

Empleo de las Componentes Principales en sus investigaciones biológicas.

II.- Obtención de variables canónicas.

Guillermo R. Pardo Cardoso*,Rafael Avilés Merens*, Guillermo J. Pardo Camacho.**

* Profesores, Facultad de Ciencias Agropecuarias Universidad de Camaguey. E-mail gparado1942@yahoo.es, gparado@cag.reduc.edu.cu .y ravilesmerens@cag.reduc.edu.cu.

** Especialista Primer Grado, Hospital Provincial Docente "Manuel Ascunce Domenech". Camaguey. E-mail gparado@shine.cmw.sld.cu.

RESUMEN

Se analizaron 435 variables en 4876 unidades experimentales, procedentes de 23 experimentos en las ciencias biológicas, realizados en el período 1990-2004, que fueron significativos a las pruebas de Kaiser-Mayer –Olkin y Bartlet, con vistas a valorar , cuando se emplea la técnica de Componentes Principales, la metodología a utilizar en la obtención de las variables canónicas. Los resultados obtenidos en el número de variables canónicas por experimento, variables independientes y variabilidad acumulada (%) fueron $2,8 \pm 0,08$; $1,5 \pm 0,51$ y $61,1 \pm 0,05$ respectivamente. El comportamiento de las variables respuestas y las Dósimas necesarias a realizar cuando I.- Se Aplica la Metodología y II.- No se Aplica se fueron de $I = 19,0$ y $II = 4,3 \pm 1,22$ ($p < 0,05$) Variables Respuesta y $I=66,8$ y $II = 14, 6 \pm 4,89$ ($p < 0,05$) Dósimas realizadas. Se concluye que se produce una disminución significativa en las pruebas estadísticas con las variables canónicas obtenidas ($p < 0,05$), obteniéndose que en el 17% de los experimentos analizados las variables transformadas no fueron independientes siendo necesario aplicar nuevamente el método propuesto, lo que exige que las pruebas de esfericidad y Kaiser-Mayer-Olkin se efectúen para garantizar la independencia de las variables a procesar en la toma de decisiones.

Palabras clave: Variables Canónicas, Componentes Principales, Metodología, Ciencias Biológicas

Introducción

Carvajal et al. (2001) proponen pruebas multivariadas para aquellos análisis estadístico que contemplen la evaluación simultánea de un grupo de variables sobre una misma unidad experimental, sugiriéndose el uso de las Componentes Principales en aquellos casos donde no se conozca con certeza su posible relación (Salvador 2003) .

Una forma de evaluar los resultados obtenidos con las componentes principales, consiste en la obtención de nuevos ejes ortogonales, a los cuales se pueden proyectar todas las variables analizadas (Sharma 1998), de esta forma se obtienen nuevas variables compuestas (Canónicas), que contemplan la inclusión de todas las variables analizadas tomando como elemento de transformación los métodos de Regresión, Bartlett y Andersson-Rubin (Visauta 1998).

De acuerdo a lo antes señalado la obtención de las variables canónicas mediante la proyección de los vectores propios de todas las variables originales sobre cada componente garantizan su independencia, no obstante si se valorará un método que permita la transformación solamente de las variables más importantes de las componentes seleccionadas, precisando el comportamiento de la independencia de las variables finales, permitirá entonces la reducción y el análisis individual de cada una de las variables transformada en el experimento.

Materiales y métodos

Se seleccionaron los datos primarios de 23 experimentos evaluados en el período 1990-2004, conformados por: 9 de Medicina Humana, 5 de Aves, 5 de Cerdos, y 2 de Farmacia, los cuales tenían 12 o más variables evaluadas sobre la misma unidad experimental y que respondían a los objetivos propuestos (Variables Respuesta) y dos o más factores en estudio (Variables que respondían a las Hipótesis formulada en cada experimento). Los 23 experimentos seleccionados fueron significativos a la prueba de esfericidad de Bartlett y de adecuación de la muestra de Kaiser-Mayer-Olkin (KMO), a cada experimento se les aplicó la técnica multivariada de Componentes Principales

La aplicación de este método contempla conformar solamente las variables canónicas con las originales que posean correlaciones altas. ($r = /0,60/$) empleando los siguientes criterios:

- 1.- Realizar prueba de Bartlett y KMO a las variables que responden a cada objetivo específico
- 2.- Seleccionar las componentes que cumplan las siguientes condiciones simultáneamente:
 - a) Variabilidad explicada acumulada del 60% o mayor (Aproximadamente dos o tres ya que estas pueden representarse gráficamente en un sistema de coordenadas rectangulares).
 - b) Que el total de componentes seleccionadas la conformen el mayor número de variables originales y no se presenten variables complejas
 - c) Participación en cada componente de las variables con aproximadamente una $r > 0,60$
- 3.- Obtener las puntuaciones o sea el valor que obtendrá cada unidad experimental en la nueva variable propuesta que está conformada por aquellas que poseen la correlación mayor o igual a la decidida por el investigador ($r \geq /0,60/$) en cada componente mediante la expresión

$$Y_k = \sum A_{ki} X_i$$

Donde Y_k = Variable Canónica Obtenida ($k = 1. \dots$ Componentes seleccionadas)

