DETERMINACIÓN DE MATRICES DE CORRELACIÓN TRANSITIVAS

ALVARO DE JESUS VILLOTA VIVEROS

UNIVERSIDAD NACIONAL DE COLOMBIA

DEPARTAMENTO DE ESTADISTICA

SAN JUAN DE PASTO

OCTUBRE 2008

DETERMINACIÓN DE MATRICES DE CORRELACIÓN TRANSITIVAS

Presentado por:

ALVARO DE JESUS VILLOTA VIVEROS

Presentado a:

DEPARTAMENTO DE ESTADISTICA

Trabajo de Grado presentado como requisito para optar el título de Especialista en Estadística

Asesor:

JORGE HUMBERTO MAYORGA ALVAREZ

UNIVERSIDAD NACIONAL DE COLOMBIA

DEPARTAMENTO DE ESTADISTICA

SAN JUAN DE PASTO

OCTUBRE 2008

CONTENIDO

Pag.

INTRODUCCION

La matriz de correlación ha sido objeto de numerosos estudios, sin embargo, hasta el momento no son conocidos trabajos que aborden la transitividad entre los signos de los coeficientes de correlación.

Con este aporte se logra determinar las matrices que cumplen con esta propiedad, fundamentándose exclusivamente en el conteo de signos negativos en la matriz. Se establecen unas leyes de signos presentes en estas matrices.

Consecuencia importante de esa precisión es la separación de las variables en dos grupos, opuestos entre si, las variables del primer conjunto aumentan a la vez, mientras las variables del segundo hacen lo mismo pero en sentido contrario.

Se han definido nuevos términos estadísticos “Matriz de Correlación transitiva” y “Matriz de Confrontación.

En la parte final del estudio, se presenta un método para determinar Submatrices de Correlación Transitiva a partir de una matriz primaria que no lo es.

La bibliografía que se relaciona en la conclusión del trabajo, es un soporte de conceptos básicos de estadística como varianza, covarianza, matriz de correlación y coeficientes de correlación, mas no de la idea y desarrollo de la investigación teórica que son inéditos.

1. MARCO TEORICO

El estudio supone el conocimiento de la definición, significado y rango del coeficiente de correlación. También de la aplicación de la propiedad transitiva en múltiples variables.

1.1 DEFINICION DE COEFICIENTE DE CORRELACIÓN

Se define coeficiente de correlación (ρ) al cociente entre la covarianza entre X e Y y el producto entre las desviaciones estándar de esas variables.

$\rho = \sigma xy / (\sigma x \,.\, \sigma y)$

1.2 SIGNIFICADO DEL COEFICIENTE DE CORRELACION

El coeficiente de correlación es la medida del grado en que dos variables aleatorias X e Y se encuentran linealmente relacionadas. Si los puntos P(x,y) se encuentran distribuidos muy cerca a una recta cuya pendiente es mayor que cero, el coeficiente de correlación se acerca al valor de 1. Si los puntos se distribuyen cerca a una recta con pendiente menor que cero el coeficiente de correlación se aproxima a -1. Si los puntos se distribuyen alejados de dichas rectas, el valor absoluto de la correlación se aproxima a cero.

1.3 RANGO DEL COEFICIENTE DE CORRELACION

La desigualdad de Schawarz permite demostrar que el coeficiente de correlación tiene valores comprendidos entre 1 y menos 1 inclusives. $-1 \leq \rho \leq 1$

1.4 TRANSITIVIDAD

Considérese 3 variables x1, x2 y x3

Si x2 aumenta cuando x1 lo hace y si x3 aumenta cuando x2 crece, la matemática deterministica concluye que x3 aumenta cuando x1 se incrementa.

Las correlaciones en estadística, son positivas para indicar que las dos variables crecen o decrecen a la vez y son negativas cuando una de las variables aumenta mientras la otra disminuye.

De acuerdo con lo dicho, era de esperarse que los signos en la matriz de correlaciones cumplieran con la ley de transitividad. Por ejemplo si el signo de la correlación entre x1 y x2 es positiva y el signo de la correlación entre x2 y x3 también lo es, una afirmación a priorí sería la siguiente: El signo de la correlación entre x1 y x3 debe ser positiva

1.5 TRANSITIVIDAD EN SIGNOS PARA FUNCIONES EN GENERAL

En Matemáticas determinística, la transitividad entre variables, se puede explicar así:

Si $X2$ crece cuando $X1$ lo hace, estaríamos frente a una función creciente y $dX2/dX1 > 0$, Si $X3$ crece cuando $X2$ lo hace, $dX3/dx2 > 0$.

Si realizamos el producto $(dX3/dX2)(dX2/dX1) > 0$, $dX3/dX1 > 0$

$X3$ crece cuando $X1$ lo hace.

Si $X2$ decrece cuando $X1$ aumenta, $dX2/dX1 < 0$, y además $X3$ disminuye cuando $X2$ aumenta,

$dX3/dX2 < 0$, Funciones decrecientes.

Realizando el producto: $(dX3/dX2)(dX2/dX1) > 0$, $dX3/dX1 > 0$

$X3$ crece cuando $X1$ lo hace. Además $x1$ y $x3$ crecen mientras $x2$ disminuye.

Realizando reflexiones similares, se pueden establecer todas las posibilidades y concluir que la transitividad en los signos en la matemática deterministica se cumple realmente.

1.6 TRANSITIVIDAD EN SIGNOS PARA FUNCIONES LINEALES

En el caso particular de funciones lineales. Si $X2$ aumenta cuando $X1$ lo hace y $X3$ aumenta cuando $X2$ lo realiza.

Se puede afirmar que:

$X2 = aX1 + b$ donde $a > 0$

$X3 = cx2 + c$ donde $c > 0$

Reemplazando X2 de la primera ecuación en la segunda:

$$X3 = c(aX1 + b) + c$$

$$X3 = acX1 + bc + c$$

Nótese que $ac > 0$ por lo tanto es verdad que X3 crece cuando X1 lo hace.

