

## Análisis de Tablas Múltiples de Datos: El Método STATIS

César Menacho Chiok<sup>1</sup>.

### Resumen

La presente investigación tiene como objetivo presentar un método para el análisis tablas múltiples de datos. El STATIS es una generalización del análisis de componentes principales (ACP), su objetivo es analizar varios conjuntos de tablas coleccionadas sobre las mismas observaciones. El objetivo del método de STATIS es 1) para comparar y analizar la relación entre los diferentes conjuntos de datos, 2) para combinar dentro de una estructura común llamada compromiso, el cual es analizado via ACP para revelar la estructura común entre las observaciones y finalmente 3) para proyectar cada conjunto de datos originales sobre el compromiso para analizar las similitudes y discrepancias.

Palabras clave:

### Abstract

In is the investigation have as goal there are a method for analyze tables multiples of data. STATIS is a generalization of principal component analysis (PCA) whose goal is to analyze several set of tables collected on the same set of observations. The goal of method of STATIS is 1) to compare and analyze the relationship between the different data sets, 2) to combine them into a common structure called compromise which is then analyzed via PCA to reveal the common structure between the observations, and finally 3) to project each of the original data sets onto the compromise to analyze communalities and discrepancies.

Key words:

### 1. Introducción<sup>1</sup>

Con frecuencia existen situaciones donde se desea analizar simultáneamente un conjunto de tablas de datos. En la actualidad se han desarrollado diversos métodos de tipo factorial con el propósito de analizar tablas múltiples de datos, con grupos de variables que pueden ser de naturaleza cuantitativa o cualitativa y que son medidas en distintas ocasiones. Estos métodos se basan en la búsqueda de una estructura común de las diversas tablas y, luego con esta estructura común explicar las diferencias entre las distintas tablas.

Para el análisis de tablas múltiples de datos, se han desarrollado una diversidad de métodos; el Método de STATIS (Structuration des Tableaux A Tríos Indices de la Statistique) propuesto por L'Hemier des Plantes (1976) y desarrollado por Lavit (1988), el Análisis Factorial Múltiple desarrollado por Escofier y Pagés (1992), el Método de Comparación de Grupos de Componentes Principales de Krzanowski (1979), el Método generalizado de Procusto de Gower (1975), el Análisis de Componentes Principales Triple (Tucker, 1966) y el Doble Análisis de Componentes Principales (Bouroche, 1975). Estos métodos difieren en la forma que se consigue una estructura común referencial llamado "*Compromiso*" que permite ubicar en un mismo subespacio los individuos de las diferentes tablas así como las variables.

Estos métodos tienen en común su relación con el Análisis de Componentes Principales (ACP). En muchas aplicaciones, cuando se tienen diferentes tablas de datos, que han sido analizadas separadamente por un ACP, se presenta el problema de compararlas, lo que no es posible generalmente de forma directa ya que los coeficientes de la transformación por componentes principales

presentan variabilidad muestral, pudiendo ocurrir que dicha transformación difiera sólo por el error muestral; de allí la importancia de estos métodos.

El método de STATIS (Structuration des Tableaux A Tríos Indices de la Statistique) propuesto por L'Hemier des Plantes (1976) y desarrollado por Lavit Ch. (1988) en la escuela Francesa, es un método de análisis multivariante para el estudio simultáneo de varias tablas de datos. Es particularmente útil para el análisis y descripción de fenómenos evolutivos a través del tiempo, obtenidos por mediciones repetidas sobre los mismos individuos.

El Método de STATIS permite el análisis simultáneo de varias tablas de datos cuantitativos con variables medidas en los mismos individuos. Estas variables describen un determinado fenómeno, que puede ser analizado en forma global, aunque las variables pueden ser distintas para cada tabla.

Las tablas de datos pueden estar formadas por los mismos individuos medidos en diferentes ocasiones con las mismas variables, los mismos individuos medidos en diferentes ocasiones por diferentes variables y diferentes conjuntos de individuos a los que se les miden las mismas variables. Los dos primero son los más comunes, que se refieren a tablas de datos que evolucionan en el tiempo y forman parte de datos longitudinales. Estos datos se organizan en tablas o matrices cúbicas, donde una entrada puede ser el tiempo. Así, sea  $X$  una matriz o tabla cúbica de datos, cuyo elemento  $X_{ijk}$  representa al individuo  $i$  con  $i = 1, \dots, n$ , para la variable  $j$  con  $j = 1, \dots, p_k$ , en la ocasión  $k$  con  $k = 1, \dots, K$ .

El objetivo del presente trabajo de investigación es mostrar el procedimiento metodológico de uno de los

métodos de análisis simultáneo de tablas conocido como STATIS; así como desarrollar una aplicación que permita mostrar su análisis e interpretación de resultados.

Carmen J. Mora C. (1992), realiza una aplicación del método STATIS con el objetivo de analizar las variaciones de las poblaciones del zooplancton en la represa de Uribantic-Caparo. Se aplicó el método STATIS por su capacidad en la descripción de fenómenos evolutivos a través del tiempo. Para el estudio se seleccionaron tres estaciones de muestreo durante el periodo marzo a diciembre de 1986; en estos puntos se tomaron muestras del zooplancton a siete niveles de profundidad. Los datos utilizados con el método STATIS estuvo conformado por 21 muestras (individuos) observadas en 10 meses (tablas) y con 10 distintas especies de zooplancton (variables). El método del STATIS permitió evidenciar grandes diferencias en las tres estaciones de muestreo; así como agrupar meses (tablas) con un comportamiento similar.

