

IMPLEMENTACION DEL MÉTODO STATIS EN R Y SU APLICACIÓN AL ESTUDIO DE SATISFACCIÓN DE USUARIOS DEL SISTEMA BIBLIOTECARIO DE LA BUAP

L. Sandoval Solís*, G. Linares Fleites **, A. Quiroz Clemente*, M.G. Romero Vázquez***

*Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla.

**Instituto de Ciencias. DICA , Benemérita Universidad Autónoma de Puebla.

***Dirección General de las Bibliotecas , Benemérita Universidad Autónoma de Puebla.

ABSTRACT

This paper presents the application of Dual STATIS method to surveys of “Customer Satisfaction” which holds twice a year by Libraries Department of the Autonomous University of Puebla, Puebla, Mexico. In particular, we show the results of the application, in November 2009, to three libraries and also we present the behavior of one of these libraries at different dates of implementation of these surveys. These surveys generate multiple tables of data and we are interested to define framework simultaneously between different libraries and to identify similarities and differences between them. To achieve this objective it was necessary to deepen the STATIS method and to develop a program R to perform the said application.

RESUMEN

En este trabajo se presenta la aplicación del método STATIS Dual a las encuestas de “Satisfacción de Usuarios” que lleva a cabo dos veces al año la Dirección General de Bibliotecas de la Benemérita Universidad Autónoma de Puebla, Puebla, México. En particular, se muestran los resultados de la aplicación realizada en noviembre de 2009 a tres bibliotecas de área y, además, se presenta el comportamiento de una de esas bibliotecas en diferentes fechas de aplicación de estas encuestas. Estas encuestas generan múltiples tablas de datos y es de interés definir un marco de referencia simultáneo entre las diferentes bibliotecas y determinar las semejanzas y diferencias entre ellas. Para alcanzar ese objetivo fue necesario profundizar en el método STATIS y desarrollar un programa R para llevar a cabo la referida aplicación.

KEY WORDS: Data Analysis, reliability, correlation

MSC: 62P99

1. INTRODUCCIÓN

Cuando se dispone de información sobre un conjunto de individuos y esta información se puede estructurar en diferentes grupos de variables, es posible plantearse diversas preguntas al realizar un estudio simultáneo de estos grupos, entre otras: ¿tendrán más influencia unos grupos que otros?, ¿cómo equilibrar esta influencia?, ¿definen estos grupos diferentes estructuras sobre el mismo conjunto de individuos?, ¿se pueden comparar estas estructuras? Estas preguntas surgen de la propia estructuración de los datos y el método STATIS, aborda el estudio de información de esta naturaleza. Su objetivo es estudiar la estructura común de las distintas tablas de datos, indicando la trayectoria de los individuos homogéneos, así como de los elementos heterogéneos, es decir, aquéllos que tienen un comportamiento diferente al resto.

En este trabajo se presenta la aplicación del método STATIS Dual a las encuestas de “Satisfacción de Usuarios” aplicadas por la Dirección General de Bibliotecas de la BUAP; en particular, se presentan los resultados de la aplicación realizada en noviembre de 2009 a las bibliotecas de área y, además, se presenta el comportamiento de una de esas bibliotecas en diferentes fechas de aplicación de estas encuestas. De lo anterior puede apreciarse que se han generado múltiples tablas y se persigue el objetivo de definir un marco de referencia simultáneo y determinar las semejanzas y diferencias entre ellas. Para alcanzar ese objetivo fue necesario profundizar en el método STATIS y desarrollar un programa R para llevar a cabo la referida aplicación.

2. EL MÉTODO STATIS

STATIS (Structuration des tableaux a trois indices de la statistique) es un método exploratorio de análisis de datos cuantitativos, desarrollado por Christine Lavit, [7, 8] que permite analizar arreglos de datos de manera simultánea. Sus bases matemáticas se desarrollaron por Yves Escoufier [1, 2, 3, 4], usando el álgebra lineal y la distancia euclidiana entre configuraciones de puntos. El método STATIS permite determinar qué arreglos tienen características en común, construir un arreglo resumen, describir las diferencias entre los arreglos y determinar las causas de las diferencias, ya sean debido a los individuos o a las variables [6]. El método se basa en el Análisis de Componentes Principales, técnica de uso común en análisis exploratorios de datos [9, 10].