Si X2 disminuye mientras X1 aumenta, y X3 disminuye cuando X2 aumenta, $a<0$ y $c<0$ y el producto $ac>0$ significa que X3 crece cuando X1 lo hace y que tanto X1 como X3 aumentan mientras X2 disminuye

1.7 TRANSITIVIDAD EN SIGNOS PARA MATRICES DE CORRELACIÓN

Los signos de las matrices de correlación no guardan en general transitividad, esto debido a que los puntos de dos variables se encuentran distribuidos en mayor o menor grado de dispersión con respecto a la recta guardando correlaciones débiles. Justamente el presente estudio permite detectar aquellas que si cumplen la propiedad.

2. METODOLOGIA

2.1 DEFINICIONES

2.1.1 MATRIZ DE CORRELACION TRANSITIVA: Es aquella matriz de correlación en cuyos coeficientes se cumple en su totalidad la transitividad en sus signos

2.1.2 MATRIZ DE CORRELACION NO TRANSITIVA: Es aquella matriz de correlación en cuyos coeficientes no se cumple en su totalidad la transitividad en sus signos.

2.2 DETERMINACIÓN DE LAS LEYES DE LOS SIGNOS EN UNA MATRIZ DE CORRELACIÓN TRANSITIVA

Este trabajo demuestra que la Matriz de Correlación Transitiva, posee unas leyes con respecto a sus signos:

Supóngase que se están considerando n variables y todas ellas aumentan (o disminuyen) a la vez, si la matriz de correlación es transitiva, esperaremos n^2 elementos con signos positivos y cero elementos con signos negativos.

Ahora asumamos que n-1 variables aumentan (o disminuyen) al tiempo, mientras 1 variable se comporta en sentido contrario. Si la matriz de correlación es transitiva, podemos realizar las siguientes reflexiones:

Cada una de las n-1 variables puede combinarse con las n-2 restantes formando $(n-1)(n-2)/2$ parejas relacionadas positivamente, pero como en la matriz de correlación las parejas se repiten, tendremos $(n-1)(n-2)$ signos positivos, son también positivas las correlaciones de las n variables consigo mismas, encontraremos $(n-1)(n-2) + n$ signos positivos en la matriz.

Si desarrollamos la expresión obtenemos $n^2 - 2n + 2$, que se pude presentar como $\mathbf{(n-1)^2 + 1^2}$.

Respecto a los signos negativos ellos serán n^2 menos el número de signos positivos, es decir: $n^2 - (n^2 - 2n + 2) =$ $\mathbf{2(1)(n-1)}$ signos negativos.

Ahora si n-2 variables aumentan (o disminuyen) a la vez, mientras 2 variables se comportan en sentido contrario, cada una de las n-2 variables hará duplas con las n-3 restantes. Las parejas serán $(n-2)(n-3)/2$. Como las parejas se repiten y debe considerase la diagonal de la matriz que también es positiva, tendremos la siguiente cantidad de signos positivos: $(n-2)(n-3) + n$. Pero debe considerarse que las dos variables que se comportaban en sentido contrario, también forman una pareja con correlación positiva que aparece 2 veces en la matriz. De esta manera, el total de signos positivos en la matriz es: $(n-2)(n-3) + n + 2$, que desarrollando y reagrupando adecuadamente se puede expresar como: $\mathbf{(n-2)^2 + 2^2}$ signos positivos.

Los signos negativos, serán $n^2 -((n-2)(n-3) +n +2)$, es decir $\mathbf{2(2)(n-2)}$ signos negativos.

Finalmente y con el propósito de establecer una regla general, supóngase que n-3 variables aumentan (o disminuyen) a la vez, mientras 3 variables se comportan en sentido contrario.

Los signos positivos por cuenta de las n-3 variables es (n-3)(n-4), por concepto de la diagonal debe sumarse n y por las duplas de las otras 3 variables se tendrán otros 6 signos. (n-3)(n-4) + n +6, que desarrollando y reagrupando resulta: **$(n-3)^2 + 3^2$** signos positivos.

Los signos negativos para este caso se calculan así: $n^2 - ((n-3)(n-4) + n + 6) =$ **$2(3)(n-3)$**

Signos negativos.

Las consideraciones hechas, permiten concluir que:

Si en una matriz de correlación transitiva, se incluyen n variables de las cuales n-k aumentan (o disminuyen) a la vez y las otras k tienen el mismo comportamiento pero en sentido contrario, se tendrá:

$(n – k)^2 + k^2$ signos positivos

$2k (n – k)$ signos negativos.

Nótese que la suma de signos positivos y negativos da como resultado n^2.

Una matriz de correlación transitiva, tiene un número máximo de signos negativos, el cual se calcula a continuación:

Número Máximo de Signos Negativos en una Matriz de Correlación Transitiva:

Llámese z al número de signos negativos en una matriz de correlación transitiva.

$z = 2k(n-K)$

$z = 2kn – 2k^2$

Asumamos que k es una variable continua (realmente no lo es, pero si el valor de k que maximiza es entero, no habrá dificultad.)

Derivemos z con respecto a k e igualemos a cero:
$dz/dk = 2n - 4k = 0$

Se obtiene $k = n/2$
k es entero cuando n es par. Por lo tanto si n es par, el número máximo de signos negativos en la matriz será:
$z = 2(n/2)(n - n/2)$

$z = n^2/2$ número máximo de signos negativos en la matriz de correlación transitiva , cuando n es par.

La segunda derivada de z con respecto a x:
$d^2z/dx^2 = -4$ índica que verdaderamente en un máximo.

Cuando n es impar, el problema se soluciona tomando los valores de k enteros más próximos a n/2. Dichos valores son $(n - 1)/2$ y $(n+1)/2$.

