UNIVERSIDAD AUTÓNOMA CHAPINGO

DIVISIÓN DE CIENCIAS FORESTALES

Departamento de Estadística Matemática y Cómputo

Métodos Multivariados

Examen

Alumna:
Arely Yazmin Vásquez González

7°5

Profesor:
Dr. Filemón Ramírez Pérez

Enero 2022

7.25 Algunos médicos recetan amitriptilina como antidepresivo. Sin embargo, también hay efectos secundarios conjeturados que parecen estar relacionados con el uso de la droga: irregularidades, latidos del corazón, presiones sanguíneas anormales y ondas irregulares en el electrocardiograma, entre otras cosas. Los datos recopilados sobre 17 pacientes que ingresaron en el hospital después de una sobredosis de amitiptilina se dan en la tabla 7.6. Dos variables de respuesta son
Y_1 = Nivel plasmático total de TCAD (TOT) Y_2 = Cantidad de amitiptilina presente en el nivel plasmático de TCAD (AMI) Los cinco predictores son
Z_1 =Sexo: 1 si es mujer, 0 si es hombre (GEN)
Z_2 =Cantidad de antidepresivos tomados en el momento de la sobredosis (AMT)
Z_3 = PR medida de onda (PR)
Z_4 = Presión arterial diastólica (DIAP)
Z_5 = medición de onda QRS (QRS)

Tabla 7.6. Datos de amitiptilina

y_1 TOT	y_2 AMI	z_1 GEN	z_2 AMT	z_3 PR	z_4 DIAP	z_5 QRS
3389	3149	1	7500	220	0	140
1101	653	1	1975	200	0	100
1131	810	0	3600	205	60	111
596	448	1	675	160	60	120
896	844	1	750	185	70	83
1767	1450	1	2500	180	60	80
807	493	1	350	154	80	98
1111	941	0	1500	200	70	93
645	547	1	375	137	60	105
628	392	1	1050	167	60	74
1360	1283	1	3000	180	60	80
652	458	1	450	160	64	60
860	722	1	1750	135	90	79
500	384	0	2000	160	60	80
781	501	0	4500	180	0	100
1070	405	0	1500	170	90	120
1754	1520	1	3000	180	0	129

(c) Realice un análisis de regresión multivariable utilizando las respuestas Y_1 y Y_2.

(i) Sugerir y ajustar modelos de regresión lineal apropiados.
Respuesta:
Ajustando un modelo completo con todas las covariables, se obtuvieron las siguientes estimaciones de los parámetros ,

Beta	
-2879.478	-2728.709
675.65078	763.02976
0.2848511	0.3063734
10.272133	8.8961977
7.2511714	7.2055597
7.5982397	4.9870508

Entonces el modelo ajustado es:
$\hat{y_1} = -2879.478 + 675.65z_1 + 0.285z_2 + 10.272z_3 + 7.251z_4 + 7.598z_5$
$\hat{y_2} = -2728.709 + 763.02976z_1 + 0.306z_2 + 8.896z_3 + 7.206z_4 + 4.987z_5$
Las sumas de cuadrados resultantes son:

SCTOTAL		SCREG		SCE	
7705940.2	7474767.2	6835931.9	6709090.7	870008.31	765676.48
7474767.2	7610377.9	6709090.7	6669669	765676.48	940708.9

Haciendo la prueba de hipótesis, de que las respuestas no dependen de las covariables, se obtuvo que el estadístico Lambda de Wilks es 0.0837 con p-value de 0.00122, por lo tanto, con un nivel de significancia del 0.05 rechazamos la hipotesis nula de que las covariables no dependen de las respuestas.

Lambda	Fo	pvalue1
0.0837239	4.9120246	0.0012226

Ahora haciendo una hipótesis acerca de que las covariables z_3, z_4, z_5 no son significativas en el modelo, se obtuvo que para esta prueba el estadístico Lambda de Wilks es 0.44 con p-value de 0.17, por lo tanto, con un nivel de significancia del 0.05, no rechamos la hipotésis nula de que z_3, z_4 y z_5 no son significativas.