Para poder utilizar esta metodología, los datos de interés deben recolectarse en situaciones específicas. Las dos situaciones más comunes son: (a) se recolectan medidas de los mismos individuos en diferentes ocasiones formando un arreglo para cada ocasión y, (b) las mismas variables se miden en grupos distintos de individuos.

Estas situaciones dan origen a dos versiones del método STATIS. En la primera los individuos son los mismos en cada tabla y las variables difieren en cada una de ellas. En la segunda versión, también conocido como el Método STATIS Dual, las variables son las mismas en cada tabla y los individuos varían. En el estudio de satisfacción de usuarios de las bibliotecas de la BUAP se presenta este caso.

Una vez definida la estrategia que se desea utilizar, es necesario determinar el peso que se le asignará a cada individuo (o variable) y la forma en que se medirá la distancia entre estos. Denotaremos por X a un arreglo de datos formado por n líneas y m columnas, donde n es el número de individuos, m el número de variables y p el número de arreglos. Al espacio vectorial de las variables lo denotaremos E , con dimensión igual a n y d es el producto escalar del espacio dual E^* de E definido a partir de los pesos de los individuos. F denotará al espacio vectorial de los individuos. Su dimensión equivale a m y su producto escalar q se utiliza para calcular las distancias entre individuos.

La terna $(X; q; d)$, formada por un arreglo de datos y los productos escalares q y d que se eligen para analizar X , se conoce como *estudio*.

A continuación se brindan los elementos fundamentales de esta metodología.

- **Intraestructura**

La idea esencial del método STATIS es encontrar una estructura común dentro de los *estudios*, conocida como *intraestructura*. De acuerdo a los intereses, buscamos una intraestructura de los individuos o de las variables. Si buscamos un análisis de individuos (situación 1), la intraestructura equivale a las distancias que separan a los individuos dentro de cada uno de los arreglos.

- **Interestructura**

Una vez determinada la intraestructura de cada uno de los p arreglos, nos interesa conocer si la estructura de individuos (o variables) es común en los estudios, es decir, si las distancias entre individuos (o variables) son estables de un estudio a otro. De esta forma es posible conocer la relación existente entre los arreglos. A esta relación se le conoce como *interestructura*. El producto escalar de Hilbert -Schmidt entre transformaciones lineales se utiliza para este fin.

- **Imagen Euclidiana**

A partir de la interestructura se construye una Imagen Euclidiana de los arreglos.

- **Compromiso**

A partir de la imagen euclidiana de los estudios, podemos construir un objeto compromiso W considerado como un arreglo de productos escalares medios entre los individuos

- **Trayectorias de los individuos:**

Las trayectorias de los individuos nos permiten describir de forma individual las desviaciones de los individuos (variables) entre ellos mismos y con el compromiso. Además, es posible conocer la evolución que han sufrido los individuos (variables) de un estudio a otro. Estas desviaciones se obtienen con la interestructura y la imagen euclidiana de los estudios. La trayectoria de cada individuo (variable) se construye, en una imagen euclidiana, al unir los puntos que representan a dicho individuo (variable) para cada estudio.

A partir de la tesis de Geyne, 2008 [5], se estructuró el Diagrama del Método STATIS (Anexo 1) y se elaboró un programa R (Anexo 2) lo que ha permitido profundizar en los estudios de satisfacción de usuarios en las Bibliotecas de la Benemérita Universidad Autónoma de Puebla, México.