Nótese que la ecuación que relaciona a z y k corresponde a una parábola con el eje focal paralelo al eje Y, que se abre hacía abajo. Por tanto los valores de z (máximos) correspondientes a los valores de k: $(n-1)/2$ y $(n + 1)/2$ deben ser iguales como se comprobará a continuación.

$z = 2((n-1)/2.(n - (n-1)/2)$
$z = (n^2 - 1)/2.$ Y en el otro caso:
$z = 2((n + 1)/2).(n - (n+1)/2)$
$z = (n^2 - 1)/2.$

Podemos concluir, que si n es par, el máximo número de signos negativos en la matriz de correlación transitiva es $n^2/2$ y si n es impar, el máximo número de signos negativos en la matriz es $(n^2 -1)/2$

Podemos también encontrar el número mínimo número de signos positivos en la matriz de correlación transitiva:

Para n par:

El mínimo número de valores positivos es t:

$t = n^2 - n^2/2 = n^2/2.$

Para n impar:

$t = n^2 - (n^2-1)/2 = (n^2 + 1)/2.$

Si en una matriz de correlación transitiva el número de variables es par, el número mínimo de signos positivos es $n^2/2$. Si el n es impar, el número máximo de signos positivos es $(n^2 + 1)/2$

A manera de ejemplo, encontraremos todas las cantidades posibles de valores negativos en una matriz de correlación transitiva, cuando $n = 11$.

$z = 2k (11 – k)$. Hágase variar k desde cero hasta 11

k	**z**
0	0
1	20
2	36
3	48
4	56
5	60
6	60
7	56
8	48
9	36
10	20
11	0

Obsérvese que justamente, para los valores $(n-1)/2$ y $(n+ 1)/2$ es decir $(11-1)/2 =5$ y $(n+1)/2$, $(11 +1)/2 = 6$ se produce el valor máximo de 60.

Leyes de los signos en una Matriz de Correlación Transitiva:

I Una Matriz de Correlación Transitiva, puede tener solamente, cualquiera de los siguientes números de signos negativos:

$z = 2k (n-k)$ donde z es el número de signos negativos, n el número de variables y k un entero que varía entre 0 y n inclusive.

II Una Matriz de Correlación Transitiva, puede tener solamente, cualquiera de los siguientes números de signos positivos:

$p = (n - k)^2 + k^2$ donde p es el número de signos positivos.

III Una Matriz de Correlación Transitiva, puede tener como máximo el siguiente número de signos negativos:

$z = n^2/2$ si n es par.

$z = (n^2 - 1)/2$ si n es impar.

Si en una matriz de correlación cualquiera, encontramos una cantidad de valores negativos posible, como los que se encontraron en la tabla, dicha matriz es candidata a clasificarse como matriz de correlación transitiva.

En el ejemplo que hemos propuesto (11 variables), si en la matriz estudiada se totalizan 48 signos negativos, podemos estar ante la presencia de una matriz de correlación transitiva. Los valores de k serán 3 u 8. Realmente es una sola respuesta porque si 3 variables evolucionan en un sentido, las ocho restantes lo hacen en sentido contrario, nótese que en cada caso, para un valor posible de z, se obtienen valores de k que sumados dan como resultado n. No habrá dificultad en adoptar como k al menor de los dos valores.

Si en una matriz de correlación, encontramos un z factible, no podemos asegurar existencia de transitividad, puesto que el azar, también puede generar una matriz con igual cantidad de signos negativos sin que se trate de una matriz de correlación transitiva.

Tan solo no se puede descartar su existencia.

2.3 MÉTODO PARA DETERMINAR UNA MATRIZ DE CORRELACION TRANSITIVA

Se expondrá un método que permite detectar matrices de correlación transitivas, fundamentándose en la ley de los signos.

El procedimiento consiste en separar las variables en los dos conjuntos, aquellas que pertenecen al grupo de n-k variables (varían en el mismo sentido) y las que pertenecen al grupo que contiene k variables que varían también en un mismo sentido pero contrario al del otro conjunto. Una vez establecidos los dos conjuntos, cada uno de ellos debe formar matriz de correlación con signos todos positivos.

Se iniciara con las n variables, si el valor de z es compatible con matriz de correlación transitiva, se probara una a una todas las variables para saber si cumplen con los valores teóricos de z para posteriormente clasificarlas n el subconjunto al que pertenecen.

Supóngase que se examina la variable i y que la matriz original posee z signos negativos que corresponden a un valor de k.

Se excluye la variable i. Si en la matriz (n-1) x (n-1) aparece z signos negativos que corresponden a k-1 variables (con respecto a la matriz nxn), la variable i pertenece al conjunto que en la matriz nxn posee k elementos.

Si por el contrario en la matriz (n-1)x(n-1) aparecen z signos negativos que corresponden a k variables (con respecto a la matriz nxn), entonces la variable i corresponde al conjunto que en la matriz original nxn tiene n-k elementos.

No obstante puede suceder que la variable a clasificar, al retirarse de la matriz de n variables, produzca una disminución de signos negativos compatible con una variable perteneciente al conjunto de k variables o al conjunto de n-k variables, pero al intentar ingresar a su conjunto correspondiente, suceda que su correlación con el resto de variables de ese conjunto no sea positiva, lo que significa que en realidad esa matriz no es de correlación transitiva.

Este chequeo es mas sencillo de lo que podría aparentar, simplemente se coteja la correlación de la variable estudiada con la primera que ingresó al subconjunto. En el caso especial que la primera variable ingresada al subconjunto no guarde correlación positiva con las que entren posteriormente, el error se detectara al chequear esa positividad con la segunda variable ingresada, en ese caso, no puede definirse si fue la primera o la segunda la “falsa” eso no interesa, puesto que la incompatibilidad presentada descarta toda la matriz.