Lambda	Fo	pvalue1
0.4405021	1.6889907	0.1755326

Entonces podemos ajustar un modelo más simple quitando estas covariables y usando solo z_1 y z_2.
El modelo ajustado usando solo z_1 y z_2 como covariables es:
$\hat{y_1} = 55.72 + 507.073z_1 + 0.329z_2$
$\hat{y_2} = -241.3479 + 606.31z_1 + 0.324z_2$

Beta	
56.720053	-241.3479
507.07308	606.30967
0.3289618	0.3242549

SCTOTAL		SCREG		SCE	
7705940.2	7474767.2	5905583.9	5928573	1800356.4	1546194.2
7474767.2	7610377.9	5928573	5989720.5	1546194.2	1620657.3

Lambda	Fo	pvalue1
0.1900647	8.4094852	0.0001713

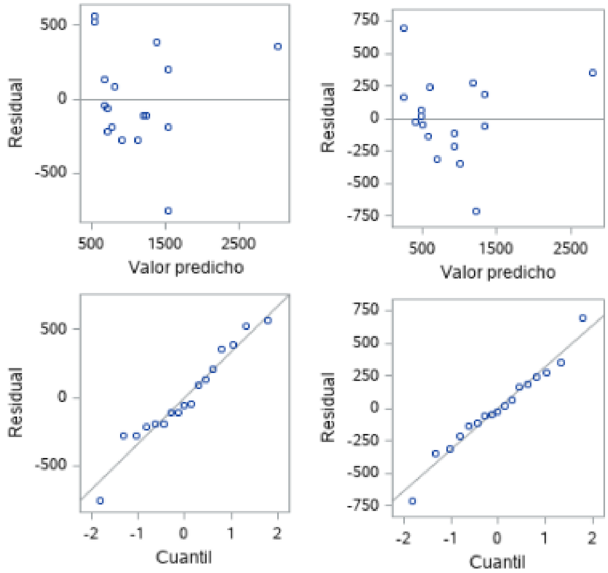
El p-value es 0.00017, por lo tanto con un nivel de significancia del 0.05 concluimos z_1 y z_2 si son significativos en nuestro modelo ajustado.

(ii) Analice los residuos.

Respuesta:

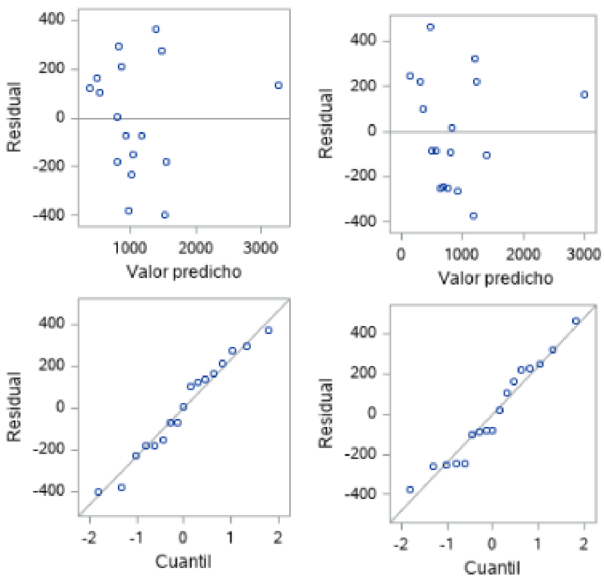
A continuación se presentan la gráfica de los residuales con el modelo ajustado usando solo las covariables z_1 y z_2.

Grafica de los residuales para Y_1 Grafica de los residuales para Y_2



Podemos observar que tanto para Y_1 como para Y_2 en la gráfica de los residuales contra los valores predichos más o menos se distribuyen aleatoriamente y en la gráfica Q-Q podemos ver que los residuales caen más o menos en una linea recta. Las siguientes gráficas de los residuales corresponden al modelo completo usando todas las covariables.