3. APLICACIÓN

La Dirección de Bibliotecas de la Benemérita Universidad Autónoma de Puebla está trabajando en la mejora continua de los servicios que brinda a la comunidad universitaria. Para estudiar la satisfacción de los usuarios de sus bibliotecas en las áreas de ciencias naturales y exactas (Biblioteca Niels Bohr), ingenierías (Biblioteca Ing. L. Barragán) y ciencias sociales (Biblioteca J. Revueltas) ha elaborado un cuestionario. (Ver Anexo 3)

El estudio de confiabilidad y validez llevado a cabo a este cuestionario que utilizan las Bibliotecas de la Universidad Autónoma de Puebla para el estudio de la calidad de los servicios bibliotecarios ha mostrado que el instrumento es adecuado y útil para la investigación sistemática y continua que permita alcanzar la excelencia de los servicios bibliotecarios en la Universidad [11]. Dentro de la norma ISO 9001-2008 se requiere efectuar cada seis meses encuestas de satisfacción de usuarios y por ello se desarrolló el Sistema de Encuestas Bibliotecario, en PHP y MySQL, que permite aplicar las encuestas simultáneamente e incluir en las bases de datos sólo las encuestas donde se contestan todos sus reactivos.

En el Anexo 4 se muestra el comportamiento que ha tenido la Biblioteca Luis Barragán en el período 2007-2009. Puede apreciarse en el gráfico “Imagen Euclidiana” que la aplicación de la encuesta realizada en Mayo 2009 (azul) mostró resultados que la alejan de las aplicaciones de la encuesta en las demás fechas. En el gráfico “Correlaciones entre las variables” se aprecia menos variabilidad en las respuestas al cuestionario en esa aplicación. En el gráfico de “Trayectorias” las preguntas que se refieren al servicio bibliotecario (préstamo, renovación, devolución y servicio de consulta) están en la parte superior del gráfico, lo que señala la importancia de estos aspectos en esa aplicación.

El gráfico de la “Imagen Euclidiana”, en el Anexo 5, muestra que las Bibliotecas Niels y Revuelta se comportaron, en la aplicación realizada en Noviembre de 2009, de manera diferente a las restantes bibliotecas. En el gráfico de “Trayectorias” se corrobora que las Bibliotecas Niels (azul) y Barragán (rojo) son las más distantes entre sí y, el gráfico “Correlaciones entre variables” revela que la biblioteca Barragán muestra la menor variabilidad.

4. CONCLUSIONES

STATIS permite el análisis simultáneo de diferentes tablas numéricas referidas a las mismas o a distintas variables y a un determinado conjunto de individuos. Compara las nubes de **individuos** a través de las **matrices de productos escalares**, teniendo mayor influencia aquéllas con un comportamiento similar al mayoritario. STATIS no permite saber cuáles son las variables que intervienen en la formación de los ejes de variabilidad.

El uso de esta técnica exploratoria le ha permitido a la Dirección de Bibliotecas de la BUAP detectar algunos de los problemas que se han presentado en sus servicios, y le ha ayudado a dirigir sus esfuerzos para lograr una mayor satisfacción de sus usuarios.