En la situación particular en que una sola variable se oponga a las n-1 restantes, se chequeara porque su correlación con la primera de las variables del otro conjunto necesariamente debe ser negativa.

Por último, el número de variables ingresadas al conjunto de k variables, justamente debe tener esa cantidad de elementos, y el número de variables ingresadas al conjunto de n-k variables debe contener esa cantidad de variables.

Cuando se trata de un numero par de variables, podría suceder que la matriz contenga el máximo valor de z, entonces k será igual a n-k, en otras palabras los dos subconjuntos deben contener igual numero de variables k, la clasificación en uno y otro conjunto, se practicará en este caso muy particular, de acuerdo al signo de la correlación con las dos primeras variables ingresadas en cada subconjunto. (Véase ejemplo #1 y #2 en resultados)

2.4 MATRIZ DE CONFRONTACION

2.4.1 Definición: Es aquella matriz de correlación transitiva, en la cual se han colocado primero las variables pertenecientes al conjunto de las n-k variables y luego las que pertenecen al conjunto de las k variables (o viceversa).

2.4.2 Características de la Matriz de Confrontación: Se trata de una matriz conformada por cuatro submatrices. En la diagonal principal dos matrices cuadradas con todos los signos positivos, son submatrices de k x k elementos la primera y n-k x n-k elementos en la segunda. En la diagonal secundaria aparecen dos submatrices con todos los signos negativos, son matrices de n-k x k la primera y k x n-k la segunda. Estas submatrices de la diagonal secundaria son transpuestas la una de la otra (véase ejemplo # 3 en resultados)

2.5 DETERMINACION DE SUBMATRICES DE CORRELACION TRANSITIVA (a partir de una matriz que no lo es)

Se puede afirmar que toda matriz de correlación posee submatrices de correlación transitivas, es evidente que absolutamente todas las submatrices de correlación entre dos variables, lo son.

Se propone el siguiente procedimiento para buscar submatrices de correlación transitiva, a partir de una matriz de correlación no transitiva.

Se busca en la matriz original la correlación de menor valor absoluto, y se retira las dos variables comprometidas.

Lo anterior considerando que esas correlaciones débiles son las que pueden oscilar fácilmente entre los signos positivos y negativos comprometiendo la transitividad.

De lograrse con esta maniobra la obtención de una submatriz de correlación transitiva, es conveniente, regresar una de ellas, la que tenga en promedio de valores absolutos, mayor correlación con las otras variables. Si con esa actividad, se logra matriz transitiva, se regresa al lote de las demás variables.

Si al excluir las dos variables, no se obtiene matriz de correlación transitiva, se descartan y se ensaya la supresión de dos nuevas variables, siguiendo las pautas expuestas. (Véase ejemplo # 4 en resultados)

3. RESULTADOS

3.1 APLICACIÓN DE DETERMINACION DE MATRIZ DE CORRELACION TRANSITIVA

Ejemplo # 1

Se utilizara como ejemplo las 7 variables tomadas de un reporte de puntajes y hábitos de los estudiantes de grado once del Colegio San Francisco de Pasto.

Puntaje es la calificación lograda por cada alumno en un examen de conocimientos generales de bachillerato.

Las variables Estudio, Televisión, Sueño, Internet, Deportes y Artes se miden en el número de horas que dedican a esa actividad en el día.

Individuo	Puntaje	Estudio	Televisión	Sueño	Internet	Deportes	Artes
1	58	0	6	6	2	0	1
2	68	1	5	7	1	1	1
3	91	2	3	9	1	1	2
4	71	3	3	5	0	0	3
5	81	2	4	8	1	1	1
6	94	4	2	7	2	2	0
7	92	5	2	7	0	1	2
8	83	3	2	6	2	0	2
9	89	6	1	6	1	0	0
10	90	5	0	5	1	0	3
11	63	1	5	6	2	1	0
12	68	4	2	3	2	0	1
13	65	3	3	4	2	1	0
14	42	2	4	2	0	2	1
15	62	1	5	6	1	1	1
16	94	5	1	7	1	2	0
17	104	6	0	8	0	3	0
18	74	0	6	9	2	1	0
19	68	1	5	7	1	1	1
20	62	2	4	5	0	1	2
21	84	4	2	6	2	1	0
22	81	3	3	7	0	1	2
23	81	2	4	8	1	1	1
24	87	2	4	9	1	1	1
25	106	5	1	9	1	2	0
26	98	6	0	6	0	0	3

27	73	2	4	6	3	0	0
28	60	1	5	5	3	0	0
29	86	3	3	7	3	0	0
30	69	0	6	8	1	0	2
31	82	2	4	9	0	3	0
32	102	4	2	9	2	1	0
33	112	5	1	9	3	0	0
34	114	6	0	9	0	1	2
35	78	2	4	8	0	2	1
36	82	3	3	7	1	1	1
37	73	3	3	5	2	0	1
38	99	4	2	8	3	0	0
39	79	2	4	8	1	2	0
40	56	0	6	6	2	1	0
41	109	5	1	9	0	0	3
42	78	2	4	7	2	0	1
43	86	3	3	7	3	0	0
44	75	1	5	8	0	0	3
45	107	5	1	9	0	1	2
46	89	4	2	7	1	1	1
47	71	2	3	6	1	2	1
48	79	1	5	8	0	1	3
49	74	1	5	8	1	1	1
50	66	0	6	7	2	2	0

La matriz de correlaciones 7x7 será:

1	0,784039	-0,786223	0,649395	-0,116235	0,0180804	0,0860377
0,784039	1	-0,980953	0,0740385	-0,178637	-0,0182858	0,0886585
-0,786223	-0,980953	1	-0,0682896	0,162418	0,0151164	-0,134368
0,649395	0,0740385	-0,0682896	1	-0,145509	0,209261	-0,0121664
-0,116235	-0,178637	0,162418	-0,145509	1	-0,389965	-0,636678
0,0180804	-0,0182858	0,0151164	0,209261	-0,389965	1	-0,336538
0,0860377	0,0886585	-0,134368	-0,0121664	-0,636678	-0,336538	1

Resulta que esta matriz posee 24 elementos con signo negativo.

