

A PROPOSAL FOR COMPARATIVE ANALYSIS OF REDUNDANCY IN MULTIBLOCKS OF QUANTITATIVE VARIABLES

UNA PROPUESTA PARA EL ANÁLISIS COMPARATIVO DE LA REDUNDANCIA EN MULTIBLOQUES DE VARIABLES CUANTITATIVAS

JUAN TRABUCCO^{1,a}, JENNYFER COMBARIZA^{2,b}, MAURA VÁSQUEZ^{3,c}

¹DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD METROPOLITANA, CARACAS, VENEZUELA

²POSTGRADO EN ESTADÍSTICA, FACULTAD DE CIENCIAS ECONÓMICAS Y SOCIALES, UNIVERSIDAD CENTRAL DE VENEZUELA, CARACAS, VENEZUELA

³POSTGRADO EN ESTADÍSTICA, FACULTAD DE CIENCIAS ECONÓMICAS Y SOCIALES, UNIVERSIDAD CENTRAL DE VENEZUELA, CARACAS, VENEZUELA

Resumen

Se hace una adaptación del método STATIS, creado para el análisis simétrico de datos cuantitativos de 3 vías, a datos no simétrico cuantitativos de tres vías. Se ilustra con una aplicación a un conjunto de datos basados en un estudio de evaluación de crecimiento y del estado nutricional de jóvenes nadadores venezolanos.

Palabras clave: Índice de redundancia, STATIS, multibloques..

Abstract

An adaptation of the STATIS method was developed, created for the symmetric analysis of three-way quantitative data, to three-way quantitative non-symmetric data. An application was carried out based on a study evaluating the growth and nutritional status of young Venezuelan swimmers.

Key words: Redundancy index, STATIS, multiblock.

1. Introducción

El presente trabajo de investigación se inserta en el análisis multivariante de datos cuantitativos no simétricos, lo que implica explicar la influencia que un conjunto de variables ejercen como determinantes del comportamiento de otro grupo de variables.

Para ello se presta especial atención a una metodología desarrollada en un contexto de análisis multivariante de datos simétrico, denominada STATIS (Lavit, [3]), cuya finalidad es explorar simultáneamente múltiples matrices de datos, bajo ciertas condiciones, mediante el análisis de las similitudes entre estructuras denominadas objetos, que resumen adecuadamente la información relativa a las distancias de los individuos o las interrelaciones de las variables.

^aProfesor Titular. E-mail: jctrabucco@unimet.edu.com

^bPostdoctoral Student. E-mail: jennyfer.combariza@gmail.com

^cProfesora Titular. E-mail: mauralvasquez@gmail.com

Con la adecuada construcción de estructuras que contengan la información estadística de interés en los pares de bloques, en este caso el índice de redundancia de Stewart y Love ([1]- [2]), así como la definición de las similaridades entre estas estructuras, se espera adaptar la metodología STATIS para obtener representaciones que permitan analizar comparativamente múltiples bloques pareados de datos cuantitativos no simétricos.

2. STATIS-R

La propuesta que se presenta tiene como propósito fundamental realizar un análisis exploratorio comparativo de la relación no simétrica existente entre los conjuntos de variables explicativas y sus respectivas contrapartes criterio, representados por los bloques pareados (X_k, Y_k) para $k = 1, \dots, K$ bajo las restricciones que se indicarán.

2.1. Análisis interno de cada bloque con un enfoque de análisis de redundancia

En principio, el análisis no simétrico de cada uno de los K bloques pareados puede ser abordado considerando el análisis de redundancia, cuyo propósito consiste fundamentalmente en analizar la variabilidad de las Y explicada por un conjunto de variables latentes del espacio de las X , utilizando para ello el índice de redundancia, una medida construida como función de un arreglo matricial que contiene información procedente de ambos bloques, en el que se respeta el rol que juegan tanto las variables explicativas como las criterio.

En el contexto del análisis de correlación canónica (ACC), Stewart y Love ([5]) introdujeron el concepto de redundancia, como una medida no simétrica de la relación entre dos bloques de variables. La redundancia se puede interpretar como la media de variabilidad de las variables del bloque Y que es explicada por una variable canónica del bloque X .

Para obtener las combinaciones lineales de las variables X que permiten optimizar el índice de redundancia, denominadas variables de redundancia (ten Berge, [6]), se buscan componentes normalizadas y no correlacionadas $v_j = X\beta_j$ que maximicen el criterio

$$\sum_{j=1}^J \sum_{i=1}^q \text{cor}^2(y_i, v_j)$$

Wollenberg ([8]) propuso un método para maximizar la redundancia, independiente de los cálculos que se obtienen a partir de la correlación canónica, generando los coeficientes de la combinación lineal del bloque X en términos de los cuales se alcanza el máximo. A raíz de esta propuesta se define el índice de redundancia (IR) mediante

$$R_{Y \cdot X}^2 = \frac{1}{q} \sum_{i=1}^q R_{Y_i \cdot X}^2 = \frac{1}{q} \text{tr}(R_{YX} R_{XX}^{-1} R_{XY})$$

que efectivamente maximiza la redundancia del bloque X sobre el bloque Y . Según el tratamiento previo de las variables utilizadas, el índice de redundancia se ofrece a través de distintas versiones.

