}} &\leq \mu_1 \leq 525.29 + 2.44\sqrt{\frac{5808.06}{87}} \\
 506.64 &\leq \mu_1 \leq 546.54 \\
 54.69 - 2.44\sqrt{\frac{126.05}{87}} &\leq \mu_2 \leq 54.69 + 2.44\sqrt{\frac{126.05}{87}} \\
 51.75 &\leq \mu_2 \leq 57.63 \\
 25.13 - 2.44\sqrt{\frac{23.11}{87}} &\leq \mu_3 \leq 25.13 + 2.44\sqrt{\frac{23.11}{87}} \\
 23.87 &\leq \mu_3 \leq 26.38
 \end{aligned}$$

Se observa que los intervalos calculados por el método de Component-Wise tienen un nivel de confianza relativamente pequeño, además de tener una longitud de intervalo mayor comparado con los intervalos de Bonferroni.

Punto 2

Los investigadores interesados en evaluar la función pulmonar en poblaciones no patológicas pidieron a los sujetos que corrieran en una cinta de correr hasta el agotamiento. Se recogieron muestras de aire a intervalos definidos y se analizaron los contenidos de gas. Los resultados se resumen en 4 medidas. El consumo de oxígeno para 25 hombres y 25 mujeres, las variables fueron:

- X_1 Volumen de reposo O_2 L/min : Variable 1
- X_2 Volumen de reposo O_2 $ml/Kg/min$: Variable 2
- X_3 Volumen de máximo O_2 L/min : Variable 3
- X_4 Volumen de máximo O_2 $ml/Kg/min$: Variable 4

- a Encuentre las diferencias de género al probar la igualdad de las medias del grupo. Utilice $\alpha = 0.05$. Si se rechaza $H_0 : \mu_1 - \mu_2 = 0$, encuentre la combinación lineal más conveniente. Verificar los supuestos. Suponga que las observaciones son independientes.

Se realiza la prueba de normalidad multivariada de Shapiro-Wilk, y se concluye con un $p\text{-valor} = 9.137 \times 10^{-7}$, con un nivel de significancia del 5 %, que los datos no provienen de una distribución normal multivariada.

Se procede a hacer la prueba M-Box para probar que las matrices de varianza-covarianza de las poblaciones son iguales. Tenemos que para un $p\text{-valor} = 8.637 \times 10^{-9}$ y un nivel de confianza del 95 %, decimos que las poblaciones tienen matriz de varianza-covarianza diferentes. Sin embargo para el ejercicio teórico asumimos este supuesto como verdadero.

Dado a que no se cumplen el supuesto de normalidad multivariada, se procede a utilizar el método bootstrap para la construcción de intervalos de confianza para evaluar la hipótesis de igualdad de medias. En primer lugar se considera la transformación propuesta por Garthwaite y Kock, dada por:

$$X \rightarrow W = A(X - \mu)$$

Donde \mathbf{X} representa una matriz 4×1 dada por la diferencia de medias entre géneros, es decir, entre los dos grupos de estudio. \mathbf{A} es una matriz construida de tal forma que $\sum_{i=1}^4 \text{corr}(X_i, W_i)$ se maximice bajo la condición de que $W^T W = (X - \mu)^T \Sigma^{-1} (X - \mu)$ se cumpla para todas las X , luego,

$$A = (G \Sigma G)^{-1/2} G$$

Tal que \mathbf{G} es una matriz diagonal, definida positiva, y en su diagonal tiene los elementos diagonales iguales al recíproco de la raíz cuadrada del elemento diagonal correspondiente de Σ , es decir, $1/\sqrt{\sigma_{ii}}$.

Cabe aclarar que para este caso se usaran la estimación de Σ , dada por $S = (n_1 - 1)S_1 + (n_2 - 1)S_2/(n - 2)$, tal que $n_1 = 25$, $n_2 = 25$, $n = n_1 + n_2$ corresponde la tamaño de la población de hombres y al tamaño de la población d mujeres respectivamente, y S_1 , S_2 , las matrices de covarianzas muestrales de cada una de las poblaciones respectivamente .