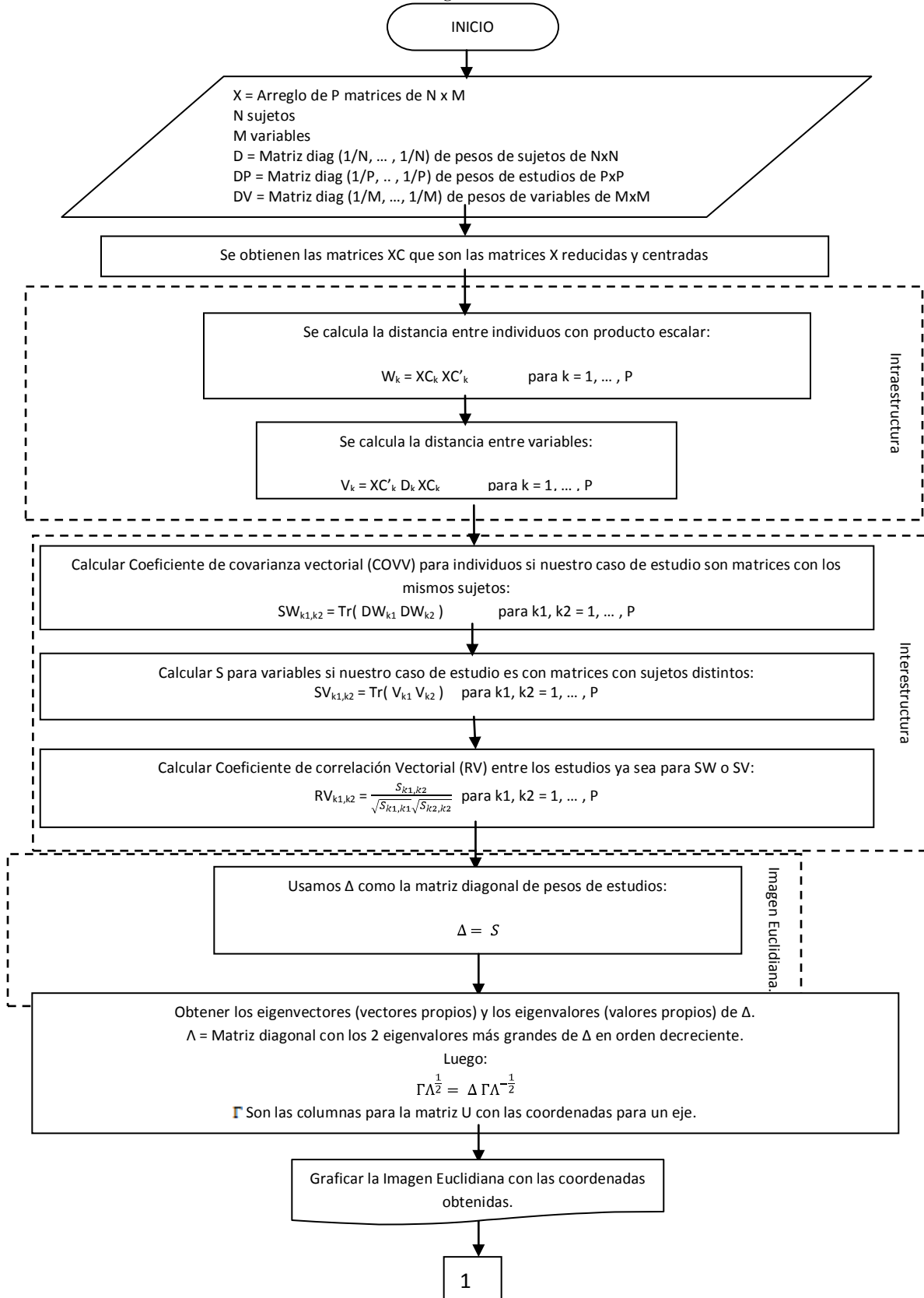
RECEIVED APRIL 2010
REVISED APRIL 2011

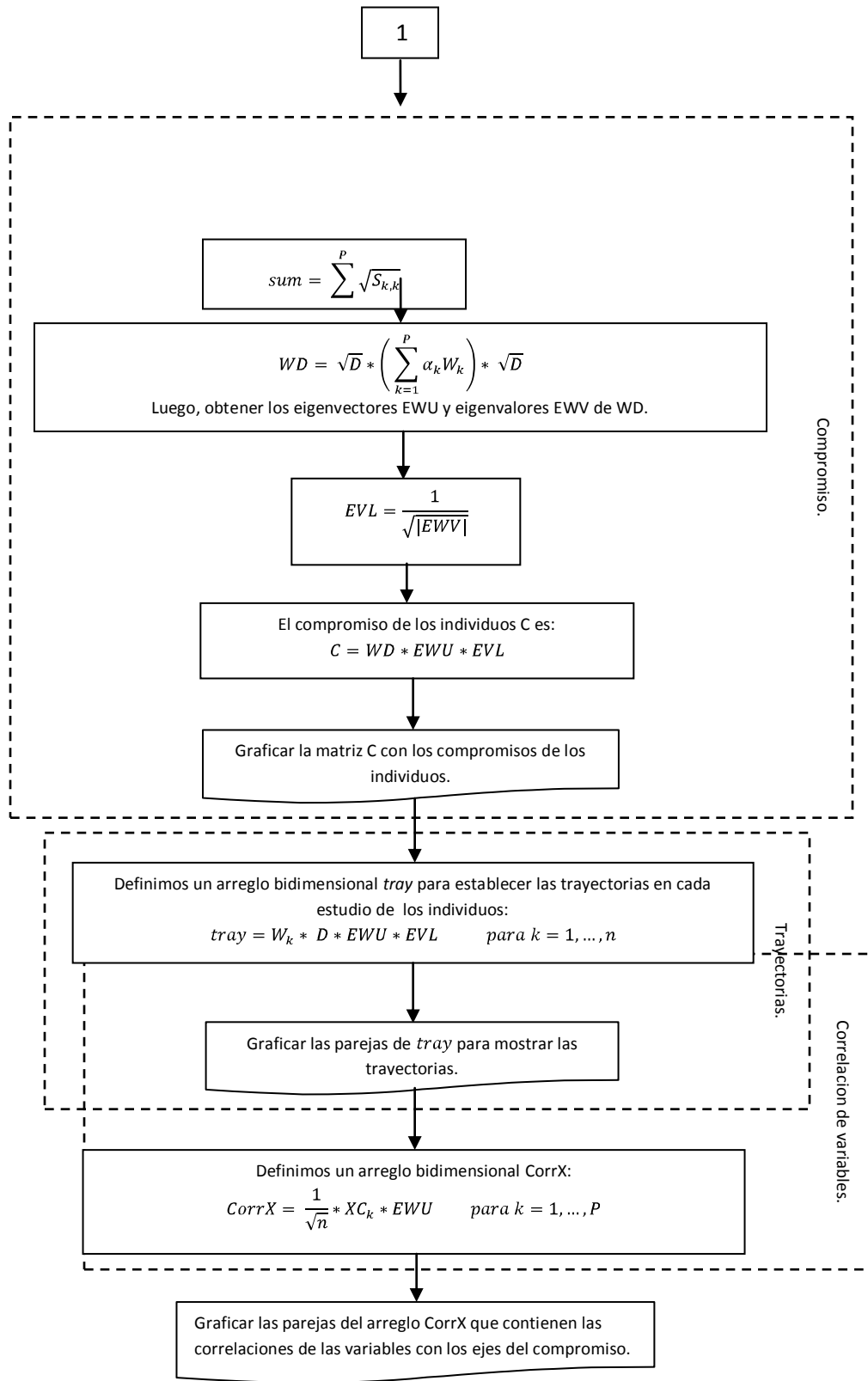
REFERENCIAS

[1] ESCOUFIER Y. (1973): Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.

- [2] ESCOUFIER Y. (1977): **Operators related to a data matrix**. Recent Developments in Statistic, North Holland , Amsterdam.
- [3] ESCOUFIER, Y., F.L. DÍAZ et M.C. BERNARD (1979): **La méthode Statis, une application à l'évolution des campagnes Ianguedociènnies-** USTL, Montpellier.
- [4] ESCOUFIER Y. (2006) Operators related to a data matrix: a survey. Compstat 2006. **Proceedings in Computational Statistic**, 17th Symposium Held in Rome Italy.
- [5] GEYNE RAJME,FRIDA. (2008): Análisis longitudinal de la marginación en las entidades federativas de México (1990 -2005): Tesis **de Licenciatura. Instituto Tecnológico Autónomo de México**.
- [6] GLAÇON, F. (1981): Analyse conjointe de plusieurs matrices de donnés. **Tesis PhD. Tercer Ciclo-USMG Grenoble**.
- [7] LAVIT, C. (1988): **Analyse conjointe de tableaux quantitatifs**. Masson, Paris.
- [8] LAVIT, CHRISTINE, ESCOUFIER, YVES, SABATIER, ROBERT and TRAISSAC, PIERRE. (1994): The ACT (STATIS method): **Computational Statistics and Data Analysis** 18, 97-119.
- [9] LINARES F., G. (2002): Representaciones gráficas de relaciones complejas. En: **Ordenando el caos**. Ed.: Redondo Botella, L., Editorial Félix Valera, La Habana, 78- 91.
- [10] LINARES, G. (2007) **Análisis de Datos Multivariados**. Editorial Benemérita Universidad Autónoma de Puebla. México.
- [11] SANDOVAL, M. DE L., ROMERO, M. G. y LINARES, G. (2008): Confiabilidad y validez de un instrumento para medir la calidad de los servicios bibliotecarios en la Benemérita Universidad Autónoma de Puebla. **Ponencia presentada en la 1ra. SIEP de FCFM**. Puebla México.