Por ejemplo, si se utilizan variables centradas pero no estandarizadas, entonces se plantea

$$S_{Y \cdot X}^2 = \frac{\text{tr}(S_{YX} S_{XX}^{-1} S_{XY})}{\text{tr}(S_{YY})}$$

En la práctica, el interés del investigador no se centra en disponer de un número que resuma la fuerza de la relación entre los bloques, como lo hace el índice de redundancia, sino en analizar las razones que explican dicha relación mediante la inspección de las diferentes combinaciones lineales de variables explicativas que ocurren como consecuencia de la búsqueda de la redundancia, así como las correlaciones entre las variables originales y los factores resultantes. Para satisfacer estos objetivos, generalmente se sigue el siguiente procedimiento (véase [4]):

1. Paso 1: preparación de las matrices que conforman los bloques originales. Se trabaja con matrices centradas por columnas, si bien dependiendo de la naturaleza de las mediciones, se aconseja el uso de matrices estandarizadas.
2. Paso 2: en este ítem nos centramos en el análisis de la matriz de las proyecciones mínimo cuadráticas de todas las variables del bloque criterio Y sobre el espacio generado por las variables explicativas contenidas en el bloque X. En general se analiza la descomposición espectral de

$$S = S_{YX} S_{XX}^{-1} S_{XY}$$

ya que el numerador de $S_{Y \cdot X}^2$, contiene toda la información relevante para el cálculo de la redundancia. Se efectúa el cálculo de autovalores y autovectores de S.

3. Paso 3: se decide el escalado a utilizar con los autovectores, dependiendo del interés en construir un biplot de distancias o un biplot de correlaciones. Para realizar una interpretación de las correlaciones, se escalan los autovectores de S de forma tal que cada uno tenga longitud igual a la raíz cuadrada de su autovalor respectivo.
4. Paso 4: se construyen los puntos o scores de los individuos, los descriptores de las variables y medidas de calidad de representación, correlaciones entre variables, etc ...
5. Paso 5: se construye el biplot y se analizan las interrelaciones entre las variables y con los grupos de individuos, considerando los cálculos del paso anterior. Se analizan los ordenamientos de los individuos en cada una de las variables.

2.2. Adaptación de la metodología STATIS para el caso cuantitativo no simétrico

Seguidamente se describen los distintos elementos requeridos para la adaptación del STATIS para el caso cuantitativo no simétrico, que denotaremos por STATIS-R o STATIS de Redundancia.

2.2.1. Etapa 1: estudio dentro del k-ésimo bloque

En el STATIS-R se define un estudio como la tripleta de la forma

$$((X_k, Y_k), M, D) \quad (1)$$

Donde:

1. $\{X_1, \dots, X_K\}$ son K matrices de datos de orden $n \times P_k$.
2. $\{Y_1, \dots, Y_K\}$ son K matrices de datos de orden $n \times Q$.
3. (X_k, Y_k) , $k = 1, \dots, K$, representan K pares de bloques pareados no simétricos, que contiene la medición de las mismas variables cuantitativas sobre n individuos.
4. M es una matriz definida positiva que define la métrica utilizada para construir las distancias entre los individuos.
5. D es una matriz definida positiva, usualmente diagonal, utilizada en el cálculo de las inter-relaciones entre esas variables.
6. Las columnas de las matrices que componen el bloque pareado (X_k, Y_k) , $k = 1, \dots, K$, se denotan respectivamente mediante $X_{1k}, \dots, X_{P_k k}$ e Y_{1k}, \dots, Y_{Qk} para cualquier k.

Para desarrollar una metodología análoga a STATIS que sirva para abordar esta investigación, se requiere identificar con claridad objetos cuya estructura permita analizar las relaciones entre los bloques no simétricos (X_k, Y_k) , $k = 1, \dots, K$.

Definición 1 (Definición de objeto no simétrico). Dado un estudio $((X_k, Y_k), M, D)$, se define el objeto mediante la siguiente expresión

$$W_k = \frac{1}{\sqrt{nQ}} P_{X_k} Y_k = \frac{1}{\sqrt{nQ}} X_k (X_k^t X_k)^{-1} X_k^t Y_k$$

donde $P_{X_k} = X_k (X_k^t X_k)^{-1} X_k^t$ es la matriz de proyección sobre las columnas de la matriz X_k , en este caso particular para cada escenario k se utiliza $M = D_{pk} = (X_k^t X_k)^{-1}$.

El objeto W_k resulta ser una matriz de orden $(n \times P_k) \times (P_k \times P_k) \times (P_k \times n) \times (n \times Q) = n \times Q$; como puede observarse el orden de este arreglo resulta independiente de P_k , el número de variables explicativas, más si depende del número de individuos y del número de variables criterio Q .