A_{ki} = Vector Propio de la componente k-ésima de la Variable i-ésima

X_i = Valor de la Variable i-ésima seleccionada de la componente k-ésima
y $r > /0,60/$

- 4.- Asignar a cada Y_k un nombre que contemple simultáneamente las variables que intervienen en ella, considerándola entonces como una nueva variable, que recibe el nombre de Canónica.

5.- Se repite la prueba de Bartlett y KMO a las nuevas variables en estudio o sea las canónicas y aquellas que no fueron seleccionadas en ninguna componente por presentar correlaciones menores a $r < /0,60/$ y si es necesario se repite el proceso nuevamente a partir del punto 2

Ya aplicado este método se calculó cuantas Dótimas se realizaron en el caso de no realizar la transformación y cuando esta se hizo (Anexo 1)., a todos los indicadores obtenidos en la metodología propuesta se les verificó su esfericidad mediante la prueba de Bartlett y Kaiser-Mayer-Olkin la cual garantiza que entre las variables analizadas no se presente correlación (Salvador 2003)

A los resultados obtenidos se les estimó con una confiabilidad del 95 % el número de variables canónicas, variables independientes y variabilidad acumulada por cada experimento, así como la proporción de experimentos en que fue necesario repetir la Prueba de Bartlett y KMO, se determinó si la metodología utilizada reduce las variables a evaluar y el número de Dótimas a efectuar, todos los resultados obtenidos fueron procesados utilizando el paquete estadístico SPSS, (2001).

Resultados y Discusión

Al aplicar el método propuesto (Tabla 1) se tiene como aspecto importante que aproximadamente en el 17 % de los casos, no se cumplió la independencia de las variables finales a evaluar, esto complica el trabajo a realizar ya que es necesario repetir nuevamente el proceso de reducción de variables lo cual consecuentemente hace perder la proporción total de la variabilidad que esperábamos explicar, no obstante es necesario mantener la prueba de significación al nuevo grupo de variables para poder garantizar su independencia y poder aplicar, cualquier método estadístico de decisión que contemple la aditividad del modelo (Huberty 1994 y Gallo 2001)

Tabla 1.- Comportamiento de la muestra analizada al aplicar la metodología

Indicador	μ	$\pm \sigma$
No de Variables Canónicas	2,8	0,08
No de Variables Independientes	1,5	0,51
Variabilidad Acumulada (%)	61,1	0,05
Significativo a la Prueba de Esfericidad (Bartlett KMO) (%)	17,4	15,49

Evidentemente las puntuaciones serán exactas sólo cuando dentro de cada componente se utilizan todas las variables originales aunque las mismas presenten una correlación baja (Sharma 1998) lo que conduce a que en este caso las componentes seleccionadas sean ortogonales.

En términos geométricos, el subespacio lineal desplegado por las primeras k componentes principales ofrece el mejor ajuste a los puntos de datos medido como la suma cuadrada de las distancias perpendiculares de cada punto respecto al subespacio, en contraste con la interpretación geométrica de la regresión, que minimiza la suma cuadrada de las distancias verticales de los puntos respecto a la media (Hair et al. 1999).

Por otra parte tal como aparece en la Tabla 2 nótese como el número de Dótimas efectuadas y la Cantidad de Variables Respuesta a los Objetivos planteados presentan una reducción significativa.

Tabla 2.- Resultado obtenidos al aplicar la metodología propuesta a los experimentos analizados

Indicador	Factor en Estudio		± ES	Sign
	Sin Aplicar Metodología	Aplicando Metodología		
No de Variables Respuesta	19,0	4,3	1,22	*
Dótimas Realizadas	66,8	14,6	4,89	*
<ul style="list-style-type: none"> * $p < 0,05$ 				

Es evidente que la reducción de variables simplifica la interpretación y análisis de los resultados, sobre todo porque se conoce que la primera componente principal da la solución mínimo cuadrática del modelo $Y = XB + E$ o sea, se busca hacer mínima la traza de $(E'E)$, esto es, la suma de los cuadrados del error. (McCullagh, y Nelder, 1989). lo que permite conocer a plenitud la causa de los resultados que puedan obtener a partir del análisis de la variable canónica que la representa.