ANEXO 1. Diagrama del Método STATIS





ANEXO 2: Programa R del método STATIS

Se decidió usar el lenguaje R por su enfoque a la estadística lo que proporciona herramientas útiles para programar las operaciones requeridas para esta aplicación. Además, por ser un software libre, no representa ningún costo su adquisición y uso.

Preparación de los datos.

R permite leer datos de archivos, los cuales se obtuvieron de las encuestas realizadas y para poder usarlos se les dio un formato simple en archivos de texto, desde los cuales la aplicación puede de forma rápida usar los valores de los estudios y aplicar el algoritmo sobre ellos.

Funciones de R.

Algunas funciones importantes usada fueron:

read.table – Permite leer de archivos de texto datos y los pone en forma de tablas.

matrix – Crea matrices con los datos leídos para las operaciones del algoritmo.

win.graph() – Operación que abre una ventana para gráficos.

eigen – Devuelve un vector de valores propios y una matriz con vectores propios ordenados.

plot – Operación básica de graficación de datos.

Además se crearon funciones para la implantación del algoritmo como reducir matrices, obtener compromisos, trayectorias y correlaciones entre otras.

Método STATIS.

Sea X un arreglo de P matrices de tamaño N renglones por M columnas, donde N son el número de sujetos estudiados y M el número de variables utilizadas.

1. Definimos D como una matriz diagonal de dimensión N x N para definir los pesos de los sujetos.

$$D = \begin{bmatrix} 1/N & & \\ & \ddots & \\ & & 1/N \end{bmatrix}$$

2. Así mismo hacemos una matriz DP de dimensión P x P para definir los pesos de los estudios.

$$DP = \begin{bmatrix} 1/P & & \\ & \ddots & \\ & & 1/P \end{bmatrix}$$

Intraestructura

3. Sean XC el arreglo de las matrices de X reducidas y centradas, calculamos la distancia entre individuos W y la distancia entre variables con los productos escalares:

$$W_k = XC_k \cdot XC'_k \text{ para } k = 1, \dots, P$$
$$WV_k = XC'_k \cdot D \cdot XC_k \text{ para } k = 1, \dots, P$$

Interestructura.

4. Calculamos el coeficiente de covarianza vectorial SW para los individuos cuando tenemos el caso de varios estudios con los mismos individuos y SV para las variables cuando nuestro caso son estudios con individuos distintos:

$$SW_{k_1 k_2} = \text{Traza}(DW_{k_1} DW_{k_2}) \text{ para } k_1, k_2 = 1, \dots, P$$

$$SV_{k_1 k_2} = \text{Traza}(WV_{k_1} WV_{k_2}) \text{ para } k_1, k_2 = 1, \dots, P$$

5. Obtener la matriz de coeficiente de correlación vectorial entre los estudios

$$RV_{k_1k_2} = \frac{S_{k_1k_2}}{\sqrt{S_{k_1k_1}} \cdot \sqrt{S_{k_2k_2}}} \text{ para } k_1, k_2 = 1, \dots, P$$

Imagen Euclideana.