Gráfica de los residuales para Y_1 Grafica de los residuales para Y_2



8.1. Determinar los componentes principales de la población Y_1 y Y_2 para la matriz de covarianza $\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$
Además, calcule la proporción de la varianza de la población total explicada por el primer componente principal.
Respuesta:
Primero obtenemos los eigenvalores y eigenvectores de Σ

$$\lambda_1 = 6, \quad e_1' = [0.894, 0.447]$$
$$\lambda_2 = 1, \quad e_2' = [0.447, -0.894]$$

El i-ésimo componente principal esta dado por:
$Y_i = \mathbf{e_i'X}$, entonces
$Y_1 = e_1'\mathbf{X} = 0.894X_1 + 0.447X_2$
$Y_2 = e_2'\mathbf{X} = 0.447X_1 - 0.894X_2$
La proporción de la varianza de la población total explicada por el primer componente principal Y_1 es,
$\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{6}{6+1} = 0.857$

8.2. Convierta la matriz de covarianza del Ejercicio 8.1 en una matriz de correlación $\boldsymbol{\rho}$.
Respuesta:
Usando el siguiente resultado obtenemos la matriz de correlación
$Cov(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\mathbf{\Sigma}(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho}$
donde $\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} = \begin{bmatrix} 2.236 & 0 \\ 0 & 1.414 \end{bmatrix}$
entonces, $\boldsymbol{\rho} = \begin{bmatrix} 1.0 & 0.6325 \\ 0.6325 & 1.0 \end{bmatrix}$

(a) Determine los componentes principales Y_1 y Y_2 de $\boldsymbol{\rho}$ y calcule la proporción de la varianza de la población total explicada por Y_1.
Respuesta:
Los eigenvalores y eigenvectores de $\boldsymbol{\rho}$ son,

$$\lambda_1 = 1.632, \; e_1' = [0.707, 0.707]$$
$$\lambda_2 = 0.368, \; e_2' = [0.707, -0.707]$$

El i-ésimo componente principal esta dada por
$Y_i = \mathbf{e_i'Z} = \mathbf{e_i'(V^{1/2})^{-1}}(\mathbf{x} - \boldsymbol{\mu})$, entonces
$Y_1 = 0.707Z_1 + 0.707Z_2 = 0.707(\frac{X_1-\mu_1}{2.236}) + 0.707(\frac{X_2-\mu_2}{1.414}) =$
$0.316(X_1 - \mu_1) + 0.5(X_2 - \mu_2)$
$Y_2 = 0.707Z_1 - 0.707Z_2 = 0.707(\frac{X_1-\mu_1}{2.236}) - 0.707(\frac{X_2-\mu_2}{1.414}) =$
$0.316(X_1 - \mu_1) - 0.5(X_2 - \mu_2)$
La proporcion de la varianza de la poblacional estandarizada explicada por Y_i es $\frac{\lambda_k}{p}$, entonces la población total explicada por Y_1 es:

$$\frac{1.632}{2} = 0.816$$

(b) Compare los componentes calculados en la Parte a con los obtenidos en el ejercicio 8.1. ¿Son iguales? ¿Deberían serlo?
Respuesta:
No, no son iguales y no deberian de serlo. Ya que cuando se estandarizan X_1 y X_2, las variables resultantes contribuyen igualmente a los componentes principales determinados a partir de $\boldsymbol{\rho}$, en cambio en 8.1 las variables no contribuyen igualmente.

(c) Calcular las correlaciones $\rho_{Y_1,Z_1}, \rho_{Y_1,Z_2}$ y ρ_{Y_2,Z_1}.
Respuesta:
$\rho_{Y_i,Z_k} = e_{ik}\sqrt{\lambda_i}$
$\rho_{Y_1,Z_1} = (0.707)\sqrt{1.632} = 0.903$
$\rho_{Y_1,Z_2} = (0.707)\sqrt{1.632} = 0.903$
$\rho_{Y_2,Z_1} = (0.707)\sqrt{0.368} = 0.429$

8.6 Los datos sobre x_1=ventas y x_2=beneficios de las 10 empresas más grandes del mundo se incluyeron en el ejercicio 1.4 del capítulo 1.
Del ejemplo 4.12
$\overline{\mathbf{x}} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix}, \; S = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$