García y E. Simón (2002), presenta un resumen para la implementación del método STATIS utilizando el software SAPD (Système Pour l'Analyse des Données) versión 4.5 para Windows. En su trabajo desarrolla en primer lugar los fundamentos del STATIS como una técnica exploratoria multidimensional para el análisis simultáneo de varias tablas de datos. El método STATIS se desarrolla en tres etapas: la interestructura, el compromiso y la intraestructura. Se formula el método STATIS para el caso cuando se mide un grupo de variables en diferentes momentos de tiempo para un mismo conjunto de individuos, con el objetivo de estudiar un fenómeno evolutivo.

W. Castillo E. y J. González V. (1994), presentan los dos métodos más utilizados por la escuela Francesa para analizar tablas múltiples de datos: el Método STATIS y el Análisis Factorial Múltiple. Se analizan los tipos de datos que pueden procesar cada metodología y la forma en que cada una define el "compromiso" para obtener la representación simultánea de los individuos/variables en un espacio principal. Además, se estudian los aspectos computacionales en vista de una implementación. El objetivo es analizar las semejanzas y diferencias entre las diferentes situaciones a través de las configuraciones de los individuos y de las relaciones entre los diferentes grupos de variables. Considerando una aplicación evolutiva en el tiempo (años), se supone que se tiene  $r$  tablas de datos correspondiente a cada año  $X_1, \dots, X_r$ , donde  $X_k$  está previsto de las métricas  $M_k$  (métrica utilizada en el  $k$ -ésimo año) y  $D$  (matriz de pesos durante los  $r$  años). La estructura espacial de los individuos en el año  $k$ -ésimo está determinada por la matriz de productos internos entre individuos  $W_k = X_k' M_k X_k$  que define la matriz de la interestructura. Comparar las diferentes situaciones de cada una de las tablas se reduce a

comparar las matrices  $W_k$ , para tal efecto se utiliza la métrica de Hilbert-Smith para construir una matriz común.

Hervé Abdi and Dominique Valentin (2003), menciona que el método STATIS es una generalización del Análisis de Componentes Principales (ACP), cuyo objetivo es analizar varios conjunto de variables coleccionadas sobre un mismo conjunto de observaciones. El método STATIS, tiene como propósito 1) comparar y analizar las relaciones entre los diferentes conjuntos de datos, 2) combinar dentro de una misma estructura llamada "compromiso" la cual es analizada vía ACP para revelar una estructura común entre las observaciones y 3) para proyectar cada conjunto de datos originales sobre el compromiso y así analizar sus similitudes y diferencias existentes. La idea principal del STATIS, es analizar la estructura individual de un conjunto de datos (es decir, la relación entre el conjunto de datos individuales) para derivar de esta estructura un conjunto de ponderaciones para calcular una representación común de las observaciones llamado "el compromiso"; los pesos usados para calcular el compromiso son elegidos para hacer los más representativos todos los conjuntos de datos. El ACP permite dar la posición de las observaciones en el espacio de compromiso y la posición de las observaciones para cada conjunto de datos pueden ser representadas en el espacio de compromiso como puntos suplementarios.

L. Lera, A. Vallejo, H. Gomzáles y A. Gutiérrez (2000), analizan diferentes técnicas de tipo factorial para el análisis simultáneo de múltiples tablas de datos, entre las que se presentan el método de comparación de grupos de Componentes Principales, el método STATIS, el método de Procusto Generalizado, el Análisis de Componentes Principales Triple y el Doble Análisis de Componentes Principales. El objetivo fundamental de estos métodos, es la búsqueda de una estructura común para las diferentes tablas de datos (definición del compromiso) y a partir de esta estructura común explicar las diferencias entre las diferentes tablas. Estos métodos tienen en común su relación directa con el Análisis de Componentes Principales. Los datos se organizan en matrices cúbicas (tablas de 3 índices de entradas). Así, se tiene  $X$  una matriz cúbica, cuyo elemento  $X_{ijk}$  representa el individuo  $i$  con  $i = 1, \dots, n$ , para la variable  $j$  con  $j = 1, 2, \dots, p$ , en la ocasión  $k$  con  $k = 1, 2, \dots, K$ . Se estudian el caso donde se tienen  $K$  tablas de datos formadas por los mismos individuos y las mismas variables cuantitativas medidas en el tiempo. La estructura (intraestructura) esta descrita por las distancias mutuas entre las observaciones dada por:  $W_k = X_k X_k'$ , la cual representa la evolución de los individuos. Se define la interestructura como las relaciones entre las

observaciones, descritas por las distancias entre las  $W_k$  y se deduce del producto escalar de Hilbert-Smith:

$$(W_{k1}, W_{k2}) = \text{traza}(DX_{k1}DX_{k2}).$$

Con la finalidad de comparar los métodos propuestos se realiza una aplicación. Para tal efecto se obtuvieron muestras de sedimentos superficiales de 11 estaciones, en 5 ocasiones diferentes (abril y septiembre de 1988, marzo y septiembre de 1989 y abril de 1990), en la bahía de Cienfuegos, Cuba. Se analizaron los metales Cu, Pb, Zn como indicadores de contaminación urbano-industrial y Co, Fe, Mn y Ni como indicadores de contaminación industrial, con el objetivo de buscar una estructura común de la contaminación así como analizar su evolución.