Así, se tiene que para los datos estudiados la transformación es:

$$X \rightarrow \hat{W} = \begin{pmatrix} 1.230 \\ -0.864 \\ 1.856 \\ 1.415 \end{pmatrix}$$

Y se puede verificar que:

$$\begin{aligned} W^T W &= 7.7098 \\ (X - \mu)^T \Sigma^{-1} (X - \mu) &= 7.7098 \end{aligned}$$

Se examina el método bootstrap para la formación de intervalos de confianza para $\theta = W_i^2 / \sum_{j=1}^m W_j^2$, esta expresión define la contribución porcentual a la distancia de Mahalanobis por cada una de las variables de estudio, y el vector obtenido se definirá como el vector $\hat{\gamma}$. El vector encontrado es:

$$\hat{\gamma} = \begin{pmatrix} 0.196 \\ 0.096 \\ 0.447 \\ 0.259 \end{pmatrix}$$

Con base en la transformación anterior y por medio del método planteado por Zillur R. Shabuz & Paul H. Garthwaite, se realizan 3 pasos para encontrar el intervalo de confianza:

1. Se realizan 50.000 iteraciones por le método de bootstrap con reemplazamiento sobre la matriz de diferencia de medias de los grupos, y así se encuentran tantas estimaciones de los $\hat{\gamma}_i$ como iteraciones se hayan hecho. De aquí, los histogramas para cada $\hat{\gamma}_i^*$, están dados por:

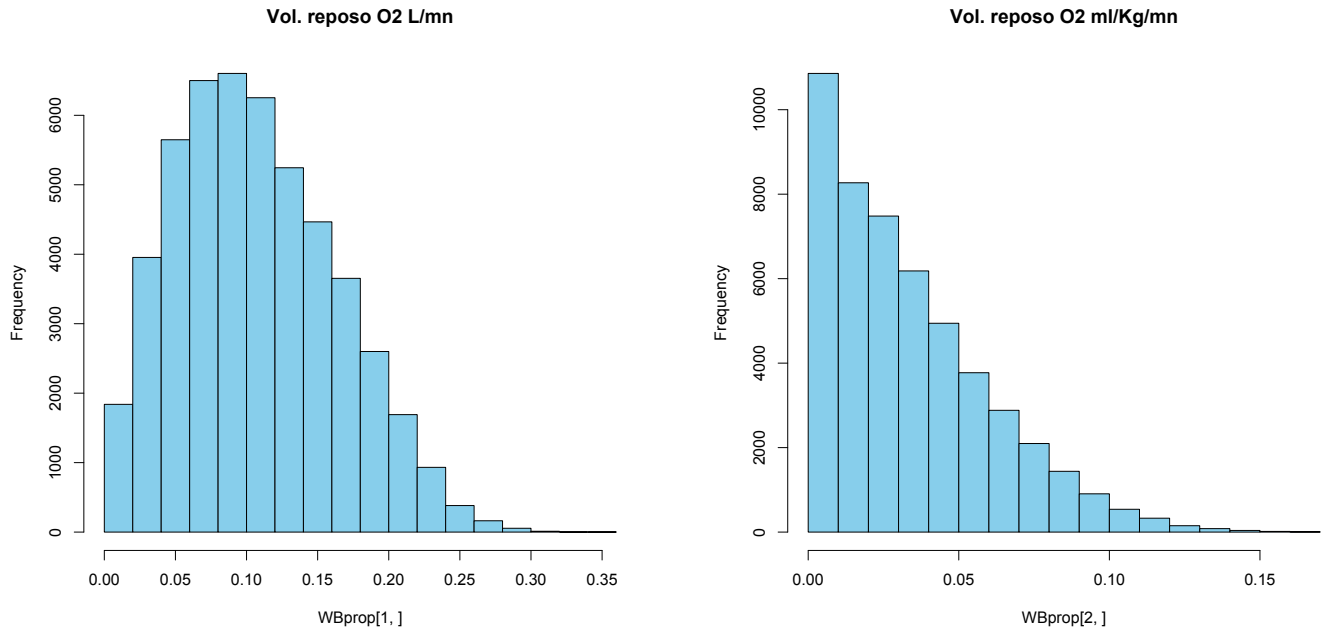


TABLA 2

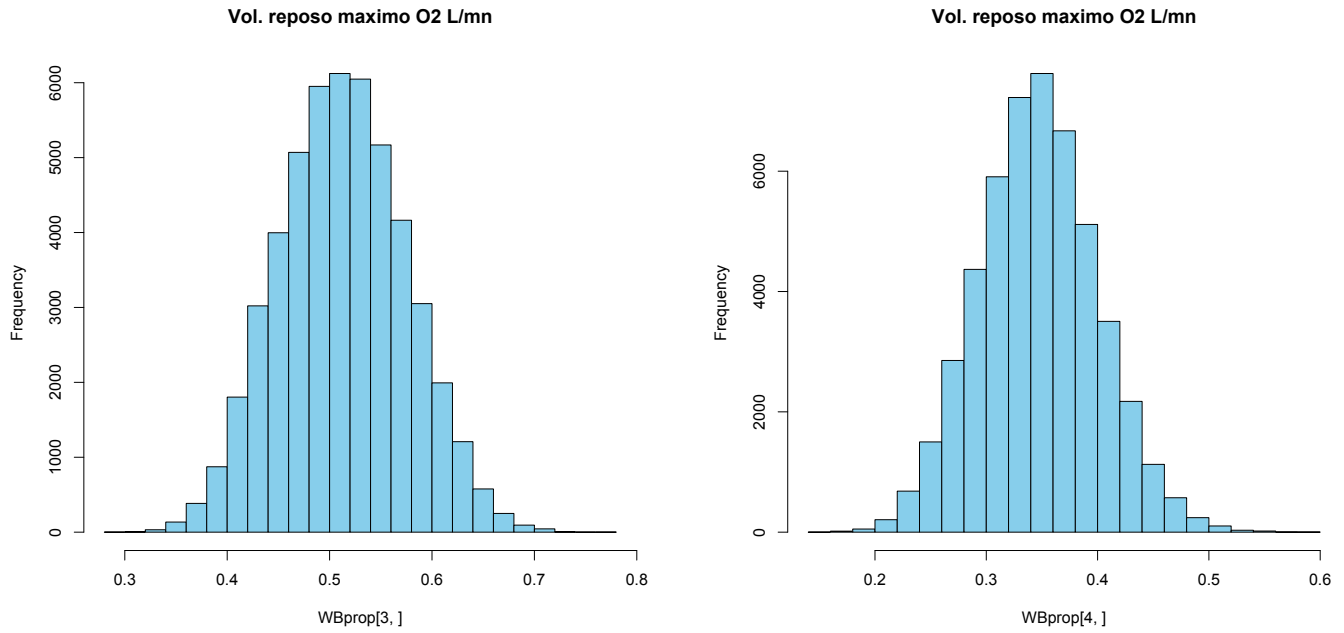


TABLA 3

2. Seguidamente, se realiza una transformación dada por: $\hat{v}_k^* = \frac{(2\hat{\gamma} - \gamma_k^*)^2}{\sum_{i=1}^4 (2\hat{\gamma} - \gamma_k^*)^2}$, con $k = 1, \dots, 50.000$.
3. Para el intervalo de confianza $100(1 - \alpha)\%$ para θ es dado por $(\hat{v}^*(\alpha_1), \hat{v}^*(1 - \alpha_2))$ donde $\hat{v}^*(q)$ es el q -ésimo cuantil muestral de la muestra $\hat{v}_1^*, \dots, \hat{v}_N^*$

Los intervalos de confianza para cada una de las variables están dados por los siguientes quantiles:

	α_1	α_2
1	(0.00	0.95)
2	(0.01	0.95)
3	(0.01	0.96)
4	(0.01	0.96)
5	(0.02	0.97)
6	(0.03	0.97)
7	(0.03	0.98)
8	(0.04	0.98)
9	(0.04	0.99)
10	(0.04	0.99)
11	(0.05	1.00)

Los intervalos de confianza encontrados para cada una de las variables son:

1	(3.189791x10 ⁻¹¹ , 0.3820623)
2	(2.134845x10 ⁻⁵ , 0.3900520)
3	(8.057511x10 ⁻⁵ , 0.3984405)
4	(1.753545x10 ⁻⁴ , 0.4075627)
5	(3.144051x10 ⁻⁴ , 0.4175662)
6	(4.810720x10 ⁻⁴ , 0.4288487)
7	(7.105732x10 ⁻⁴ , 0.4414222)
8	(9.796042x10 ⁻⁴ , 0.4574010)
9	(1.251663x10 ⁻³ , 0.4756311)
10	(1.554935x10 ⁻³ , 0.4999899)
11	(1.907899x10 ⁻³ , 0.5545118)

TABLA 4: Intervalos de confianza-Variable 1

1	(2.506588x10 ⁻¹² , 0.1089509)
2	(3.125529x10 ⁻⁶ , 0.1110925)
3	(1.573880x10 ⁻⁵ , 0.1133257)
4	(3.714593x10 ⁻⁵ , 0.1160501)
5	(6.513927x10 ⁻⁵ , 0.1190924)
6	(9.986842x10 ⁻⁵ , 0.1222690)
7	(1.429668x10 ⁻⁴ , 0.1253408)
8	(1.964539x10 ⁻⁴ , 0.1290730)
9	(2.626208x10 ⁻⁴ , 0.1328890)
10	(3.384400x10 ⁻⁴ , 0.1384013)
11	(4.156397x10 ⁻⁴ , 0.1473770)

TABLA 5: Intervalos de confianza-Variable 2

1	(6.972280x10 ⁻⁸ , 0.9089506)
2	(2.006137x10 ⁻² , 0.9131998)
3	(3.807182x10 ⁻² , 0.9180306)
4	(5.368373x10 ⁻² , 0.9225100)
5	(6.767717x10 ⁻² , 0.9272813)
6	(8.088395x10 ⁻² , 0.9333955)
7	(9.325959x10 ⁻² , 0.9391318)
8	(1.040202x10 ⁻¹ , 0.9461682)
9	(1.153271x10 ⁻¹ , 0.9531513)
10	(1.262270x10 ⁻¹ , 0.9621139)
11	(1.374160x10 ⁻¹ , 0.9903093)

TABLA 6: Intervalos de confianza-Variable 3

1	(9.540542x10 ⁻¹⁴ , 0.6135771)
2	(3.344381x10 ⁻⁵ , 0.6226117)
3	(1.302515x10 ⁻⁴ , 0.6316682)
4	(2.709149x10 ⁻⁴ , 0.6408670)
5	(4.826665x10 ⁻⁴ , 0.6512244)
6	(7.222829x10 ⁻⁴ , 0.6635766)
7	(1.009665x10 ⁻³ , 0.6759216)
8	(1.372642x10 ⁻³ , 0.6905225)
9	(1.772285x10 ⁻³ , 0.7072490)
10	(2.253803x10 ⁻³ , 0.7283409)
11	(2.794059x10 ⁻³ , 0.7778317)

TABLA 7: Intervalos de confianza-Variable 4

Luego, el intervalo de confianza más corto para cada variable es:

	Intervalo de	confianza	α_1	α_2
$Variable_1$	[0.00000 ,	0.3820623)	0.00	0.95
$Variable_2$	[0.00000 ,	0.1089509)	0.00	0.95
$Variable_3$	(1.262270x10 ⁻¹ ,	0.9621139)	0.04	0.99
$Variable_4$	[0.00000 ,	0.7778317)	0.00	0.95

TABLA 8: Intervalos más cortos

Así, mediante una corrección en la que se incluye el 0 sobre los intervalos cuando $\alpha_1 = 0.00, \alpha_2 = 0.95$, se tiene que para la variable 3, Volumen de máximo O_2 L/min , se rechaza la hipótesis nula de diferencia de medias entre grupos. Así mismo, para las variables, 1, 2 y 4 no se rechaza la hipótesis nula de igualdad de medias entre grupos.

- b Construya los intervalos de confianza simultáneos del 95 % para cada $\mu_{1i} - \mu_{2i} = 0, i = 1, 2, 3, 4$. Compare con los otros tipos de intervalos.