Se calcula los valores posibles de z en caso que se trate de una matriz de correlación transitiva, para n=7

K	Z
0	0
1	12
2	20
3	24
4	24
5	20
6	12
7	0

24 es un número z compatible con matriz de correlación transitiva, también se podría aplicar 2k(n-k) = z, 2k(7-k) = 24

Ecuación de segundo grado que entrega como resultados los valores enteros k=3 y k= 4.

Se puede aceptar que k=3 y n-k =4

Si se excluye la variable 7, la matriz de correlación será:

1	0,784039	-0,786223	0,649395	-0,116235	0,0180804
0,784039	1	-0,980953	0,0740385	-0,178637	-0,0182858
-0,786223	-0,980953	1	-0,0682896	0,162418	0,0151164
0,649395	0,0740385	-0,0682896	1	-0,145509	0,209261
-0,116235	-0,178637	0,162418	-0,145509	1	-0,389965
0,0180804	-0,0182858	0,0151164	0,209261	-0,389965	1

Esta matriz posee 16 elementos negativos

2k(n-k) = 16, n=6

2k(6-k) = 16

Los resultados solución para esta ecuación de segundo grado son:

K= 2 y k= 4, obviamente si el k para la matriz 7x7 era 3, el nuevo k será k=2, si k disminuye al quitar la variable 7, esta variable se incluye en el conjunto I de k variables.

La variable 7 pertenece al conjunto de k variables.

Excluyamos de la matriz de correlaciones 7x7 la variable 6, la matriz así modificada será:

1	0,784039	-0,786223	0,649395	-0,116235	0,0860377
0,784039	1	-0,980953	0,0740385	-0,178637	0,0886585
-0,786223	-0,980953	1	-0,0682896	0,162418	-0,134368
0,649395	0,0740385	-0,0682896	1	-0,145509	-0,0121664
-0,116235	-0,178637	0,162418	-0,145509	1	-0,636678
0,0860377	0,0886585	-0,134368	-0,0121664	-0,636678	1

18 = 2k(6-k)

La ecuación de segundo grado tiene solución única para k=3 , significa que k no disminuyó con respecto a la matriz original 7x7, por lo tanto la variable 6 pertenece al conjunto que contiene n-k variables.

Exclúyase la variable 5, tendremos la matriz:

1	0,784039	-0,786223	0,649395	0,0180804	0,0860377
0,784039	1	-0,980953	0,0740385	-0,0182858	0,0886585
-0,786223	-0,980953	1	-0,0682896	0,0151164	-0,134368
0,649395	0,0740385	-0,0682896	1	0,209261	-0,0121664
0,0180804	-0,0182858	0,0151164	0,209261	1	-0,336538
0,0860377	0,0886585	-0,134368	-0,0121664	-0,336538	1

Esta matiz posee 14 elementos negativos

14= 2k(6-k)

Esta ecuación de segundo grado no tiene soluciones enteras, por lo tanto la variable 5 no puede incluirse en ninguno de los dos conjuntos. En este momento ya se puede descartar la matriz como de correlación transitiva, no obstante y con fines ilustrativos continuaremos el procedimiento.

Exclúyase la variable 4 de la matriz:

1	0,784039	-0,786223	-0,116235	0,0180804	0,0860377
0,784039	1	-0,980953	-0,178637	-0,0182858	0,0886585
-0,786223	-0,980953	1	0,162418	0,0151164	-0,134368
-0,116235	-0,178637	0,162418	1	-0,389965	-0,636678
0,0180804	-0,0182858	0,0151164	-0,389965	1	-0,336538
0,0860377	0,0886585	-0,134368	-0,636678	-0,336538	1

Esta matriz tiene 18 elementos negativos, k=3, si ello es así y siguiendo reflexiones anteriores la variable 4 pertenece al conjunto que posee n-k variables.

Nótese que la correlación de la variable 4 con la variable 6 (primera en ingresar al conjunto de las n-k) variables es positiva. No hay dificultad con este chequeo.

Ahora excluimos la variable 3:

1	0,784039	0,649395	-0,116235	0,0180804	0,0860377
0,784039	1	0,0740385	-0,178637	-0,0182858	0,0886585
0,649395	0,0740385	1	-0,145509	0,209261	-0,0121664
-0,116235	-0,178637	-0,145509	1	-0,389965	-0,636678
0,0180804	-0,0182858	0,209261	-0,389965	1	-0,336538
0,0860377	0,0886585	-0,0121664	-0,636678	-0,336538	1

La matriz contiene 16 variables, siguiendo reflexiones anteriores, la variable 3 pertenece al conjunto que contiene k variables.

Chequeamos ahora el signo de correlación de la variable 3 con la variable 7 su valor es negativo, estas dos variables no pueden estar en el mismo conjunto. Razón también suficiente para descartar la matriz como de correlación transitiva. Continuaremos con el estudio solo por propósitos explicativos.

Excluyamos la variable 2

1	-0,786223	0,649395	-0,116235	0,0180804	0,0860377
-0,786223	1	-0,0682896	0,162418	0,0151164	-0,134368
0,649395	-0,0682896	1	-0,145509	0,209261	-0,0121664
-0,116235	0,162418	-0,145509	1	-0,389965	-0,636678
0,0180804	0,0151164	0,209261	-0,389965	1	-0,336538
0,0860377	-0,134368	-0,0121664	-0,636678	-0,336538	1

Esta matriz posee 18 variables, entonces la variable 2 pertenece al conjunto de n-k variables.