S. Margado, J. Moncada y M. Purificación (2006), en este estudio se presenta una breve revisión de las metodologías AFM (Análisis Factorial Múltiple) y el STATIS. Se dispone para la aplicación de un conjunto de datos relativos al número de turistas que han visitado diversas regiones portuguesas entre los años 1993 y 2001. El objetivo es estudiar la evolución registrada de los visitantes y comparar globalmente las distintas regiones.

## 2. Materiales y métodos<sup>1</sup>

### Materiales y Equipo

Para la aplicación del método STATIS, se tomará los datos de las estadísticas ambientales presentados por el Instituto Nacional de Estadística e Informática (INEI) No. 08 Agosto 2006. Los datos corresponden al informe de la situación ambiental sobre la calidad del aire, producción de agua, calidad del agua y reservorio del río Rimac. Las estadísticas son registros de datos mensuales del periodo 2002-2005. Los resultados de la aplicación del método STATIS serán obtenidos a través del Software XLSTAT, programa computacional que permite realizar el análisis estadístico de varias técnicas multivariadas, entre las que figura el análisis de tablas múltiples de datos.

### Metodología

La metodología del método STATIS esta estrechamente ligada al ACP, se basa en comparar y analizar las relaciones entre los diferentes conjuntos de datos (tablas) para luego combinar dentro de una misma estructura llamada "compromiso" la cual es analizada vía ACP, para revelar una estructura común entre las observaciones y, luego proyectar cada conjunto de datos originales sobre el compromiso y así analizar sus similitudes y diferencias existentes.

El Método STATIS, esta caracterizada por tres etapas fundamentales: 1) Etapa de representación global (interestructura); consiste en elegir un objeto o individuo representativo para cada estudio (tabla). 2) Etapa búsqueda del compromiso; consiste en elegir una distancia (una métrica) para los objetos o individuos representativos obtenido en la primera etapa. 3) Etapa de representación detallada (intraestructura); consiste en encontrar una imagen

euclidiana de los objetos representativos, asociada a los productos escalares introducidos en la segunda etapa. Se iniciará con una serie de notaciones respecto a la estructura de datos.

Estructura de las tablas de datos

Los datos analizados por el método STATIS, están representados por  $K$  tablas  $X_1, \dots, X_k$  (centradas) que pueden estar expresadas como matrices de datos con  $j = 1, 2, \dots, P_k$  variables cuantitativas observadas en un mismo grupo de  $n$  individuos. Las variables pueden ser diferentes en cada una de las tablas. Así, tenemos las siguientes notaciones:

$K$  Número de tablas.

$n$  Número de individuos por tabla

$P_k$  Número de variables en la tabla  $k$ .

$X_k$  Tabla de datos  $k$  centrada, formada por  $n$  individuos y  $P_k$  variables

$X$  Tabla global resultante de la yuxtaposición de las  $K$  tablas  $X_k$ .

$D$  Matriz diagonal de orden  $n \times n$ , cuyos elementos de la diagonal se encuentran los pesos de los individuos. Definiéndose los pesos de los individuos  $p_1, \dots, p_n$  tal que

$$p_i > 0, \forall i = 1, 2, \dots, n \text{ y } \sum_i p_i = 1.$$

Se supone que los individuos tienen el mismo peso en cada una de las tablas.

$M_k$  Son  $K$  matrices diagonales de orden

$P_k \times P_k$  cada una, en cuya diagonal se encuentran los pesos de las variables para cada una de las tablas. Definiéndose los pesos de las variables para la  $k$ -ésima tabla, como

$$m_1^k, m_2^k, \dots, m_{j_k}^k \text{ tal que}$$

$$\sum_i m_i^k \neq 0, \forall j = 1, 2, \dots, j_k.$$

Se supone que los individuos tienen el mismo peso en cada una de las tablas.

### Análisis de la Intraestructura

El propósito de esta etapa es describir los individuos según sus estructuras en cada una de las tablas. Para lograr esto, el STATIS proyecta a los individuos, tales como son vistos en cada tabla, en el mismo subespacio factorial de los individuos medios o compromisos. Esto permite estudiar el comportamiento de cada individuo alrededor de su posición media o compromiso, definida como una media ponderada de los individuos parciales según cada tabla.

La estructura espacial de los individuos en la  $k$ -ésima tabla esta determinada por la matriz de productos internos entre individuos, definida por:

$$W_k = X_k' M_k X_k$$

Se tiene que:  $W_j D = X_j X_j' D$

$$\|W_j D\| = \sqrt{\sum_i (\lambda_i^j)^2}$$

Donde :

$X_j$  es la matriz asociada al grupo  $j$  de variables

$D$  una matriz diagonal que recoge los pesos de los individuos

$\lambda_i^j$  es el  $j$ -ésimo autovalor del ACP de la tabla  $j$ .

A partir de esto, se construye una imagen euclidiana de los individuos, se define la matriz  $W$ , denominada matriz de la interestructura, simétrica, semidefinida positiva y de orden  $K \times K$ , cuyos elementos son los coeficientes de asociación entre los individuos  $k, l$  de la matriz  $W$ .