Definición 2 (Producto interno de Frobenius). Se define el producto interno de tipo Frobenius entre dos matrices W_h y W_k , de dimensión $n \times P$, de la forma

$$\langle W_h | W_k \rangle_F = \text{traza}(W_h^t W_k). \quad (2)$$

2.2.2. Etapa 2: producto escalar entre objetos de distintos bloques

Definición 3 (Matriz de productos internos S). Se denotará mediante S a la matriz de orden $K \times K$ que contiene los productos internos de Frobenius entre los objetos dos a dos.

$$S = \begin{pmatrix} \langle W_1|W_1 \rangle_F & \cdots & \langle W_1|W_j \rangle_F & \cdots & \langle W_1|W_k \rangle_F \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \langle W_i|W_1 \rangle_F & \cdots & \langle W_i|W_j \rangle_F & \cdots & \langle W_i|W_k \rangle_F \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \langle W_k|W_1 \rangle_F & \cdots & \langle W_k|W_j \rangle_F & \cdots & \langle W_k|W_k \rangle_F \end{pmatrix} \quad (3)$$

Sustituyendo las expresiones que definen a W_h y W_k , el término general queda:

$$S_{hk} = \frac{1}{nQ} \text{traza}(Y_h^t P_{X_h}^t P_{X_k} Y_k) = \frac{1}{nQ} \text{traza}(\tilde{Y}_h^t \tilde{Y}_k) \quad (4)$$

Donde $\tilde{Y}_h = (\tilde{Y}_{(1,h)}, \dots, \tilde{Y}_{(q,h)})$ e $\tilde{Y}_k = (\tilde{Y}_{(1,k)}, \dots, \tilde{Y}_{(q,k)})$ son las proyecciones de las columnas de Y_h y Y_k sobre los respectivos espacios columna X_h y X_k .

Un importante resultado al que conducen los desarrollos anteriores es que sobre la diagonal de la matriz S se encuentra la información fundamental para cada uno de los índices de redundancia correspondientes a cada uno de los K pares de bloques:

$$S_{kk} = \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k^t Y_k \right). \quad (5)$$

La construcción de la matriz S se hace para obtener por un lado un gráfico comparativo del nivel de asociación de los bloques pareados, y por otro lado ser una herramienta que facilita la reconstrucción de las distancias entre los objetos en el espacio euclídeo usual, con una interpretación de interés a los efectos del análisis.

La traza de la matriz S es en sí misma una medida de la asociación contenida entre los K bloques bajo estudio. *El elemento S_{hk} fuera de la diagonal principal de la matriz S , constituye una medida agregada de las covarianzas entre las estimaciones mínimo-cuadrática de las variables criterio en las ocasiones h y k .*

2.2.3. Etapa 3: distancia entre los objetos de distintos bloques

Definición 4 (Distancia entre objetos). Se define una distancia entre objetos basada en la norma dada de forma natural por el producto interno de Frobenius:

$$\begin{aligned} d_F^2(W_k, W_h) &= \|W_k - W_h\|_F^2 = \langle W_k - W_h | W_k - W_h \rangle_F \\ &= \|W_k\|_F^2 + \|W_h\|_F^2 - 2\langle W_h | W_k \rangle_F. \end{aligned}$$

Donde el último término es una medida de la covarianza entre las estimaciones de y en ambos bloques. De modo que mientras mayor sea está covariabilidad, menor será la distancia entre las representaciones de los bloques. Es decir, es posible cuantificar en una medida de distancia F , las diferencias existentes entre las estructuras definidas por las estimaciones mínimo-cuadrática del

bloque de las Q variables criterio que son obtenidas en función de los correspondientes bloques de variables explicativas.

En resumen, la medida de distancia propuesta para comparar los objetos de interés será mayor cuanto más grandes sean las respectivas medidas de redundancia en las dos ocasiones consideradas y tanto más pequeña cuanto mayor sea la medida agregada de las covarianzas entre las estimaciones mínimo-cuadrática de la variable criterio en las ocasiones h y k .

2.2.4. Etapa 4: espacio de representación de la interestructura

La matriz que define las representaciones en el STATIS es la matriz S que contienen los productos escalares de Frobenius entre objetos de los distintos bloques.

La representación deseada se obtendrá al efectuar un análisis de componentes principales sobre la matriz de productos interno de interés, en este caso \mathbf{S} , con la finalidad de construir un espacio de representación óptimo para los objetos \mathbf{W}_k , determinado por las direcciones principales que definen la medida de redundancia total:

$$\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k Y_k \right) \quad (6)$$

De esta manera el posicionamiento de un bloque pareado en ese espacio, determinado por su distancia al origen de coordenadas, deberá ser medida por una variación del índice de redundancia en cuestión

$$\|W_k\|_F^2 = \frac{1}{nQ} \text{traza} (W_h^t W_k) = \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k Y_k \right). \quad (7)$$