Como un ejercicio académico, se desarrollan los intervalos de confianza de bonferroni para cada diferencia e intervalo univariado para cada diferencia. Sin embargo, aclaramos que el supuesto de normalidad no se tiene y por el tamaño de muestra no es pertinente aplicar resultados asintóticos para ajustes distribucionales, con lo cual estos intervalos carecen de sentido en un caso de la vida real.

1. Intervalos Univariados

$$\mu_{1i} - \mu_{2i} : (\bar{x}_{1i} - \bar{x}_{2i}) \pm \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p, n_1+n_2-p-1}(\alpha) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{iipooled}}$$

Los valores del intervalo están dados en la siguiente tabla:

	<i>Variable</i> ₁	<i>Variable</i> ₂	<i>Variable</i> ₃	<i>Variable</i> ₄
inferior	0.02	-0.78	1.02	7.11
superior	0.14	1.08	1.73	15.42

Se observa en este caso que la variable 2 al poseer el cero dentro del intervalo nos permite rechazar H_0 , caso contrario en los demás intervalos, con lo que se tiene un resultado similar al obtenido por el método bootstrap.

2. Intervalos de Bonferroni

Los intervalos confidenciales simultaneos $100(1 - \alpha) \%$ para las diferencias de la población p-ésima está dado por:

$$\mu_{1i} - \mu_{2i} : (\bar{x}_{1i} - \bar{x}_{2i}) \pm t_{n_1+n_2-2}\left(\frac{\alpha}{2p}\right) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{iipooled}}$$

Los valores del intervalo están dados en la siguiente tabla:

	<i>Variable</i> ₁	<i>Variable</i> ₂	<i>Variable</i> ₃	<i>Variable</i> ₄
inferior	-0.15	-3.48	-0.02	-5.00
superior	0.32	3.79	2.77	27.53

TABLA 9: Intervalos de Bonferroni

A partir de los intervalos de bonferroni se tiene que H_0 se rechaza en todas las variables, con lo cual diríamos que no es tan acertados los intervalos de bonferroni para este caso.

Anexos

```

rm(list=ls())
library(mvnormtest)
library(boot)
library(resample)
library(graphics)
library(car)
library(sandwich)
library(effects)
library(clusterSim)
library(Matrix)
library(copula)
library(expm)
setwd("//Users//jonathanmedina//Documents//Liz//Multivariado")

# First Item ####
mu<-c(500,50,30)
alpha=0.05
n<-length(data[,1])
p<-length(data[1,])-1
Xbar<-colMeans(data[, -1])
S<-cov(data[, -1])
eigenvv<-eigen(S)

mshapiro.test(t(as.matrix(data[, -1])))
T2<-n*t(Xbar-mu)%*%solve(S)%*%(Xbar-mu)
T2quant<-((n-1)*p/(n-p))*qf(1-alpha,p,n-p)
T2>T2quant

lambda1 <-eigenvv$values[1]
lambda2 <-eigenvv$values[2]
lambda3 <-eigenvv$values[3]

vp1<- eigenvv$vectors[,1]
vp2<- eigenvv$vectors[,2]
vp3<- eigenvv$vectors[,3]

eje1 <- sqrt(lambda1)*sqrt(p*(n-1)/(n*(n-p))*qf(1-alpha,p,n-p))
eje2 <- sqrt(lambda2)*sqrt(p*(n-1)/(n*(n-p))*qf(1-alpha,p,n-p))
eje3 <- sqrt(lambda3)*sqrt(p*(n-1)/(n*(n-p))*qf(1-alpha,p,n-p))

pnt <- ((p*(n-1))/(n-p))*qf(1-alpha,p,n-p)
mu1_inf <- Xbar[1]-sqrt(pnt)*sqrt(S[1,1]/n)
mu1_sup <- Xbar[1]+sqrt(pnt)*sqrt(S[1,1]/n)

mu2_inf <- Xbar[2]-sqrt(pnt)*sqrt(S[2,2]/n)
mu2_sup <- Xbar[2]+sqrt(pnt)*sqrt(S[2,2]/n)

mu3_inf <- Xbar[3]-sqrt(pnt)*sqrt(S[3,3]/n)
mu3_sup <- Xbar[3]+sqrt(pnt)*sqrt(S[3,3]/n)

x<- data$X1
y<- data$X2
z<- data$X3
r1 <- sqrt((n-1)^2 * qbeta(0.997, 1, (n-3)/2) / n)