6. Obtener los eigen vectores U (vectores propios) y eigen valores V (valores propios) de SW o SV según el caso que estemos estudiando.

7. En una matriz diagonal V1 de 2 x 2 guardamos las raíces cuadradas de los 2 valores más grandes de V en orden decreciente.

8. Obtener las coordenadas con:

$$coord = U2 * V1 \text{ donde } U2 \text{ es una matriz con las 2 primeras columnas de } U$$

9. Graficar las parejas de coord.

Compromiso.

10. Primero obtenemos la siguiente sumatoria de SW o SV representada con S:

$$sum = \sum_{k=1}^P \sqrt{S_{kk}}$$

11. Luego obtenemos un vector de la siguiente forma:

$$\alpha_i = \frac{1}{\sqrt{1_{1,1}}} * sum * DP * U(:,1)$$

Donde U(:,1) significa hacer la multiplicación por cada columna de la matriz de vectores propios.

12. Ahora obtenemos la siguiente matriz:

$$WD = \sqrt{D} * \left(\sum_{k=1}^P \alpha_k W_k \right) * \sqrt{D}$$

13. Y de nuevo obtenemos los vectores (EWU) y valores (EWW) propios de WD con la siguiente modificación:

$$EVL = \frac{1}{\sqrt{|EWW|}}$$

14. Obtenemos la matriz compromiso, recordando que usamos nuevamente una matriz con los 2 eigen valores más grandes:

$$C = WD * EWU * EVL$$

15. Graficamos la matriz C con el compromiso de los individuos.

Trayectorias.

16. Definimos un arreglo bidimensional tray para establecer las trayectorias en cada estudio de los individuos:

$$tray = W_k * D * EWU * EVL \quad \text{para } k = 1, \dots, N$$

17. Graficamos las parejas de tray.

Correlación de las variables.

18. Definimos un arreglo bidimensional CorrX:

$$CorrX = \frac{1}{\sqrt{N}} * XC_k * EWU \text{ para } k = 1, \dots, P$$

19. Graficamos las parejas de CorrX que contienen las correlaciones de las variables con los ejes del compromiso.

ANEXO 3: Cuestionario

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA
VICERRECTORÍA DE DOCENCIA/ Dirección General de Bibliotecas

FOLIO:

Biblioteca: _____

Estimado usuario:
 Con el objetivo de mejorar el servicio que se ofrece en la biblioteca, estamos solicitando la opinión sobre ciertos aspectos a través del presente cuestionario, por lo que le pedimos responder de manera honesta y sincera, gracias por su colaboración.

1.- De la siguiente lista de servicios que ofrece esta biblioteca, haga el espacio de la calificación que crea conveniente del 1 al 10.

a) Préstamo de Libros	<input type="radio"/>	1
b) Renovación de préstamo	<input type="radio"/>	2
c) Devolución de libros	<input type="radio"/>	3
d) Servicio de consulta	<input type="radio"/>	4
e) Sala de cómputo	<input type="radio"/>	5
f) Búsqueda en Revistas Electrónicas	<input type="radio"/>	6

2.- De los siguientes servicios que presta la biblioteca valore en orden de importancia del 1 al 10, la forma en que busca la información.

a) Catálogo en línea	<input type="radio"/>	7
b) Con el bibliotecario responsable	<input type="radio"/>	8
c) Directamente en la estantería	<input type="radio"/>	9
d) Internet	<input type="radio"/>	10
e) Búsqueda en Revistas Electrónicas	<input type="radio"/>	11

3.- En una escala del 1 al 10, ¿Qué calificación le darías al personal bibliotecario que te atiende en esta biblioteca en cuanto a...?

a) Amabilidad	<input type="radio"/>	12
b) Trato	<input type="radio"/>	13
c) Aclaración de dudas	<input type="radio"/>	14
d) Atención personal	<input type="radio"/>	15
e) Eficiencia	<input type="radio"/>	16
f) Capacidad	<input type="radio"/>	17

4.- En lo general, ¿Qué calificación del 1 al 10 le darías al servicio que ofrece esta biblioteca?