Al valorar los diferentes métodos para la obtención de las puntuaciones (Pérez 2001) se tiene que:

a) En el método de regresión las puntuaciones resultantes tienen de media 0 y varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Las puntuaciones pueden estar correlacionadas incluso cuando los factores son ortogonales.

b) En el método de Bartlett las puntuaciones resultantes tienen una media de 0. Se minimiza la suma de cuadrados de los factores únicos sobre el rango de las variables.

c) El método Andersson-Rubin es una modificación del método de Bartlett, que asegura la ortogonalidad de los factores estimados. Las puntuaciones resultantes tienen una media 0, una desviación típica de 1 y no correlacionan entre sí.

En el método propuesto al no contemplar todas las variables originales dentro de cada componente, no se puede precisar su ortogonalidad, si embargo, como se desechan aquellas que aportan poco a la componente por poseer una correlación baja, prácticamente se puede valorar que las mismas son perpendiculares entre sí, lo cual se verifica con la repetición de la prueba de KMO y Bartlett, que fueron las que precisamente se seleccionaron para aplicar esta técnica multivariada Mendoza (2001), en el anexo 2 se muestra de forma simplificada la aplicación de este método.

Conclusiones

- 1.- Se produce una disminución significativa de Dósimas a efectuar al realizar las pruebas estadísticas con las variables canónicas obtenidas
- 2.- En el 17% de los experimentos analizados fue necesario aplicar nuevamente el método propuesto al ser significativo a la esfericidad las variables obtenidas, lo que exige que esta evaluación sea realizada para garantizar la independencia de las variables a procesar en la toma de decisiones.
- 3.- El método propuesto para la obtención de las variables canónicas permite el cumplimiento de una de las restricciones básicas de los modelos estadísticos propuestos para la toma de decisiones, su independencia..
- 4.- Solamente es posible aplicar esta técnica Multivariada, si las variables obtenidas han sido controladas y medidas sobre una misma unidad experimental.

REFERENCIAS

CARVAJAL P., TREJOS A.A. Y SOTO J.: Búsqueda de la relación entre áreas ICFES en Matemáticas, Física, Lenguaje y Rendimiento en Matemática I y Matemática II. Mediante Componentes Principales. [http:// planea.utp.edu.co/](http://planea.utp.edu.co/) planea. 2001

- GALLO J.P. "La vaquita marina y su habitat crítico en el alto del Golfo de California" [http:// www.ine.gob.mx /veajer/ publicaciones/ gacetas/ 153/ vaquita 2001](http://www.ine.gob.mx/veajer/publicaciones/gacetas/153/vaquita2001).
- HAIR, J., ANDERSON, R., TATHAM, R. Y BLACK, W. Análisis Multivariante. 5ª Edición. Prentice Hall.1999
- HUBERTY, C.J Applied Discriminant Analysis. Wiley. Interscience1994
- MCCULLAGH, P. AND NELDER, J.A.. Generalized Linear Models. Second Edition. Chapman & Hall.1989
- MENDOZA C.E. "Estadística Multivariada". CD2000. [http:// w3.mor.itesm.mx /cmendoza/cd2000/multi](http://w3.mor.itesm.mx/cmendoza/cd2000/multi). 2001.
- PEREZ C. "Técnicas estadísticas con SPSS". Prentice Hall. 2001.
- SALVADOR M.: Análisis de correspondencias. [http:// www.5campus.com/ lección/ anamul](http://www.5campus.com/lección/anamul). 2003.
- SHARMA, S. Applied Multivariate Techniques. John Wiley and Sons.1998
- SPSS V 10.1 "Manual del Usuario" 2001
- VISAUTA, B.: Análisis Estadístico con SPSS para WINDOWS (Vol II. Análisis Multivariante). Mc-Graw Hill.1988

Anexo 1 .- Características de las Variables Obtenidas y Procesadas Estadísticamente al aplicar la metodología.

Experim	Var Independ	Var Transf.	2da Bart KMO	Total Var. Analizar	Dósimas Iniciales	Dósimas Finales
1	0	3	NS	3	66	9
2	7	3	*	10	72	40
3	0	2	NS	2	40	4
4	0	3	NS	3	45	9
5	0	3	NS	3	68	12

6	0	3	NS	3	54	9
7	8	3	*	11	56	22
8	0	2	NS	2	51	6
9	3	3	NS	6	56	24
10	0	2	NS	2	48	6
11	1	3	NS	4	100	16
12	1	3	NS	4	95	20
13	0	3	NS	3	63	9
14	0	3	NS	3	80	12
15	0	3	NS	3	75	9
16	3	3	NS	6	24	12
17	4	3	*	7	76	28
18	6	3	*	9	39	27
19	0	2	NS	2	48	8
20	1	3	NS	4	51	12
21	1	3	NS	4	144	24
22	0	3	NS	3	66	9
23	0	2	NS	2	120	10

Var. Independ. = Variables Independientes .

Var. Transf. = Variables Transformadas (Canónicas)

2da Bart KMO = "2da prueba de Bartlet y KMO

Anexo 2.- Esquema simplificado de la metodología propuesta.