Cotejada esta variable con la primera en ingresar al conjunto de n-k variables, (variable 6) encontramos correlación negativa, las dos variables no pueden pertenecer al mismo conjunto. otra razón que por si sola descarta la matriz como de correlación transitiva.

Finalmente se excluye la variable 1:

1	-0,980953	0,0740385	-0,178637	-0,0182858	0,0886585
-0,980953	1	-0,0682896	0,162418	0,0151164	-0,134368
0,0740385	-0,0682896	1	-0,145509	0,209261	-0,0121664
-0,178637	0,162418	-0,145509	1	-0,389965	-0,636678
-0,0182858	0,0151164	0,209261	-0,389965	1	-0,336538
0,0886585	-0,134368	-0,0121664	-0,636678	-0,336538	1

La matriz posee 20 elementos negativos:

$20 = 2k(6 – k)$ La ecuación de segundo grado no tiene solución entera, por lo tanto la variable 1 no puede incluirse en ninguno de los dos conjuntos.

No pueden formarse los conjuntos buscados.

Demasiadas razones para descartar la matriz como de correlación transitiva.

Seguramente una o más variables no se adaptan al modelo de las matrices transitivas.

A continuación se expondrá el ejemplo de una matriz de correlaciones que si cumple con las condiciones de transitividad.

Ejemplo # 2

Considérese el siguiente cuadro de datos:

Fueron tomados del Libro de Partos del Hospital San Pedro de Pasto y corresponden a los nacimientos sucedidos durante la primera quincena del mes de diciembre de 2007.

Las variables se midieron en las siguientes unidades: Edad, es la edad de la madre en años. Peso, es el peso del recién nacido en gramos, Ingreso, mide el ingreso en miles de pesos por persona en el hogar cada mes. Hijos, es el número de hijos anteriores. Trabajo de Parto, es el número de horas experimentadas por la madre en esta etapa. Años de Estudio, mide el número de años cursados después de la primaria por la madre. Edad Gestacional mide el número de semanas que duró el embarazo.

Edad	Peso	Ingreso	Hijos	Trab. Parto	Años Estudio	Edad Gest
16	3500	400	0	1	5	43
18	3200	350	0	3	7	43
25	3000	200	3	5	6	41
40	2500	100	0	10	5	38
30	3000	200	3	8	6	40
22	3200	220	2	4	6	42
35	2400	100	1	9	3	40
29	2300	100	2	8	2	41
24	3500	350	2	5	10	42
35	2600	120	4	9	6	40
39	2500	100	1	8	2	39
22	3300	230	0	3	6	42
17	3500	350	0	2	8	43
32	2200	90	2	8	0	40
42	2000	80	4	10	1	38
29	2800	180	0	8	5	41
33	2700	170	0	8	5	40
25	3400	380	1	5	9	41
16	3700	400	0	4	5	44
34	2200	90	2	8	0	40
17	4000	500	1	2	6	43
36	2200	100	0	8	2	39
20	3800	400	1	3	8	42
35	2500	100	1	8	2	39
40	2400	90	0	9	1	38
25	3500	420	2	6	11	41
25	3600	430	0	5	11	41
19	3800	400	0	3	8	43
38	2000	70	5	8	0	40

Se tratan de 7 variables cuya matriz de correlación es:

1	-0,889056	-0,871264	0,369523	0,945219	-0,650496	-0,966197
-0,889056	1	0,958178	-0,406618	-0,861289	0,834368	0,816003
-0,871264	0,958178	1	-0,370262	-0,845813	0,809662	0,803517
0,369523	-0,406618	-0,370262	1	0,377157	-0,261636	-0,291975
0,945219	-0,861289	-0,845813	0,377157	1	-0,586406	-0,898693
-0,650496	0,834368	0,809662	-0,261636	-0,586406	1	0,559133
-0,966197	0,816003	0,803517	-0,291975	-0,898693	0,559133	1

Esta matriz posee 24 signos negativos, valor que si es posible en una matriz transitiva, ello sucede para k=3 o k=4 en la expresión

Z = 2k(7-k). Esto indica que 4 variables evolucionan en un sentido y 3 en el contrario, no hay dificultad en tomar k=3 (siempre se toma el menor valor)

Al suprimir variables una por una, solo es factible encontrar en cada etapa valores de k=3 o valores de k=2. Si el valor es 3, la variable excluida entrará a formar parte del conjunto de variables iníciales con n-k elementos, conjunto que llamaremos B, si el valor es 2 la variable excluida formará parte del conjunto de variables iníciales con k variables al cual llamaremos A. Cualquier otro valor en esas submatrices de n-1 elementos, descarta matriz de correlación transitiva.

Suprimamos entonces la variable 7, ahora la matriz es:

1	-0,889056	-0,871264	0,369523	0,945219	-0,650496
-0,889056	1	0,958178	-0,406618	-0,861289	0,834368
-0,871264	0,958178	1	-0,370262	-0,845813	0,809662
0,369523	-0,406618	-0,370262	1	0,377157	-0,261636
0,945219	-0,861289	-0,845813	0,377157	1	-0,586406
-0,650496	0,834368	0,809662	-0,261636	-0,586406	1

Matriz que posee 18 elementos negativos.

18=2k(6-k) ecuación de la cual es factible el valor k=3, entonces la variable 7 entra a formar parte del conjunto B.

Suprímase la variable 6:

1	-0,889056	-0,871264	0,369523	0,945219	-0,966197
-0,889056	1	0,958178	-0,406618	-0,861289	0,816003
-0,871264	0,958178	1	-0,370262	-0,845813	0,803517
0,369523	-0,406618	-0,370262	1	0,377157	-0,291975
0,945219	-0,861289	-0,845813	0,377157	1	-0,898693
-0,966197	0,816003	0,803517	-0,291975	-0,898693	1

Esta matriz contiene 18 signos negativos, k=3, la variable 6 pertenece al conjunto B.