<input type="radio"/>	18
-----------------------	----

5.- En una semana, aproximadamente, ¿cuántas veces estás a la biblioteca?

<input type="radio"/>	19			
DE 1 A 2	DE 3 A 4	DE 5 A 6	DE 7 A 8	MÁS DE 8

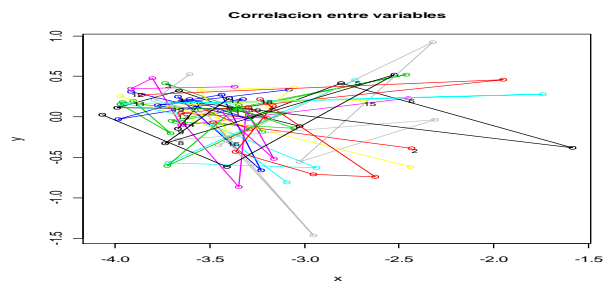
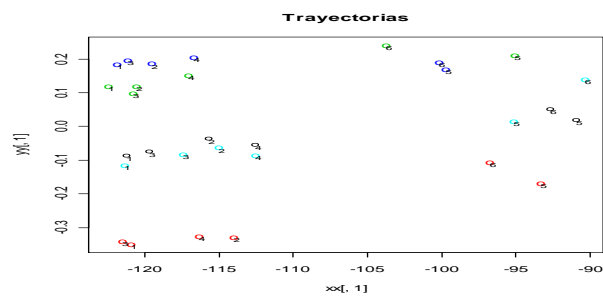
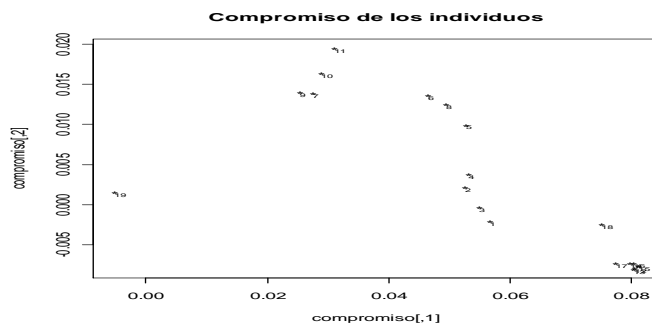
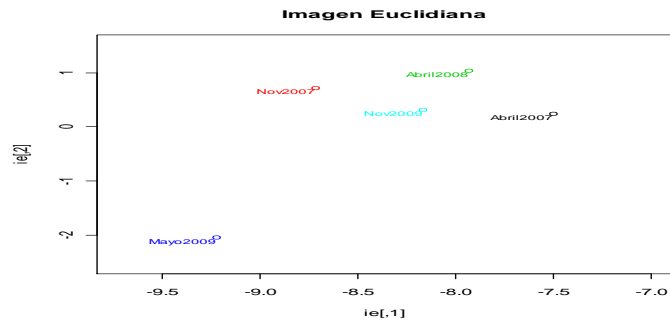
DATOS DEL USUARIO

a) Sexo	<input type="radio"/>	20			
	MASCULINO	FEMENINO			
b) Edad	<input type="radio"/>	21			
	17 a 20 años	21 a 60 años			
c) Usuario	<input type="radio"/>	22			
	ALUMNO	DOCENTE	INVESTIGADOR	TRABAJADOR	OTRO / ESPECIFICAR
d) Facultad o escuela	_____				23
e) Nombre de la Licenciatura o Posgrado:	_____				24
Especifica el cuatrimestre que cursas	<input type="radio"/>	25			

Agradecemos cualquier sugerencia o comentario que nos ayude a brindarte un mejor servicio

Actualización: noviembre 2007/ RSSP 7.5.1, 7A.

ANEXO 4: Biblioteca Luis Barragán



Anexo 5: Bibliotecas de Áreas Noviembre 2009