Los resultados deseados se consiguen al efectuar la descomposición espectral de \mathbf{S} , descrita mediante la expresión $\mathbf{S} = \mathbf{G}\mathbf{T}\mathbf{G}^t$, siendo $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_K)$ la matriz cuyas columnas son los autovectores de la matriz \mathbf{S} asociados con los correspondientes autovalores ordenados de forma descendente y organizados sobre la diagonal de la matriz

$$\begin{pmatrix} t_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_k \end{pmatrix} \quad (8)$$

Un resultado fundamental que se deriva de esta factorización, es la posibilidad de descomponer la redundancia global en la forma:

$$\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k Y_k \right) = \sum_{k=1}^K t_k. \quad (9)$$

Es decir, al diagonalizar la matriz S (3) se obtiene que la traza, es una medida agregada del índice de redundancia en cada ocasión, donde cada autovalor está asociado con una dirección principal del análisis del índice de redundancia.

La asociación de los objetos capturada por el k -ésimo eje se mide mediante la relación

$$\frac{t_k}{\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k Y_k \right)} \quad (10)$$

En este orden de ideas, las coordenadas de proyección γ^{α_k} , de los bloques sobre el eje a^{α_k} quedan de la forma

$$\gamma^{\alpha_k} = \sqrt{\lambda_{\alpha_k}} a^{\alpha_k}, \quad \alpha_k, k = 1, \dots, K, \quad (11)$$

En la representación gráfica se tendrá que en los bloques más alejados del origen, la capacidad de Y_k para explicar Y_k será mayor (mayor será el índice $\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k^t Y_k \right)$).

2.2.5. Estrategia denominada ICI

En el caso de esta investigación en particular se pretende comparar la variabilidad de las variables X_K en función de las matrices Y_k , $k = 1, \dots, K$, en términos del índice de redundancia $\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k^t Y_k \right)$. Para ello se procederá a construir un marco semejante al de la metodología STATIS, que permita extender el análisis a objetos que contengan toda la información necesaria para reconstruir el índice de redundancia $\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k^t Y_k \right)$ en cada bloque no simétrico (X_k, Y_k) , $k = 1, \dots, K$. Así las cosas, se propone un símil de la estrategia denominada ICI (Interestructura - Compromiso - Intraestructura) para este caso en particular. Esto quiere decir que se lleva a cabo un procedimiento de tres fases que responde a los objetivos de:

1. Identificar los bloques de información (X_k, Y_k) , $k = 1, \dots, K$, que son similares (interestructura).
2. Construir un marco de representación común a todos los bloques de información (X_k, Y_k) .
3. Describir las diferencias o similitudes entre los individuos sobre un espacio compromiso (intraestructura). Utilizando el compromiso como pieza fundamental se va a analizar el comportamiento de los individuos y los bloques que de alguna manera expliquen las razones o causas de las semejanzas y/o diferencias detectadas en el análisis de la interestructura.

Cabe destacar, que un objeto definido de acuerdo a los principios de la metodología STATIS, sería una matriz cuadrada simétrica, con características que resultan beneficiosas en algunos aspectos. Sin embargo, los objetos definidos en esta investigación resultan ser matrices rectangulares no simétricas. Sobre la base del producto interno de Frobenius, y con medidas de asociación y correlación matricial establecidas bajo los principios que orientan este concepto, se construyen matrices de similaridad entre objetos que aceptan representaciones gráficas sobre espacios de baja dimensión, las cuales permiten apreciar gráficamente el parecido y/o diferencia en relación con el índice de redundancia $\sum_{k=1}^K \frac{1}{Q} \text{traza} \left(Y_k^t X_k \frac{(X_k^t X_k)^{-1}}{n} X_k^t Y_k \right)$, entre los distintos bloques no simétricos.

3. Aplicación del STATIS-R

La muestra utilizada para ilustrar esta propuesta de representación gráfica, forma parte de un estudio de evaluación de crecimiento y del estado nutricional de jóvenes nadadores venezolanos, llevado a cabo en el año 2004. Dicho estudio se enmarca en un proyecto de investigación en colaboración realizado por la Universidad Central de Venezuela y Fundacredesa titulado Crecimiento, estado nutricional, evaluación de la ingesta alimentaria y de condiciones socioeconómicas de los

jóvenes nadadores venezolanos". Los datos fueron amablemente cedidos para su uso en esta investigación por la Profesora Betty M. Pérez, del Instituto de Investigaciones Económicas y Sociales, FACES de la Universidad Central de Venezuela.

El diseño original, de tipo transversal, se apoyó en mediciones realizadas sobre cien niños y adolescentes, nadadores masculinos de edades comprendidas entre 7 y 17 años, quienes integraban parte de la selección del estado Miranda, Venezuela. Estos jóvenes atletas además fueron clasificados según las categorías de maduración sexual de los genitales en: pre-púberes, púberes-iniciales y púberes-avanzados, siguiendo criterios reconocidos internacionalmente. Los grupos estuvieron integrados por 20, 53 y 27 individuos respectivamente.