(a) Determine los componentes principales muestrales y sus varianzas para estos datos. (Es posible que necesite la fórmula cuadrática para resolver los valores propios de S).
Respuesta:
Los eigevalores y eigenvectores de S son:

$$\hat{\lambda_1} = 7488.8, \;\; \hat{e}_1' = [0.999, 0.041]$$
$$\hat{\lambda_2} = 13.837, \;\; \hat{e_2}' = [-0.041, 0.999]$$

La iésima componente principal muestral esta dado por

$$\hat{y}_i = \hat{\mathbf{e}}_i^t \mathbf{x}$$

$\Rightarrow \hat{y}_1 = 0.999X_1 + 0.041X_2$
$\hat{y}_2 = -0.041X_1 + 0.999X_2$
La varianza muestral esta dada por
$varianza\ muestral(\hat{y}_k) = \hat{\lambda}_k$, entonces
$varianza\ muestral(\hat{y}_1) = \hat{\lambda}_1$=7488.8
$varianza\ muestral(\hat{y}_2) = \hat{\lambda}_2$=13.837

(b) Encuentre la proporción de la varianza muestral total explicada por $\hat{y}_1$.
Respuesta:
La proporción de varianza muestral total explicada por $\hat{y}_1$ esta dada por

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_1+\hat{\lambda}_2} = \frac{7488.8}{7488.8+13.837} = 0.998$$

10.1 Considere la matriz de covarianza dada en el ejemplo 10.3:

$$\mathrm{Cov}\left(\begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \cdots \\ X_1^{(2)} \\ X_2^{(2)} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \cdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 100 & 0 & \vdots & 0 & 0 \\ 0 & 1 & \vdots & .95 & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ 0 & .95 & \vdots & 1 & 0 \\ 0 & 0 & \vdots & 0 & 100 \end{bmatrix}$$

Verifique que el primer par de variables canónicas sea $U_1 = X_2^{(1)}, \; V_1 = X_1^{(2)}$ con correlación canónica $\rho_1^* = .95$.
Respuesta:
$\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2} = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}^{-1/2} \begin{bmatrix} 0 & 0 \\ 0.95 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0.95 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}^{-1/2} =$

$\begin{bmatrix} 0 & 0 \\ 0 & 0.9025 \end{bmatrix}$

Los eigenvalores ρ_1^{*2}, ρ_2^{*2} y eigenvectores $\mathbf{e_1}, \mathbf{e_2}$ asociados a la matriz obtenida son:

$\rho_1^{*2} = 0.9025 \Rightarrow \ \rho_1^* = \sqrt{0.9025} = 0.95 \qquad \mathbf{e_1'} = [0, 1]$

$\rho_2^{*2} = 0.0 \quad \mathbf{e_2'} = [1, 0]$

Se tiene que los vectores $\mathbf{a}$ y $\mathbf{b}$ en la c.l U y V que permiten obtener la k-ésima variable canonica son:

$\mathbf{a} = \mathbf{e}_k^t \Sigma_{11}^{-1/2}$ y $\mathbf{b} = \mathbf{f}_k^t \Sigma_{22}^{-1/2}$

Entonces, $\mathbf{a}^t = [0 \ \ 1] \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}^{-1/2} = [0 \ \ 1]$, $\quad \mathbf{b}^t = [1 \ \ 0] \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}^{-1/2} = [1 \ \ 0]$