#### Análisis de la Interestructura

La finalidad de esta etapa es comparar las configuraciones de los mismos individuos, obtenidas en las diferentes tablas (L'Hermier, 1976; Lavit, 1988; Dazy, 1996). Para evaluar la relación entre los grupos o tablas se utiliza un coeficiente de asociación que se basa en el coeficiente  $RV$  de Dazy (1996). Así, se define el coeficiente de asociación entre los individuos de las tablas  $k$  y  $l$  a partir del producto escalar de Hilbert-Smitjh y considerando que los individuos se toman normalizados por la siguiente expresión:

$$RV(k, l) = \frac{(W_k D, W_l D)}{\|W_k D\| \|W_l D\|} = \frac{\text{Traza}(W_k D W_l D)}{\|W_k D\| \|W_l D\|}$$

El valor de este coeficiente se encuentra entre 0 a 1.

Un valor cercano de  $RV(k, l)$  a 1 indicaría una relación (equivalencia o similitud) entre las tablas

$k$  y  $l$ ; es decir, los individuos u observaciones tienen la misma estructura interior para las tablas

$X_k$  y  $X_l$ . Entonces en esta etapa de la interestructura se realiza un análisis de las semejanzas y diferencias entre los individuos y variables de cada una de las tablas; obteniéndose una matriz de la interestructura, simétrica, semidefinida positiva y de orden  $K \times K$ .

#### Análisis del Compromiso

El objetivo de esta etapa es comparar los individuos (o variables) desde el punto de vista global. Para esto, se define en el método STATIS una matriz de resumen de las  $K$  tablas de datos, denominada

matriz de compromiso (tabla global). El método STATIS utiliza el compromiso propuesto por H. L'Hermier des Plantes (1976), que permite obtener una estructura de referencia común en el cual se representen adecuadamente las diferencias estructuradas aportadas por cada tabla. La expresión a considerar es:

$$C = \sum_{i=1}^K \alpha_i W_i$$

La forma de elegir los  $\alpha_i$ , es buscar el punto  $C$  respecto al cual la nube  $\{W_1, \dots, W_k\}$ , provista de los pesos  $p_1, \dots, p_k$  tenga inercia mínima:

$$I_C = \sum_{i=1}^K p_i \|W_i - C\|^2$$

para conseguir esto se tiene que  $\alpha_i = p_i$ , para  $i = 1, \dots, K$ ; es decir el mínimo se alcanza cuando

$C$  es el centro de gravedad de la nube o proyecciones. Así, si los pesos son iguales, el compromiso es la media aritmética de la nube.

Esta matriz de compromiso es simétrica y semidefinida positiva, por ser una combinación lineal positiva de matrices semidefinidas positivas y resume las  $K$  tablas de datos en una representación del conjunto, constituyendo el mejor resumen de las matrices.

### 3. Resultados<sup>1</sup>

#### Aplicación y resultados

Los datos utilizados en esta investigación corresponden de la concentración mensual (Miligramos por litro – mg/l de agua) para el periodo 2002-2005 de 5 compuestos (Hierro, Plomo, Cadmio, Aluminio, Material Orgánico y Nitrato) registrados por SEDAPAL en el río Rimac. El propósito es buscar una estructura común de la contaminación, así como analizar su evolución.

#### Estructura de las tablas de datos

La estructura de los datos para el ejemplo de aplicación del método STATIS será: 4 tablas (4 años), 6 variables y 12 datos (12 meses). Cada tabla esta compuesta por una matriz de 12 filas (12 meses) y 6 columnas (6 variables).

Los datos analizados por el método STATIS, están representados por 4 tablas

$X_1, X_2, X_3, X_4$  (centradas) que pueden estar expresadas como matrices de datos con  $j = 1, 2, 3, 4$  variables cuantitativas observadas en un mismo grupo de 12 individuos. Así tenemos las siguientes notaciones:

$K = 4$  Número de tablas.

$n = 12$  Número de individuos por tabla

$P_k = 6$  Número de variables en la tabla  $k$ .

$X_k$  Tabla de datos  $k$  centrada, formada por 12 individuos y 6 variables

$X$  Tabla global resultante de la yuxtaposición de las 4 tablas  $X_k$ .

Análisis de la Interestructura

El análisis de la interestructura permitirá el estudio comparativo de las distintas tablas. Se trata de identificar que tablas (años) o grupos de variables

(compuestos) son semejantes (tienen la misma estructura) y cuales son más diferentes.

En la Tabla 1, se muestra un resumen de los valores propios y los porcentajes de inercia asociados a cada uno de los factores obtenidos por los ACP aplicados a cada uno de las tablas (años). Se observa que los cuatro años tienen un comportamiento bastante similar; con un primer eje factorial preponderante, que recoge entre un 60.1% y un 70.7% de la variabilidad total y un segundo eje factorial que oscila entre el 14.2% y el 23.4%.

La semejanza entre las tablas (años) permite realizar el estudio comparativo, el análisis de la infraestructura.

**Tabla 1.** Valores propios y Porcentaje (%) de Inercia en cada año (tabla).

Años	Valores propios					% de Inercia				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
2002	3.606	1.403	0.522	0.323	0.134	60.103	23.388	8.697	5.381	2.233
2003	3.767	1.098	0.760	0.368	0.005	62.791	18.301	12.666	6.126	0.090
2004	4.245	0.950	0.783	0.021	0.000	70.753	15.838	13.051	0.348	0.007
2005	4.127	0.855	0.675	0.231	0.108	68.786	14.246	11.250	3.848	1.794

En la Tabla 2, se presenta los coeficientes de la matriz RV, que indican la relación o asociación entre las tablas (años), cuando el coeficiente se aproxima a la unidad indica que mayor es la similitud de las tablas. En el cuadro, se aprecia que el coeficiente RV para los años 2004 y 2005 es de 0.539, podemos

indicar que en ambos años presenta una estructura similar de los datos. Es decir, las concentraciones mensuales de los metales (Fe, Pb, Cd, Al, MO y Ni) encontradas en el Río Rimas fueron similares para los años 2004 y 2005.