Nótese que la correlación entre la variable 6 y la 7 es positiva, por lo tanto es verdad que ambas pertenecen al conjunto B.

Suprímase la variable 5:

1	-0,889056	-0,871264	0,369523	-0,650496	-0,966197
-0,889056	1	0,958178	-0,406618	0,834368	0,816003
-0,871264	0,958178	1	-0,370262	0,809662	0,803517
0,369523	-0,406618	-0,370262	1	-0,261636	-0,291975
-0,650496	0,834368	0,809662	-0,261636	1	0,559133
-0,966197	0,816003	0,803517	-0,291975	0,559133	1

Esta matriz contiene 16 signos negativos.

$16= 2k (6 – k)$ ecuación que se satisface para $k=2$. Por lo tanto la variable 5 pertenece al conjunto A.

Suprímase la variable 4.

1	-0,889056	-0,871264	0,945219	-0,650496	-0,966197
-0,889056	1	0,958178	-0,861289	0,834368	0,816003
-0,871264	0,958178	1	-0,845813	0,809662	0,803517
0,945219	-0,861289	-0,845813	1	-0,586406	-0,898693
-0,650496	0,834368	0,809662	-0,586406	1	0,559133
-0,966197	0,816003	0,803517	-0,898693	0,559133	1

La matriz contiene 16 signos negativos, como ya se indico, es compatible con k=2, la variable 4 ingresa al conjunto A.

La variable 4 y la 5 poseen correlación positiva, si pertenecen al mismo conjunto.

Suprímase la variable 3.

1	-0,889056	0,369523	0,945219	-0,650496	-0,966197
-0,889056	1	-0,406618	-0,861289	0,834368	0,816003
0,369523	-0,406618	1	0,377157	-0,261636	-0,291975
0,945219	-0,861289	0,377157	1	-0,586406	-0,898693
-0,650496	0,834368	-0,261636	-0,586406	1	0,559133
-0,966197	0,816003	-0,291975	-0,898693	0,559133	1

La matriz posee 18 signos negativos compatibles con k=3, la variable 3 entra a formar parte del conjunto B

La variable 3 y la 7, tienen correlación positiva, efectivamente pertenecen al mismo conjunto.

Suprímase la variable 2

1	-0,871264	0,369523	0,945219	-0,650496	-0,966197
-0,871264	1	-0,370262	-0,845813	0,809662	0,803517
0,369523	-0,370262	1	0,377157	-0,261636	-0,291975
0,945219	-0,845813	0,377157	1	-0,586406	-0,898693
-0,650496	0,809662	-0,261636	-0,586406	1	0,559133
-0,966197	0,803517	-0,291975	-0,898693	0,559133	1

La matriz posee 18 elementos negativos, k=3, la variable 2 será elemento del conjunto B

La variable 2 y la 7, poseen correlación positiva eso significa que pertenecen al mismo conjunto.

Suprímase finalmente la variable 1.

1	0,958178	-0,406618	-0,861289	0,834368	0,816003
0,958178	1	-0,370262	-0,845813	0,809662	0,803517
-0,406618	-0,370262	1	0,377157	-0,261636	-0,291975
-0,861289	-0,845813	0,377157	1	-0,586406	-0,898693
0,834368	0,809662	-0,261636	-0,586406	1	0,559133
0,816003	0,803517	-0,291975	-0,898693	0,559133	1

Matriz que tiene 16 elementos negativos, k=2, la variable 1 pertenece al conjunto A.

La variable 1 y 5 tienen correlación positiva, si pertenecen al conjunto A

El conjunto A está formado por las variables: 1, 4 y 5.

El conjunto B está formado por las variables 2, 3, 6 y 7.

Chequeo: Efectivamente el conjunto A posee 3 variables (K=3) y el conjunto B posee 4 variables (n-k = 4)

De ser la matriz original 7x7, una matriz de correlación transitiva, el conjunto A de variables, tendrá una matriz de correlación con todos sus signos positivos. También el conjunto B tendrá una matriz de correlación con todos sus signos positivos.

Matriz de correlación para el conjunto A:

1	0,369523	0,945219
0,369523	1	0,377157
0,945219	0,377157	1

Matriz de correlación con todos sus signos positivos.

Matriz de correlación para el conjunto B:

1	0,958178	0,834368	0,816003
0,958178	1	0,809662	0,803517
0,834368	0,809662	1	0,559133
0,816003	0,803517	0,559133	1

Matriz de correlación con todos los signos positivos.

Esta matriz 7x7 verdaderamente y sin lugar a dudas era una matriz de correlación transitiva.

3.2 APLICACION MATRIZ DE CONFRONTACION

Ejemplo # 3. Se emplearan los resultados del Ejemplo # 2

Para la que se denominará Matriz de Confrontación, se colocaran las variables en el siguiente orden: primero las del conjunto A y luego las del conjunto B el orden será 1,4,5, 2,3,6,7.

Es interesante observar la matriz conformada por dos submatrices cuadradas diagonales entre si con signos de correlación positiva y otras dos submatrices también diagonales con signos de correlación negativa, donde una de ellas es la transpuesta de la otra, esta es una expresión estética y llamativa de la Matriz de Correlación Transitiva.

1	0,369523	0,945219
0,369523	1	0,377157
0,945219	0,377157	1

-0,208298	-0,889056	-0,871264	-0,650496
-0,0376114	-0,406618	-0,370262	-0,261636
-0,196462	-0,861289	-0,845813	-0,586406

-0,208298	-0,0376114	-0,196462
-0,889056	-0,406618	-0,861289
-0,871264	-0,370262	-0,845813
-0,650496	-0,261636	-0,586406

1	0,141647	0,168926	0,134623
0,141647	1	0,958178	0,834368
0,168926	0,958178	1	0,809662
0,134623	0,834368	0,809662	1

3.3 APLICACIÓN DETERMINACION DE SUBMATRICES DE CORRELACION TRANSITIVAS

Ejemplo # 4

A manera de ejemplo retomaremos la matriz de correlación no transitiva derivada de los datos de los estudiantes del Colegio San Francisco cuya matriz primitiva, no es de correlación transitiva.