Entonces el primer par de variables canónicas es

$U_1 = \mathbf{a}^t \mathbf{X^{(1)}} = [0 \ 1] \mathbf{X^{(1)}} = X_2^{(1)}$

$V_1 = \mathbf{a}^t \mathbf{X^{(2)}} = [1 \ 0] \mathbf{X^{(2)}} = X_1^{(2)}$

CODIGO SAS PARA PROBLEMA 7.25,

```
data uno;
input y1 y2 z1 z2 z3 z4 z5;
cards;
3389 3149 1 7500 220  0 140
1101  653 1 1975 200  0 100
1131  810 0 3600 205 60 111
 596  448 1  675 160 60 120
 896  844 1  750 185 70  83
1767 1450 1 2500 180 60  80
 807  493 1  350 154 80  98
1111  941 0 1500 200 70  93
 645  547 1  375 137 60 105
 628  392 1 1050 167 60  74
1360 1283 1 3000 180 60  80
 652  458 1  450 160 64  60
 860  722 1 1750 135 90  79
 500  384 0 2000 160 60  80
 781  501 0 4500 180  0 100
1070  405 0 1500 170 90 120
1754 1520 1 3000 180  0 129
;
proc IML;
use uno;
start multregre;
read all VAR {y1 y2} into Y;
read all VAR {z1 z2 z3 z4 z5 } into X;
/*read all VAR {z1 z2 } into X; */ para hacer ajustar modelo con solo z1 y z2
N=nrow(X);
m=ncol(Y);
M=J(N,1,1);
X=M||X;
K=ncol(X);
q=K-1;
```

```
XX=X'*X;
IXX=inv(XX);
Beta=IXX*X'*Y;
PX=X*IXX*X';
PX1=M*inv(M'*M)*M';
/*PX1=X[,1:3]*inv(X[,1:3]'*X[,1:3])*X[,1:3]';*/ para hacer prueba
acerca de z3,z4,z5
QM=I(N)-PX;
fc=Y'*(1/N)*J(N,N,1)*Y;
SCREG=(Y'*PX*Y)-fc;
SCE=Y'*QM*Y;
SCTOTAL=Y'*Y-fc;
R2=SCREG/SCTOTAL;
print Beta;
print SCTOTAL SCREG SCE;
S1=(1/N)*Y'*(I(N)-PX1)*Y;
S=(1/N)*Y'*(I(N)-PX)*Y;
DS1=det(S1);
DS=det(S);
Lambda=DS/DS1;
vh=5
/*vh=3;*/
/*vh=2;*/
vE=N-q-1;
constgl=(vE-1)/vh;
Fo=(1-sqrt(Lambda))*constgl/sqrt(Lambda);
dfn=2*vh;
dfd=2*(vE-1);
pvalue1=1-probF(Fo,dfn,dfd);
razon=log(DS/DS1);
EMG=-(N-vh-1-0.5*(-2))*razon;
pvalue2=1-probChi(EMG,dfn);
print S1, S;
print Lambda Fo pvalue1;
print EMG pvalue2;
finish;
run multregre;
quit;
/*Para obtener los graficos en ii)*/
data uno;
input y1 y2 x1 x2 x3 x4 x5;
cards;
3389 3149 1 7500 220  0 140
1101  653 1 1975 200  0 100
1131  810 0 3600 205 60 111
 596  448 1  675 160 60 120
 896  844 1  750 185 70  83
1767 1450 1 2500 180 60  80
 807  493 1  350 154 80  98
```

```
1111   941  0  1500  200  70   93
 645   547  1   375  137  60  105
 628   392  1  1050  167  60   74
1360  1283  1  3000  180  60   80
 652   458  1   450  160  64   60
 860   722  1  1750  135  90   79
 500   384  0  2000  160  60   80
 781   501  0  4500  180   0  100
1070   405  0  1500  170  90  120
1754  1520  1  3000  180   0  129
;
proc reg;
model y1 y2=x1 x2 x3 x4 x5;
mtest  x1,x2,x3,x4,x5/print;
```

CÓDIGO R

```
###8.1
sigma<-matrix(c(5,2,2,2),ncol=2)
eigenv<-eigen(sigma)
###8.2
v<-diag(c(sqrt(5),sqrt(2)),2,2)
invv<-solve(v)
rho<-invv%*%sigma%*%invv
eigenv2<-eigen(rho)
###8.6
S<-matrix(c(7476.45,303.62,303.62,26.19),ncol=2)
eigenvv<-eigen(S)
##10.1
library("expm")
S11<-matrix(c(100,0,0,1),ncol=2)
S12<-matrix(c(0,0.95,0,0),ncol=2)
S21<-matrix(c(0,0,0.95,0),ncol=2)
S22<-matrix(c(1,0,0,100),ncol=2)
p<-(solve(sqrtm(S11)))%*%S12%*%(solve(S22))%*%S21%*%(solve(sqrtm(S11)))
eig<-eigen(p)
a<-eig$vector[,1]%*%(solve(sqrtm(S11)))
b<-eig$vector[,2]%*%(solve(sqrtm(S22)))
```