**Tabla 2.** Matriz de valores RV entre los años (tablas).

Años	2002	2003	2004	2005
2002	1.000	0.361	0.153	0.196
2003	0.361	1.000	0.366	0.490
2004	0.153	0.366	1.000	0.539
2005	0.196	0.490	0.539	1.000

En la Tabla 3, se muestra el ACP resultante del análisis global de las tablas (años). El porcentaje de variabilidad que es explicada por el primer eje

factorial global es de 42.5%, el segundo eje explica sólo el 16.7% y el tercer el 13.9% de variabilidad global del conjunto de las tablas.

**Tabla 3.** Valores propios del análisis global.

	F1	F2	F3	F4	F5
Valor propio	2.601	1.024	0.854	0.514	0.434
Variabilidad (%)	42.470	16.715	13.949	8.394	7.080
% acumulado	42.470	59.185	73.134	81.529	88.609

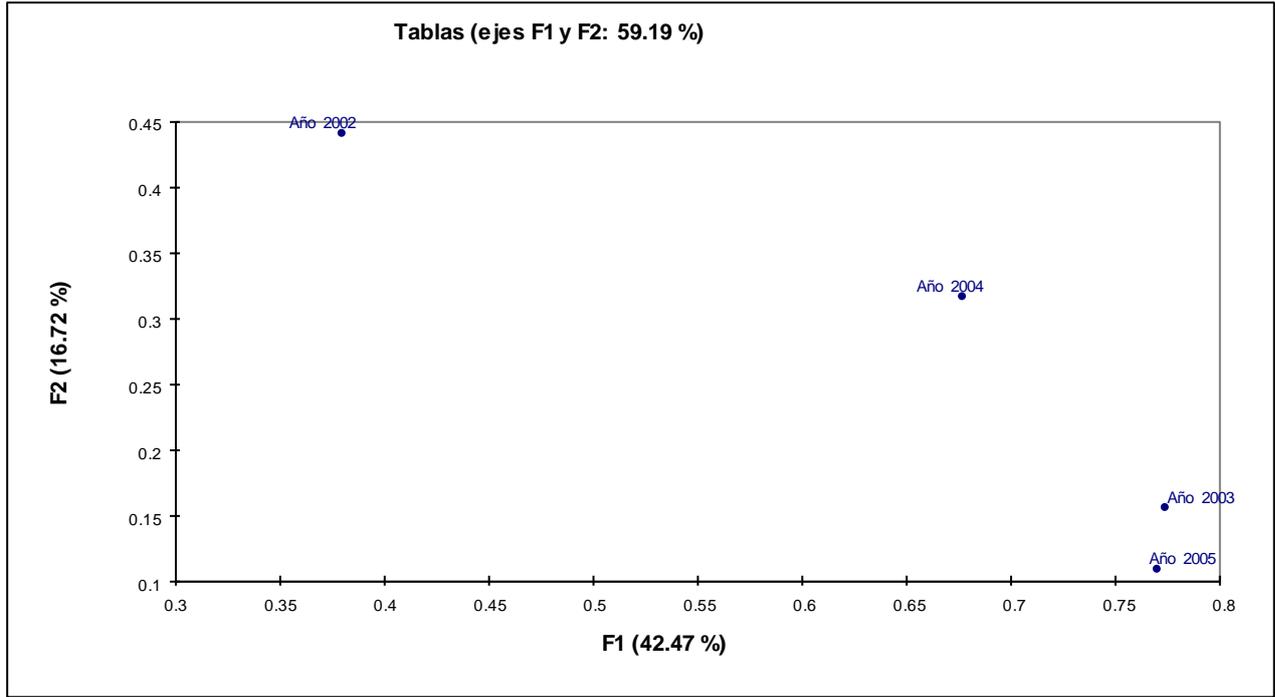
En la Tabla 4, se presenta las contribuciones de las tablas (años) a la formación de los primeros ejes globales. Se nota que los años 2003 y 2005 son los que mayor participan en la formación del primer eje

global con 29.7% y 29.6% respectivamente; los años 2002 y 2004 participan con el 43.2% y 30.8% en la conformación del segundo eje global respectivamente.

**Tabla 4.** Coordenadas y contribuciones (%) de cada año.

Años (Tabla)	Coordenadas			Contribuciones		
	F1	F2	F3	F1	F2	F3
2002	0.380	0.442	0.471	14.602	43.166	55.173
2003	0.774	0.156	0.187	29.764	15.266	21.942
2004	0.677	0.316	0.042	26.023	30.893	4.864
2005	0.770	0.109	0.154	29.611	10.674	18.021

**Figura 1.** Contribuciones de los años a la formación del primer y segundo eje global.



En la Tabla 5, permite analizar las correlaciones de las variables con el primer y segundo eje factorial global. Se nota que los metales Fe, Al y Pb son los que más están relacionados con el primer eje, siendo

el porcentaje de variabilidad que explica este eje del 42.5%. El segundo eje está relacionado con las variables Materia orgánica.