```

```

el1 <- ellipse(Xbar[-3], S[-3,-3], eje1 , draw = FALSE)
plot(el1, type='l', col="red",
      xlab="X1", ylab="X2",
      main = "Elipse de confianza variables X1 y X2")

el2 <- ellipse(Xbar[-1], S[-1,-1], eje2, draw = FALSE)
plot(el2, type='l', col="red",
      xlab="X2", ylab="X3",
      main = "Elipse de confianza variables X2 y X3")

el3 <- ellipse(Xbar[-2], S[-2,-2], eje3, draw = FALSE)
plot(el3, type='l', col="red",
      xlab="X1", ylab="X3",
      main = "Elipse de confianza variables X1 y X3")

rgl.points(x, y, z, color = "lightgray")
rgl.points(iris[,-1])

rgl_init <- function(new.device = FALSE, bg = "white", width = 640) {
  if( new.device | rgl.cur() == 0 ) {
    rgl.open()
    par3d(windowRect = 50 + c( 0, 0, width, width ) )
    rgl.bg(color = bg )
  }
  rgl.clear(type = c("shapes", "bboxdeco"))
  rgl.viewpoint(theta = 15, phi = 20, zoom = 0.7)
}

rgl_init()
rgl.spheres(x, y, z, r = 0.2, color = "yellow") # Scatter plot
rgl.bbox(color = "#333377") # Add bounding box decoration

rgl_add_axes <- function(x, y, z, axis.col = "grey",
                        xlab = "", ylab="", zlab="", show.plane = TRUE,
                        show.bbox = FALSE, bbox.col = c("#333377", "black"))
{
  lim <- function(x){c(-max(abs(x)), max(abs(x))) * 1.1}
  # Add axes
  xlim <- lim(x); ylim <- lim(y); zlim <- lim(z)
  rgl.lines(xlim, c(0, 0), c(0, 0), color = axis.col)
  rgl.lines(c(0, 0), ylim, c(0, 0), color = axis.col)
  rgl.lines(c(0, 0), c(0, 0), zlim, color = axis.col)

  # Add a point at the end of each axes to specify the direction
  axes <- rbind(c(xlim[2], 0, 0), c(0, ylim[2], 0),
               c(0, 0, zlim[2]))
  rgl.points(axes, color = axis.col, size = 3)

  # Add axis labels
  rgl.texts(axes, text = c(xlab, ylab, zlab), color = axis.col,
            adj = c(0.5, -0.8), size = 2)

```

```

# Add plane
if(show.plane)
  xlim <- xlim/1.1; zlim <- zlim /1.1
  rgl.quads( x = rep(xlim, each = 2), y = c(0, 0, 0, 0),
            z = c(zlim[1], zlim[2], zlim[2], zlim[1]))

# Add bounding box decoration
if(show.bbox){
  rgl.bbox(color=c(bbox.col[1],bbox.col[2]), alpha = 0.5,
           emission=bbox.col[1], specular=bbox.col[1], shininess=5,
           xlen = 3, ylen = 3, zlen = 3)
}
}

rgl.spheres(x, y, z, r = 0.2, color = "#D95F02")
rgl_add_axes(x, y, z, show.bbox = TRUE)
ellips <- ellipse3d(cov(cbind(x,y,z)),
                   centre=c(mean(x), mean(y), mean(z)), level = 0.95)
shade3d(ellips, col = "#D95F02", alpha = 0.05, lit = FALSE)
wire3d(ellips, col = "#D95F02", lit = FALSE)
aspect3d(1,1,1)

#95% Confidence Ellipse
library(ellipse)
plot(data)
radio<-sqrt(2*qt(0.95,n-1,n-p))
radio<-sqrt(qchisq(0.95,2))

confelip<-ellipse(Xbar[-3],S[-3,-3],radio)
plot(confelip,type = "l")

library(car)
library(jocre)remove(list = ls())
setwd("C:/Users/dport/Desktop/DiegoAlejandro/2018-2/1) Multivariate Analyse/2 Corte/derWorkshop3")
data<-read.table(file = "college.txt",header = T)
data1<-read.table(file = "OxCoDat.txt",header = T)

#Simultaneous confidence intervals
li1<-Xbar[1]-qt(1-alpha/2,n-1)*sqrt(S[1,1]/n)
ls1<-Xbar[1]+qt(1-alpha/2,n-1)*sqrt(S[1,1]/n)

li2<-Xbar[2]-qt(1-alpha/2,n-1)*sqrt(S[2,2]/n)
ls2<-Xbar[2]+qt(1-alpha/2,n-1)*sqrt(S[2,2]/n)

li3<-Xbar[3]-qt(1-alpha/2,n-1)*sqrt(S[3,3]/n)
ls3<-Xbar[3]+qt(1-alpha/2,n-1)*sqrt(S[3,3]/n)

prob <-(1-alpha)^3
(sqrt((3*86/84)*2.713227)-qt(1-alpha/2,n-1))/qt(1-alpha/2,n-1)