3.1. Variables analizadas

Todos los participantes fueron evaluados en las 35 variables antropométricas que se proponen en el Manual ISAK (siglas de International Society for the Advance of the Kineantropometry). 29 de estas variables fueron seleccionadas y clasificadas de acuerdo a su naturaleza, en cuatro bloques de interés, que se detallan a continuación:

- El Bloque X1 conformado por 7 variables que miden pliegues cutáneos (en milímetros) para estimar la dimensión Grasa Subcutánea: Subescapular (gsub), Biceps (gbic), Cresta ilíaca (gcre), Supraespinal (gsup), Abdominal (gabd), Muslo frontal (gmus), Pantorrilla medial (gpan). Para referirnos de forma abreviada al concepto subyacente en las variables componentes de este bloque, denominaremos Bloque GRA, objeto GRA, etc.
- El Bloque X2 contempla 7 variables que definen la Robusticidad Esquelética en su dimensión Longitudinal, medidas en centímetros: Brazo (lbra), Antebrazo (lant), Mano (lman), Trochanterion tibial lateral (lttl), Tibial lateral (ltla), Tibial lateral esfriotibial (ltle), Pie (lpie). La abreviatura para este bloque es LON.
- El Bloque X3 representa la Robusticidad Esquelética en su dimensión Diametral, en términos de las 7 variables que se presentan a continuación, medidas en centímetros: Biacromial (dbia), Biliocristal (dbiil), Tórax transverso (dttr), Tórax profundidad anterior-posterior (dtap), Húmero (dhum), Muñeca (dmuñ), Fémur (dfem). La abreviatura para este bloque será DIA.
- El Bloque X4 representa la dimensión Musculatura Relativa, descrita en términos de las 8 variables que se señalan a continuación: Brazo flexionado y tenso (pbrf), Antebrazo (pant), Muñeca (pmuñ), Tórax (ptor), Cintura (pcin), Cadera (pcad), Pantorrilla (ppan), Tobillo (ptob). La abreviatura para este bloque será PER.

Además, se utilizaron los tres indicadores del estado nutricional: Índice de Masa Corporal (IMC), Área Grasa (AG), Área Muscular (AM), definidos a continuación:

- Índice de Masa Corporal,

$$IMC = \frac{Tpes}{Test^2}$$

siendo Tpes el Peso y Test la Estatura;

- Área Grasa,

$$AG = \frac{(Pbrr - Gtrip)^2}{4\pi}$$

donde P_{brr} representa el perímetro del brazo relajado, y G_{tri} representa el pliegue cutáneo del tríceps;

- Área Muscular, $AM=AT-AG$, donde AT representa el Área Total del brazo y se calcula mediante

$$AT = \frac{(P_{brr}^2)}{4\pi}$$

Estas variables constituirán el Bloque Y, el cual de ser necesario se referenciará como bloque NUT (estado nutricional).

3.2. Descripción de la muestra analizada

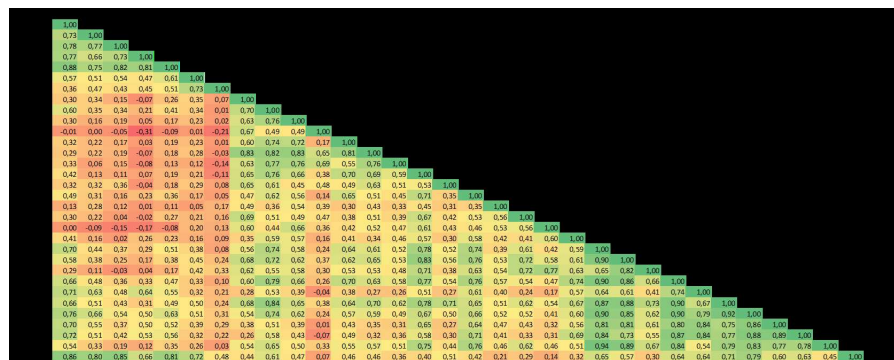
En principio, los investigadores deseaban conocer la intensidad de la fuerza con la que las variables de cada bloque X pueden explicar a las variables del bloque de indicadores nutricionales NUT de forma conjunta, y a cuáles de las variables de este bloque explican mejor.

Para simplificar algunos dispositivos gráficos, se optó por analizar el grupo de los nadadores púberes avanzados, constituido por 27 individuos. No obstante, todas estas técnicas permiten manejar números superiores de filas.

Indagando sobre las relaciones de las variables dos a dos, se observan las asociaciones que se esperarían en relación a las variables del bloque GRA con la variable AG, o de las variables del bloque PER con las variables IMC y AM. No obstante esta revisión es tediosa, por lo que se sugiere el uso de un dispositivo grafico que permita iniciar la exploración de las relaciones entre los diferentes bloques.

3.3. Resultados obtenidos

FIGURA 1: Matriz de correlación.