**Tabla 5.** Correlaciones de las variables con el primer y segundo eje factorial.

Variables	2002		2003		2004		2005	
	F1	F2	F1	F2	F1	F2	F1	F2
Hierro (Fe)	0.688	-0.526	0.955	0.021	0.734	0.613	0.735	0.023
Plomo (Pb)	0.341	-0.526	0.855	-0.298	0.855	0.448	0.833	-0.106
Cadmio (Cd)	0.274	-0.405	0.544	-0.558	0.732	0.614	0.683	0.540
Aluminio (Al)	0.776	-0.415	0.923	-0.083	0.740	0.612	0.583	0.018
Materia Orgánica	-0.132	0.177	-0.064	-0.101	0.646	-0.073	0.856	0.372
Nitratos (Ni)	-0.291	0.820	-0.346	0.413	-0.320	-0.099	-0.636	0.093

**Análisis de la Intraestructura**

En este análisis se estudia las semejanzas y diferencias entre las tablas a partir de los individuos (filas). Permite detectar que meses son los que hacen la semejanza o diferencias entre las tablas (años).

En la Tabla 6, se tiene la representación global de los 12 meses. Este cuadro permite evaluar y comparar el comportamiento de la contaminación mensual respecto concentración de los metales (mg/l de agua).

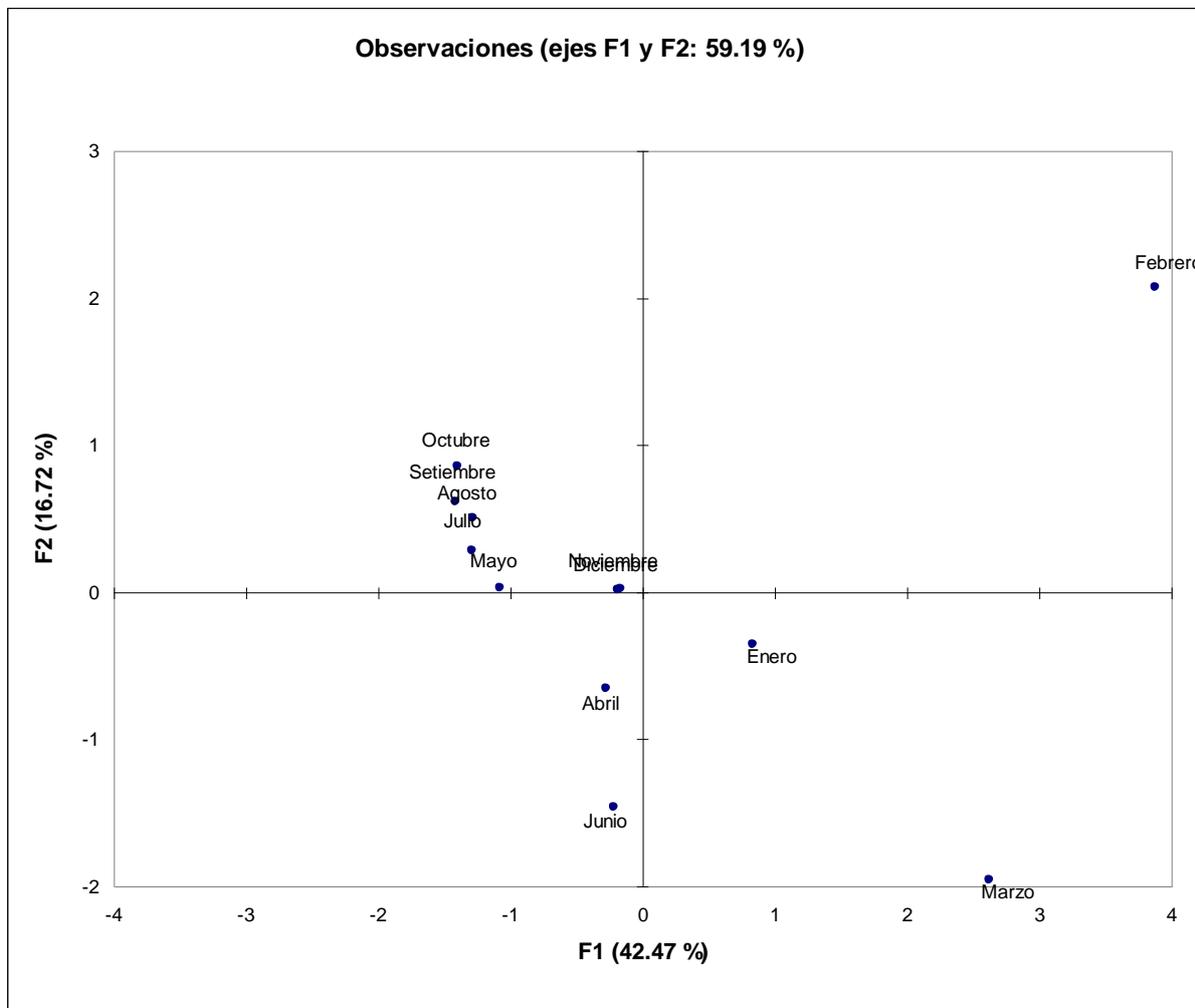
**Tabla 6.** Coordenadas de los meses (individuos) sobre los ejes factores.