```

```

#Bonferroni

lib1<-Xbar[1]-qt(1-alpha/(2*p),n-1)*sqrt(S[1,1]/n)
lsb1<-Xbar[1]+qt(1-alpha/(2*p),n-1)*sqrt(S[1,1]/n)

lib2<-Xbar[2]-qt(1-alpha/(2*p),n-1)*sqrt(S[2,2]/n)
lsb2<-Xbar[2]+qt(1-alpha/(2*p),n-1)*sqrt(S[2,2]/n)

lib3<-Xbar[3]-qt(1-alpha/(2*p),n-1)*sqrt(S[3,3]/n)
lsb3<-Xbar[3]+qt(1-alpha/(2*p),n-1)*sqrt(S[3,3]/n)

#-----
#           Second Point
#-----
#           Item a)
#-----
data1<-read.table("Tabla62.txt", sep="", header=F)
set.seed(1234)

# Test Assumption ####
mshapiro.test(t(as.matrix(data1[,-1])))
boxM(data1[,-1],grouping = data1$V1)

male<-data1[data1$V1==1,]
male<-male[,c(2,3,4,5)]
female<-data1[data1$V1==2,]
female<-female[,c(2,3,4,5)]
men<-male
women<-female
dif<-men-women
inm<-indepTestSim(25, 4, N = 1000, verbose = interactive())
inmm<-indepTest(diff, inm, alpha=0.05)

Xbar1<-colMeans(men)
Xbar2<-colMeans(women)
S1<-cov(men)
eigen(S1) ; solve(S1) #S1 is a positive definite matrix.
S2<-cov(women)
eigen(S2) ;solve(S2) #S2 is a positive definite matrix.
S<-(S1+S2)/2
D12<-t(Xbar1-Xbar2)%*%solve(S)%*%(Xbar1-Xbar2)
G<-diag(1/sqrt(diag(S)))
A<- solve(sqrtm(G%*%S%*%G))%*%G
Wg<-A%*%(Xbar1-Xbar2) # Gamma ? estimado # Wgorro #
propWg<-Wg*100/sum(Wg)
propoWg<-Wg/sum(Wg)

(t(Wg))%*%Wg

# Method A
# Step One

samp_size=25
inter = 50000
samp1<-matrix(sample(1:nrow(diff), samp_size*inter, replace=T),ncol=inter, byrow=T)