Al realizar análisis de redundancia individualizados, para los cuatro pares de bloques (GRA y NUT, LON y NUT, DIA y NUT, PER y NUT) se aprecian los siguientes resultados:

La inercia restringida (variabilidad total explicada por el bloque Y) para cada bloque resultado de moderada a alta (GRA: 69.6 %, LON: 50.7 %, DIA: 60.1 %, PER: 85.1 %). En cada caso, se cuenta con un primer eje que representa entre 79 % y 87 % de la variabilidad de las variables de NUT explicada por cada uno de los bloques explicativos (véase figura 2). Al realizar análisis de redundancia individualizados, para los cuatro pares de bloques (GRA y NUT, LON y NUT, DIA

The figure consists of four RDA biplots arranged in a 2x2 grid, each showing the relationship between environmental variables (vectors) and species distribution (points) for a specific morphological trait.

- Top Left: Mapa RDA - GRASAS (ejes F1 y F2: 95.39 %)**
 - X-axis: F1 (79.61 %)
 - Y-axis: F2 (15.79 %)
 - Species: gcre, gcrs, gcrd, gsub, gcrs, gcrs.
 - Environmental variables: AG, IMC, AM.
- Top Right: Mapa RDA - LONGITUDES (ejes F1 y F2: 95.26 %)**
 - X-axis: F1 (87.53 %)
 - Y-axis: F2 (7.73 %)
 - Species: lttl, lbpa, lpla, lltl, lltl, lltl, lltl, lltl.
 - Environmental variables: AM, AG, IMC.
- Bottom Left: Mapa RDA - DIAMETROS (ejes F1 y F2: 93.81 %)**
 - X-axis: F1 (86.82 %)
 - Y-axis: F2 (6.99 %)
 - Species: dbill, dhum, dbia, dtrr, dfem, dtrr.
 - Environmental variables: AG, AM, IMC.
- Bottom Right: Mapa RDA - PERIMETROS (ejes F1 y F2: 94.77 %)**
 - X-axis: F1 (85.20 %)
 - Y-axis: F2 (9.57 %)
 - Species: pcin, pcin, pcin, pcin, pcin, pcin, pcin, pcin.
 - Environmental variables: AG, IMC, AM.

La inercia restringida (variabilidad total explicada por el bloque Y) para cada bloque resulto de moderada a alta (GRA: 69.6 %, LON: 50.7 %, DIA: 60.1 %, PER: 85.1 %). En cada caso, se cuenta con un primer eje que representa entre 79 % y 87 % de la variabilidad de las variables de NUT explicada por cada uno de los bloques explicativos.

Las posiciones relativas de las variables de interés, IMC, AG y AM, no permiten una comparación simultanea mediante la cual ubicar las capacidades explicativas de los diferentes bloques X (GRA, LON, DIA, PER).

XXX Simposio de Estadística (2021)

3.3.1. Estudio de la Interestructura

La elaboración de los objetos de interés corresponde a VGRA, VLON, VDIA y VPER, los cuales contienen por construcción la información inherente a un análisis de redundancia (véase figura ??). Al utilizar el producto escalar de Frobenius, especialmente adaptado para que se reconstruya la

FIGURA 3: Matriz de correlación.