Meses	F1	F2	F3	F4	F5
Enero	0.832	-0.355	-1.074	1.481	-0.823
Febrero	3.876	2.078	0.262	-0.531	-0.354
Marzo	2.618	-1.956	-0.899	-0.302	1.218
Abril	-0.274	-0.653	-0.104	0.070	-0.977
Mayo	-1.081	0.031	-0.853	-0.066	-0.337
Junio	-0.219	-1.458	1.461	-1.207	-0.840
Julio	-1.292	0.286	-0.750	-0.248	-0.012
Agosto	-1.282	0.506	-0.407	-0.177	0.333
Setiembre	-1.419	0.619	-0.506	-0.362	0.271
Octubre	-1.399	0.858	0.087	-0.384	0.725
Noviembre	-0.193	0.018	0.898	0.495	0.157
Diciembre	-0.166	0.026	1.884	1.232	0.638

En el Gráfico 2, se muestra el resultado de la Tabla 6, considerando los dos primeros ejes factoriales. Este gráfico, permite identificar los meses cuyas concentraciones de los metales han sido similares y diferentes. Se aprecia, que los meses de febrero y marzo las concentraciones de los metales han sido muy altas y distintas respecto a los demás meses; en menor grado se tiene los meses de enero, abril y

junio. Así mismo, se observa grupos de meses que muestra similar comportamiento (estructura) en términos de los años. Así, los grupos de meses de mayo, Julio, agosto, setiembre y octubre, y noviembre y diciembre; tuvieron similares concentraciones de los metales durante los años 2002 a 2005.

**Figura 2.** Coordenadas de los dos primeros ejes y la concentración mensual.



**Análisis de las Trayectorias**

Es el análisis de los individuos y de las variables proyectadas a la estructura común (compromiso). Este análisis permitirá evaluar la evolución de los

meses y de las variables proyectadas en los años. Esto es, se identificarán la evolución de la concentración de los metales a lo largo de los meses y años.

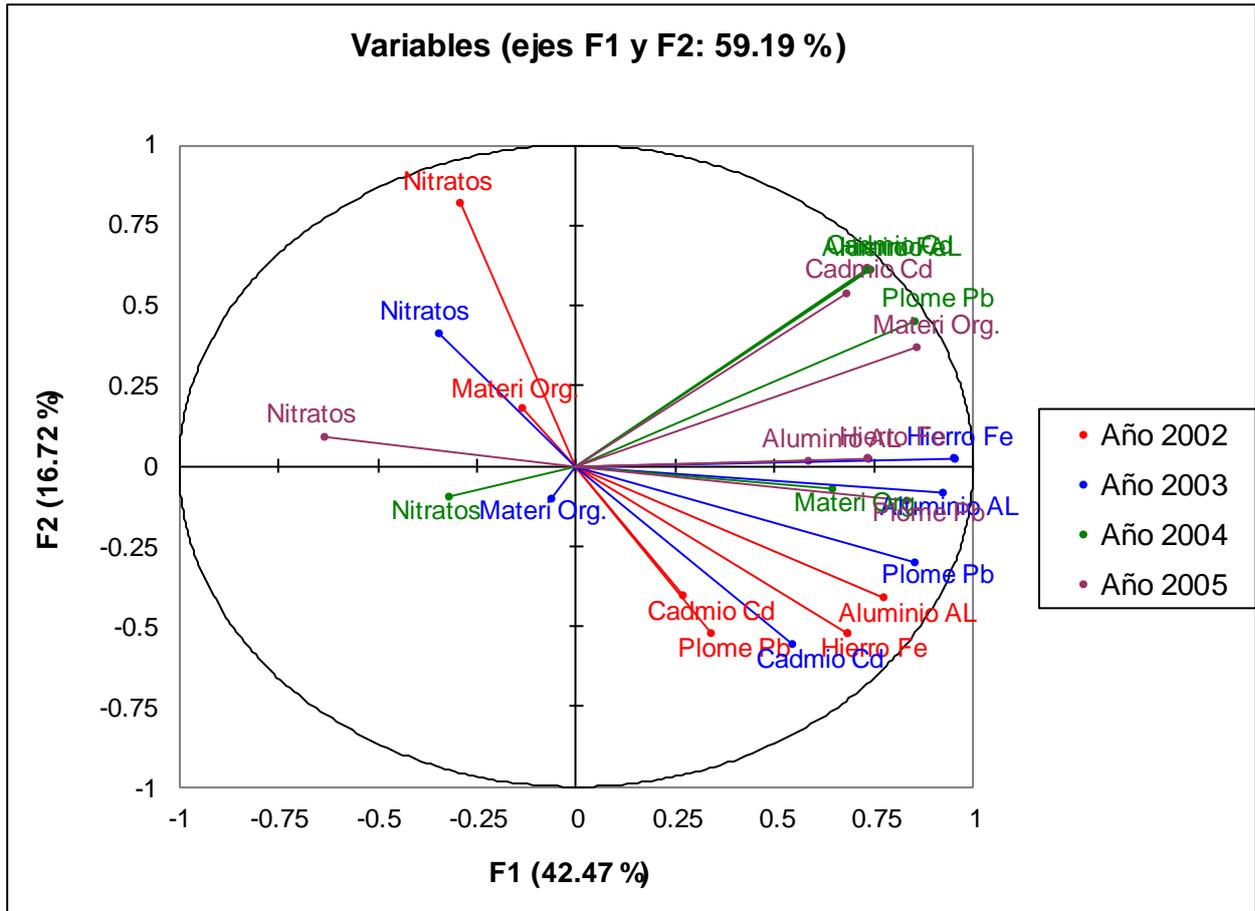
**Tabla 7.** Contribuciones parciales de las variables a los ejes factoriales.