1	0,784039	-0,786223	0,649395	-0,116235	0,0180804	0,0860377
0,784039	1	-0,980953	0,0740385	-0,178637	-0,0182858	0,0886585
-0,786223	-0,980953	1	-0,0682896	0,162418	0,0151164	-0,134368
0,649395	0,0740385	-0,0682896	1	-0,145509	0,209261	-0,0121664
-0,116235	-0,178637	0,162418	-0,145509	1	-0,389965	-0,636678
0,0180804	-0,0182858	0,0151164	0,209261	-0,389965	1	-0,336538
0,0860377	0,0886585	-0,134368	-0,0121664	-0,636678	-0,336538	1

La menor correlación, se encuentra entre las variables 4 y 7, por lo tanto se excluyen:

1	0,784039	-0,786223	-0,116235	0,0180804
0,784039	1	-0,980953	-0,178637	-0,0182858
-0,786223	-0,980953	1	0,162418	0,0151164
-0,116235	-0,178637	0,162418	1	-0,389965
0,0180804	-0,0182858	0,0151164	-0,389965	1

1 2 3 5 6

Esta matriz posee 12 signos negativos,

12 = 2k(5- k), se obtiene k = 2. Podríamos estar frente a una submatriz de correlación transitiva

Eliminemos la variable 6

1	0,784039	-0,786223	-0,116235
0,784039	1	-0,980953	-0,178637
-0,786223	-0,980953	1	0,162418
-0,116235	-0,178637	0,162418	1

1 2 3 5

Se tienen 8 signos negativos

8 = 2k(4-k) Tiene una única solución para k=2. Por lo tanto la variable 6 ingresa al conjunto de n-k variables.

Eliminemos de la matriz de 5 variables la variable 5

1	0,784039	-0,786223	0,0180804
0,784039	1	-0,980953	-0,0182858
-0,786223	-0,980953	1	0,0151164
0,0180804	-0,0182858	0,0151164	1
1	2	3	6

Existen 6 signos negativos

6 = 2k(4-k) se cumple para k=1

La variable 5 entra al conjunto de k variables.

Excluyamos la variable 3:

1	0,784039	-0,116235	0,0180804
0,784039	1	-0,178637	-0,0182858
-0,116235	-0,178637	1	-0,389965
0,0180804	-0,0182858	-0,389965	1
1	2	5	6

Existen 8 signos negativos ingresa al conjunto de n-k variables la correlación entre las variables 3 y 6 es positiva, se cumple el chequeo.

Eliminemos la variable 2.

1	-0,786223	-0,116235	0,0180804
-0,786223	1	0,162418	0,0151164
-0,116235	0,162418	1	-0,389965
0,0180804	0,0151164	-0,389965	1

1 3 5 6

Esta matriz posee 6 signos negativos.

Ingresa al conjunto de k variables, pero la correlación entre las variables 2 y 5 es negativa, argumento suficiente para descartar la submatriz de 5 variables como submatriz de correlación transitiva.

A partir de la submatriz de 5 variables, eliminaremos las variables 3 y 6 que poseen la menor correlación en valor absoluto. Obtendremos la siguiente submatriz:

1	0,784039	-0,116235
0,784039	1	-0,178637
-0,116235	-0,178637	1

1 2 5

Esta submatriz muestra directamente la forma de matriz de confrontación, por lo tanto es de correlación transitiva, pero es posible que hayamos eliminado una variable de más. De las dos eliminadas, se retorna la de mayor valor promedio de correlaciones en valor absoluto, es decir la variable 3.

1	0,784039	-0,786223	-0,116235
0,784039	1	-0,980953	-0,178637
-0,786223	-0,980953	1	0,162418
-0,116235	-0,178637	0,162418	1

1 2 3 5

En este caso también se presenta directamente una forma de submatriz de confrontación que obviamente es de correlación transitiva.

Hemos obtenido así, una submatriz de correlación transitiva a partir de una matriz de correlación que no lo es, se trata en este caso de la submatriz de correlación transitiva más grande posible. Con 4 variables.

CONCLUSIONES

1. Es posible determinar si una matriz de correlación es transitiva, simplemente realizando conteos de signos negativos, sometiendo dichos valores a las leyes de los signos y aplicando las metodologías descritas.

2. Si una matriz es de correlación transitiva, pueden separarse las variables en dos subconjuntos A y B, al interior de cada uno de esos subconjuntos, las correlaciones son positivas. Cada uno de los elementos de A tiene correlaciones negativas con los elementos de B.

3. Los subconjuntos de variables A y B, permiten formar la Matriz de Confrontación constituida por dos matrices cuadradas de signos positivos en la diagonal mayor y dos matrices negativas una transpuesta de otra en la diagonal menor.

4. Toda matriz de correlación posee submatrices de correlación transitivas, las cuales pueden obtenerse siguiendo metodología explicada.

5. Se abre la discusión sobre la aplicabilidad de las Matrices de Correlación Transitiva.

6. Debate que debe partir del hecho que las variables comprometidas pueden dividirse en dos grupos que apuntan en sentido contrario.

BIBLIOGRAFIA

- Cuadras C. M. (1993) Métodos de Análisis Multivariante, Editorial Universitaria de Barcelona.

- Morris H. Degroot (1988), Probabilidad y Estadística, segunda edición, Addison-Wesley Iberoamericana, S.A. Wilmington, Delaware, E.U.A.

- Peña Daniel (2001) , Fundamentos de Estadística, Editorial Alianza, Madrid.

- Sheldon Ross (1993), A First Course in Probability, Sixth Edition, Prentice Hall, Upper Saddle River, New Jersey. USA.