VGRA	IMC	AM	AG	VLON	IMC	AM	AG	VDIA	IMC	AM	AG	VPER	IMC	AM	AG
1	-0.4471	-0.7680	-3.7978	1	0.7465	-2.1579	-2.7985	1	0.0243	-2.8926	-2.3688	1	-0.1268	-2.9119	-2.7761
2	-1.4951	-10.1191	-1.3618	2	-1.4839	-10.7727	-2.3627	2	-2.8306	-19.9890	-2.1681	2	-3.3423	-16.2719	-1.2930
3	-0.0756	-3.4855	0.6926	3	1.6980	-0.4340	0.5676	3	1.6737	5.7585	2.4130	3	2.4008	5.5149	0.8081
4	0.3958	7.3005	-2.1511	4	0.6587	10.1651	1.8075	4	0.1001	14.5492	1.5869	4	0.4042	11.7980	0.7468
5	-0.7148	-9.3891	-0.2380	5	-0.7001	-2.5789	-1.6196	5	-0.3810	0.7715	1.0080	5	-1.1322	-3.4904	-0.0526
6	-0.0143	6.6161	6.0570	6	1.7208	9.8081	5.4679	6	0.3897	8.7637	1.5514	6	0.8305	11.6007	4.5432
7	1.9411	6.8701	10.1548	7	-0.4562	1.1788	2.1560	7	-1.3985	1.4666	2.1903	7	1.0911	5.7028	5.6267
8	1.7778	10.3172	-1.3349	8	1.0990	0.1815	2.2991	8	0.4378	0.6996	-0.0985	8	0.1010	4.4698	-3.2646
9	-1.0760	-2.8583	-3.0263	9	0.1031	0.1356	0.2332	9	0.8086	2.2638	0.2951	9	-0.9947	-2.6840	-0.6339
10	-1.7457	-3.1350	-4.2319	10	-0.6538	-4.7861	-3.3228	10	0.6790	-1.6295	-0.2820	10	-1.6122	-6.1707	-2.9993
11	-1.9345	-9.5197	-2.4628	11	-1.6690	-1.0871	-1.3502	11	-0.5621	-5.6165	-0.8340	11	0.0076	-5.3259	0.4157
12	0.6081	3.9529	2.9698	12	0.2442	-1.2842	-0.6056	12	-0.1537	1.1468	1.9378	12	-0.8723	-0.2928	-1.5307
13	-1.5044	-6.3561	-0.1882	13	-2.6412	-7.4235	-3.9526	13	-1.8279	-13.0274	-1.2829	13	-2.1482	-13.4142	-1.3685
14	0.0891	-0.0600	-0.8331	14	-0.1130	4.4072	-0.6458	14	1.6044	9.8828	1.9029	14	-0.7688	-4.2843	0.3330
15	0.3645	5.9838	-0.9128	15	-2.4467	-2.1200	-3.6400	15	1.3387	6.1878	0.0911	15	0.3881	8.6910	-1.5418
16	1.7926	10.9362	2.3492	16	1.6401	3.7117	3.4153	16	1.7468	8.9953	-1.0003	16	3.6414	18.2613	4.2732
17	-0.7291	-3.6209	-3.5472	17	-0.0058	3.2125	0.1488	17	0.1044	2.3690	0.5166	17	-1.1174	-6.3323	-3.1627
18	0.3346	4.6936	0.8258	18	0.4414	3.7678	-0.1313	18	0.9056	5.2905	1.7062	18	0.7984	11.1702	-0.3096
19	-4.7710	16.5962	9.4235	19	5.2654	22.7018	8.6642	19	5.6823	16.6153	7.1626	19	5.8180	20.2092	10.1038
20	-0.8991	-0.6697	-2.4151	20	-0.2012	2.3321	0.0960	20	0.7775	5.1316	1.3596	20	2.4557	7.8014	1.5421
21	-1.7026	-5.6319	-3.2690	21	-0.7066	-4.7413	-0.6743	21	-2.0471	-8.3987	-1.8122	21	-2.2735	-14.3251	-1.9997
22	-1.7972	-5.9343	-1.5738	22	0.0820	0.6093	0.6241	22	-3.1184	-13.1440	-4.6471	22	-2.9417	-14.6687	-6.5498
23	3.0853	-1.0962	3.2712	23	1.3214	-5.1160	1.2266	23	1.5858	0.0280	-1.5126	23	2.9021	-5.6179	2.0139
24	1.5209	8.6737	0.3810	24	-0.0851	-0.9043	1.8669	24	0.1169	-1.1817	0.7304	24	0.1716	0.6116	-0.0156
25	0.8983	4.3838	-0.2214	25	-1.1850	-4.0178	-3.3404	25	-1.5524	-7.0864	-0.2162	25	-0.4489	1.6706	0.9192
26	-1.2533	-17.2814	-1.2624	26	-1.3651	-12.6419	-1.7374	26	-1.6529	-8.2674	-4.8693	26	-0.9667	-5.5523	-2.3206
27	-2.1901	-6.4188	-3.2973	27	-1.3079	-2.1459	-2.0942	27	-2.4509	-8.0869	-3.1399	27	-2.2647	-6.1593	-1.4074

redundancia en la matriz S, se obtiene:

$$S = \begin{pmatrix} 24,9 & 15,22 & 17,08 & 23,41 \\ 15,22 & 18,53 & 17,06 & 18,39 \\ 17,08 & 17,06 & 26,62 & 25,5 \\ 23,41 & 18,39 & 25,5 & 35,52 \end{pmatrix}$$

La descomposición de esta matriz permite obtener la información relevante para ubicar en un plano de representación a los objetos GRA, LON, DIA y PER a fin de compararlos y determinar sus semejanzas y diferencias (véase figura 4).

La medida de redundancia total ($MRT = 105.6$), obtenida como la suma de los elementos de la diagonal de S, está repartida en cada uno de los ejes que se construyen con los autovectores de dicha matriz. Los dos primeros autovalores de la descomposición de la matriz S indican que el plano refleja el 90.3 % de la medida total de redundancia capturada en los cuatro bloques pareados, lo que puede considerarse una buena representatividad de la redundancia total.

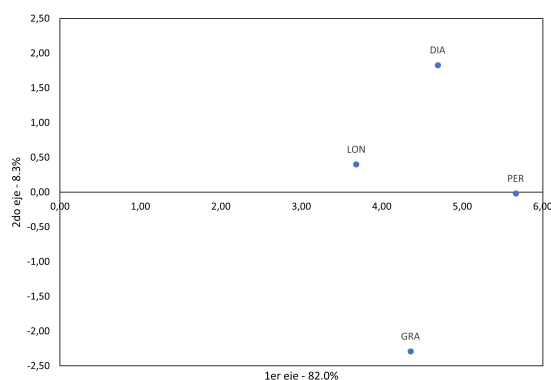
Los siguientes vectores columna permiten representar los objetos en un primer plano:

$$\begin{array}{ll} OBJ & 1 \quad 2 \\ GRA & 4.35 \quad -2.29 \\ LON & 3.68 \quad 0.40 \\ DIA & 4.69 \quad 1.83 \\ PER & 5.66 \quad -0.02 \end{array} \quad (12)$$

Particularmente, el bloque PER es responsable de un 37 % de la definición del primer eje, en tanto que el segundo eje está definido prácticamente por los bloques GRA (60 %) y DIA (38 %), que apuntan hacia cuadrantes diferentes. El bloque LON no es fundamental en la definición de los ejes (16 % y 2 % respectivamente).