Variables	2002		2003		2004		2005	
	F1	F2	F1	F2	F1	F2	F1	F2
Hierro Fe	5.051	7.509	9.317	0.011	4.884	8.649	5.033	0.012
Plomo Pb	1.241	7.489	7.459	2.307	6.622	4.616	6.472	0.265
Cadmio Cd	0.798	4.447	3.025	8.078	4.847	8.676	4.344	6.902
Aluminio AL	6.420	4.666	8.698	0.179	4.966	8.607	3.171	0.008
Materi Org.	0.187	0.847	0.042	0.264	3.775	0.121	6.828	3.284
Nitratos	0.905	18.208	1.223	4.429	0.929	0.224	3.764	0.204

En la Tabla 7, se presenta las contribuciones de las variables para conformar los dos primeros ejes

factoriales a través de los años. Una mejor representación se muestra en el Gráfico 3.

Figura 3. Contribuciones parciales de las variables a través de los años.



#### 4. Conclusiones

Las conclusiones y recomendaciones que pueden resultar del presente trabajo de investigación son:

El método STATIS, es uno de los métodos más usados con el propósito es analizar conjuntos de datos recogidos de individuos a los cuales se les han medido un grupo de variables cuantitativas.

La metodología presentada para STATIS, permitió encontrar una estructura común (compromiso) para el conjunto de tablas, con una ponderación parcial para los individuos de cada tabla.

El método STATIS brinda un análisis del conjunto de datos bajo la perspectiva de las variables, individuos y tablas; así como las proyecciones internas de variables e individuos encada tabla.

El método STATIS, permitió analizar los datos sobre la concentración de compuestos en el Río Rimac en el periodo de 2002 a 2005. Así se puede mencionar las siguientes conclusiones:

La concentración de los compuestos (Fe, Pb, Cd, Al, Mo y Ni) muestran un comportamiento similar a través de los 4 años; con una mayor similitud para los últimos dos años 2004 y 2005.

Toda la variabilidad puede ser recogida por los dos primeros factores, que explican el 59.2%. El primer factor se identifica con los compuestos de Fe y Al y el segundo con Pb, Cd, MO y Ni.

En los años 2002, 2003 y 2005 hubo una fuerte concentración de los compuestos Fe y Al (hay una evidente están representadas por el primer factor y el año 2002 por el segundo factor).

En el análisis de la estructura común (compromiso) para los 12 meses, se identificó tres grupo con similitudes en la concentración de los compuestos. Un grupo por los meses mayo, julio, agosto, setiembre; un segundo por noviembre y diciembre y un tercero enero, marzo y febrero.

Se recomienda comparar este método con otros, tales como el Análisis Factorial Múltiple, el Método de Comparación de Grupos de Componentes Principales, el Método generalizado de Procusto, el Análisis de Componentes Principales Triple y el Doble Análisis de Componentes Principales.

#### 5. Referencias bibliográficas

Carmen J. Mora C.; "Aplicación del método de STATIS: Análisis del zooplancton del embalse uribante" Tesis para optar el título de Licenciada en Estadística. 1992, Universidad de los Andes.

Ignacio García Lautre y Estela Simón Nicó; "Implementación del Método STATIS con el SPAD". Metodología de Encuestas. Vol.4, Número 1, 2002. Pag.129-139.

- Hervé Abdi & Dominique Valentin; "The STATIS Method". Encyclopedia of Measurement and Statistics. 2006.
- William Castillo Elizondo y Jorge Gonzáles Varela; "Análisis de Tablas Múltiples de datos". Revista de Matemática. 1994.
- Lydia Lera Márquez, Amparo Vallejo Arboleda, Humberto Gonzáles y Armando Gutiérrez; "Análisis de la Estructura Común de Varias Tablas de datos a través de diferentes técnicas factoriales". Revista Integración Escuela de matemáticas. Vol. 16, No. 2. 1998.
- Margado S.N., Moncada J.H y Galindo P.; "Análisis Factorial Múltiple y STATIS: una aplicación a datos de Afluencia Turística". XXIX Congreso Nacional de Estadística e Investigación Operativa. Mayo 2006.
- Escofier B y Pagés J; "Analices Factorielles Simples et Multiples; Objectifs, Methodes et Interpretation". Dunod. París. 1988.
- Lavit Ch.; "Analyse Conjointe de Tableaux Quantitatifs". Ed. Mason, París. 1988.
- L'Hermier des Plantes H.; "Structuration des tableaux á trios indices de la statistique: théorie et application d'une methode d'analyse conjointe". Tesis de Doctorado. Université des Sciences et Tecnique du Languedoc. Motpellier. 1976.
- Gower J. C.; "Generalized Procusto Análisis", Psychometrika, 40, pag.33-51. 1975
- Krzanowski W. J.; "Between-Groups Comparison of Principal Components", Jornal of the American Statistical Association", 74, No374. Pag.703-707. 1979.
- Tuckey L. R. ; "Some Mathematical Notes on Three Mode Factor Analysis", Psychometrika, 31, No.3, pag.279-311. 1966.
- Dasy F. y Le Barzic J.F. "L'Analyse des données evolutives, méthodes et applications", Editions Tecnip. 1996.