```

```

samp2<-matrix(sample(1:nrow(diff), samp_size*inter, replace=T),ncol=inter, byrow=T)
WB<-matrix(0,4,inter)
for(i in 1:inter){ # B hace referencia a Bootstrap.
  B1<-men[samp1[,i],]
  B2<-women[samp2[,i],]
  X1BAR<-colMeans(B1)
  X2BAR<-colMeans(B2)
  S1B<-cov(B1)
  S2B<-cov(B2)
  SB<-(S1B+S2B)/2
  GB<-diag(1/sqrt(diag(SB)))
  AB<-solve(sqrtm(GB%*%SB%*%GB))%*%GB
  WB[,i]<-AB%*%(X1BAR-X2BAR)
}
WBabs<-WB^2
WBprop<-matrix(0,4,inter)
for (n in 1:inter) {
  WBprop[,n]<-t(WBabs[,n]/sum(WBabs[,n]))
}
vkge<-matrix(0,4,inter)
for (k in 1:inter) {
  vkge[,k]<-(2*Wprop-WBprop[,k])^2/sum((2*Wprop-WBprop[,k])^2) # V_kgorro estrella
}
hist(WBprop[1,], main= "Vol. reposo 02 L/mn", col="skyblue")
hist(WBprop[2,], main="Vol. reposo 02 ml/Kg/mn", col="pink")
hist(WBprop[3,], main= "Vol. reposo maximo 02 L/mn", col="skyblue")
hist(WBprop[4,], main= "Vol. reposo maximo 02 L/mn", col="pink")

#-----
#      intervalos de confianza
#-----

a<-seq(0,0.050,0.005)
b<-seq(0.050,0,-0.005)
alphas<-cbind(a,(1-b))
CI1<-matrix(0,11,2)
CI2<-matrix(0,11,2)
CI3<-matrix(0,11,2)
CI4<-matrix(0,11,2)
for (p in 1:11) {
  CI1[p,<-quantile(vkge[1,],alphas[p,])
}
for (p in 1:11) {
  CI2[p,<-quantile(vkge[2,],alphas[p,])
}
for (p in 1:11) {
  CI3[p,<-quantile(vkge[3,],alphas[p,])
}
for (p in 1:11) {
  CI4[p,<-quantile(vkge[4,],alphas[p,])
}

IS1<-order(CI1[,2]-CI1[,1])
IS2<-order(CI2[,2]-CI2[,1])
IS3<-order(CI3[,2]-CI3[,1])

```

```

IS4<-order(CI4[,2]-CI4[,1])

#-----
#                               Item b)
#-----
# Item b ####
alphab=0.05
pb=4

# Confidence Region ####
T2b<-t(Xbar1-Xbar2)%*%solve((2/50)*S)%*%(Xbar1-Xbar2)
ccuad<-(48*4/45)*qf(alphab,pb,45,lower.tail = F)
T2b>ccuad # We reject H_0.

# Bonferroni CI ####
eins<-c((Xbar1[1]-Xbar2[1])-qt(alphab/(2*pb),
48,lower.tail = F)*sqrt(S[1,1]),(Xbar1[1]-Xbar2[1])+qt(alphab/(2*pb),48,lower.tail = F)*sqrt(S[1,1]))
zwei<-c((Xbar1[2]-Xbar2[2])-qt(alphab/(2*pb),
48,lower.tail = F)*sqrt(S[2,2]),(Xbar1[2]-Xbar2[2])+qt(alphab/(2*pb),48,lower.tail = F)*sqrt(S[2,2]))
drei<-c((Xbar1[3]-Xbar2[3])-qt(alphab/(2*pb),
48,lower.tail = F)*sqrt(S[3,3]),(Xbar1[3]-Xbar2[3])+qt(alphab/(2*pb),48,lower.tail = F)*sqrt(S[3,3]))
vier<-c((Xbar1[4]-Xbar2[4])-qt(alphab/(2*pb),
48,lower.tail = F)*sqrt(S[4,4]),(Xbar1[4]-Xbar2[4])+qt(alphab/(2*pb),48,lower.tail = F)*sqrt(S[4,4]))
BonfCI<-cbind(eins,zwei,drei,vier)

#Univariate CI ####
ein<-c((Xbar1[1]-Xbar2[1])-sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[1,1]),
(Xbar1[1]-Xbar2[1])+sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[1,1]))
zwe<-c((Xbar1[2]-Xbar2[2])-sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[2,2]),
(Xbar1[2]-Xbar2[2])+sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[2,2]))
dre<-c((Xbar1[3]-Xbar2[3])-sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[3,3]),
(Xbar1[3]-Xbar2[3])+sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[3,3]))
vie<-c((Xbar1[4]-Xbar2[4])-sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[4,4]),
(Xbar1[4]-Xbar2[4])+sqrt((48*4/45)*qf(alphab,pb,45,lower.tail = F)*(2/50)*S[4,4]))
UniCI<-cbind(ein,zwe,dre,vie)

```