Referente a la calidad de representación de los bloques en el biplot, se encuentran muy bien

FIGURA 4: Representación de los bloques pareados.



representados los bloques GRA (97 %), DIA (95 %) y PER (90 %), y razonablemente el bloque LON (74 %). En esencia, la información de cada uno de los cuatro bloques pareados esta suficientemente representada en el diagrama de representación de las redundancias, si bien los bloques DIA y GRA están contrapuestos en el segundo eje, mediados por los bloques PER y LON.

Como estos puntos están representando la información inherente a la redundancia de cada bloque pareado, es decir, a una medida de la capacidad explicativa global de cada bloque X en relación al bloque Y, podemos interpretar la información como que aunque las variables relacionadas a grasa (GRA) y diámetro (DIA) tienen una capacidad explicativa similar hacia el bloque Y, lo hacen en forma diferente. Las variables relacionadas a perímetros (PER) son mas explicativas globalmente, y definen el primer eje.

3.3.2. Elaboración del Objeto Compromiso

Utilizando la información del primer autovector de la descomposición de S, se obtiene el compromiso de los cuatro objetos originales.

Este compromiso se puede interpretar como un promedio ponderado de la información de redundancia de los bloques GRA, LON, DIA y PER. A continuación, se utiliza esta información resumida para construir el análisis de intraestructura.

La revisión de la descomposición espectral del objeto compromiso permite garantizar que con las dos primeras direcciones se capturara el 93 % de la variabilidad contenida en el mismo. El gráfico (5) refleja en una sola representación, toda la información de la interrelación entre las variables, pero tomando en consideración la estructura de redundancia de las variables involucradas en los diferentes bloques hacia las variables IMC, AM y AG.

Como en todos los biplots de representación de variables o de distancias, se debe analizar la información completa, considerando la calidad de representación. Por ejemplo, la variable dbiil apunta en la dirección de IMC en el diagrama de los dos primeros ejes principales, aunque la correlación (véase figura 1) entre estas variables es pobre (0.30). Sin embargo, una inspección del tercer eje permite apreciar un ángulo importante entre ellas.

FIGURA 5: Espacio compromiso obtenido con la metodología STATIS-R.



4. Conclusiones

1. Se propusieron arreglos matriciales definidos como los objetos de análisis, análogos a los establecidos en la metodología STATIS, que fundamentalmente incorporan la información necesaria para reproducir la medida de redundancia de un bloque explicativo X sobre un bloque criterio Y.
2. Se obtuvieron medidas de similitud orientadas a la comparación de los objetos de análisis, basadas en un producto escalar definido de forma conveniente para reproducir la medida de redundancia en cada bloque pareado, y medir las relaciones entre la información contenida en dos bloques pareados.
3. Se construyeron espacios óptimos de representación que permiten simbolizar gráficamente la información contenida en cada objeto, de acuerdo a las similitudes que existan entre ellos en función de las medidas definidas.
4. Se conceptualizaron las distancias estadísticas, basadas en las medidas de similitud, que permiten la interpretación de las distancias entre los objetos, observables en la representación gráfica en un espacio euclídeo de baja dimensión.
5. Se ubicaron escenarios en los cuales es posible esbozar las razones que expliquen los parecidos y/o diferencias detectados entre los objetos con las medidas estadísticas definidas.

Referencias

- [1] Israels, A. Z. (1984). Redundancy analysis for qualitative variables. *Psychometrika*, 49, 331-346.

- [2] Israels, A. Z. (1987). Eigenvalue techniques for qualitative data. DSWO Press, Leiden.
- [3] Lavit, C. (1985). Application de la méthode statis. Statistique et Analyse des Données, Vol 10, n.1.
- [4] Legendre, P., Legendre, L. (1998). Numerical Ecology. 2nd ed. Elsevier.
- [5] Stewart, D., Love, W. (1968). A general canonical correlation index. Psychological Bulletin, 70, 160-163.
- [6] ten Berge, J. M. F. (1985). On the relationship between Fortier's simultaneous linear prediction and Van den Wollenberg's redundancy analysis. Psychometrika, 50, 121 - 122.
- [7] ter Braak, C. (1994). Canonical community ordination. Part I: Basic theory and linear methods. Ecoscience, vol 1, n.2, 127-140
- [8] Wollenberg A. L. (1977). Redundancy analysis. An alternative for canonical correlation analysis. Psychometrika, Vol. 42, no. 2.