

UNIVERSIDAD DE SALAMANCA

Departamento de Estadística

Doctorado en Estadística Aplicada



**CONTRIBUCIONES A LOS METODOS STATIS
BASADOS EN TECNICAS DE APRENDIZAJE NO
SUPERVISADO**

AUTORA:

CARMEN CECILIA RODRÍGUEZ MARTÍNEZ

DIRECTORAS

M^a PURIFICACIÓN GALINDO-VILLARDÓN

M^a PURIFICACIÓN VICENTE-GALINDO

2020



Departamento de Estadística

Universidad de Salamanca

CONTRIBUCIONES A LOS METODOS STATIS BASADOS EN TECNICAS DE APRENDIZAJE NO SUPERVISADO

Trabajo para optar al título de Doctor por la
Universidad de Salamanca.

Presenta

Carmen Cecilia Rodríguez Martínez

Salamanca

2020



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

DEPARTAMENTO DE ESTADÍSTICA

DRA. M^a PURIFICACIÓN GALINDO-VILLARDÓN

Catedrática del Área de Estadística e Investigación Operativa de la Universidad de Salamanca

DRA. M^a PURIFICACIÓN VICENTE-GALINDO

Profesora Titular del Área de Estadística e Investigación Operativa de la Universidad de
Salamanca

CERTIFICAN que **Dña. Carmen Cecilia Rodríguez Martínez** ha realizado en la Universidad de Salamanca, bajo su dirección, el trabajo que para optar al título de Doctor presenta con el título: **CONTRIBUCIONES A LOS METODOS STATIS BASADOS EN TECNICAS DE APRENDIZAJE NO SUPERVISADO**, autorizando expresamente su lectura y defensa.

Y para que conste, firma el presente certificado en Salamanca a

D^a. M^a Purificación Galindo-Villardón

D^a. M^a Purificación Vicente-Galindo

Siempre parece imposible hasta que se hace.

Nelson Mandela

DEDICATORIA

A mi Diego André

A mis padres.

AGRADECIMIENTOS

Primero y ante todo a Dios, por darme fuerzas para seguir adelante y por estar siempre a mi lado en cada paso que doy.

A la Dra. Purificación Galindo-Villardón por su confianza y guía para culminar este proyecto, que sin ella no hubiese sido posible. Gracias Puri.

A la Dra. Purificación Vicente-Galindo, por aceptar ser mi tutora en este proyecto tan importante para mí.

A la Dra. Mitzi Isabel Cubilla Montilla, gracias por tu gran apoyo amiga, por estar siempre y por tus consejos para la culminación de este trabajo, has sido parte fundamental para el logro de este objetivo. Eres extraordinaria.

A las doctoras Ana Belén Nieto Librero y Nerea González por sus atinadas guías que fueron fundamentales para terminar este proyecto. Ambas son maravillosas.

A todos mis compañer@s y ahora amig@s que me acompañaron durante estos años y siempre me dieron su voz de aliento: Cinthia, Greibin, Joel, Nathalia, Nerea, María, Mariela, Victoria y Zaira. Especialmente quiero agradecerles a Zaira y Victoria por estar conmigo durante aquellas fiestas navideñas que pasamos lejos de nuestras familias, las quiero mucho.

A mis colegas y amigos de Panamá por sus continuas palabras de apoyo: Adela, Anitzia, Aurora, Chachi, Elena, Estelina, Gonzalo, Lamed y Parker.

INDICE DE CONTENIDO

DEDICATORIA	IX
AGRADECIMIENTOS.....	XI
INTRODUCCIÓN	XVIII
CAPITULO I: MÉTODOS PARA ANÁLISIS DE DATOS DE TRES MODOS	1
1.1. INTRODUCCIÓN	2
1.2. ARREGLOS MULTIVÍA	3
1.2.1. Arreglos de Tres Vías.....	3
1.2.2. Matriciación.....	6
1.3. DATOS DE TRES MODOS	7
1.3.1. Datos de Tres Vías.....	7
1.3.2. Datos de Conjuntos Múltiples	8
1.4. MÉTODOS PARA ANALISIS DE DATOS DE TRES MODOS.....	11
1.4.1. Escuela Anglosajona	13
1.4.2. Escuela Francesa.....	17
1.4.3. Escuela Salmantina	20
CAPÍTULO II: METODOS STATIS	26
2.1. INTRODUCCIÓN	27
2.2. ESTADO DEL ARTE DE LOS MÉTODOS STATIS	28
2.3. PRIMEROS DESARROLLOS DE LA FAMILIA DE MÉTODOS STATIS.....	52
2.3.1. STATIS O STATIS-ACT	54
2.3.1.1. Notación.....	55
2.3.1.2. Análisis de la Inter-estructura.....	57
2.3.1.3. Análisis del compromiso	60
2.3.2. STATIS-dual.....	64
2.3.2.1. Interestructura	66
2.3.2.2. Intraestructura	68
2.4. NUEVOS APORTES A LOS MÉTODOS STATIS.....	70
2.4.1. X-STATIS o PTA.....	71
2.4.2. Análisis factorial múltiple (AFM).....	74
2.4.3. STATIS-CoA	77
2.4.4. STATICO.....	77
2.4.5. DoubLe-STATIS (Do-ACT)	79
2.4.6. (k+1)-STATIS o STATIS externo	82
2.4.7. DISTATIS	83
2.4.8. CANOSTATIS O STATIS-CANONICO/Y STATIS CANONICO DUAL	85
2.4.9. STATIS-4.....	90
2.4.10. Kernel-STATIS.....	91
2.4.11. COVSTATIS.....	93
2.4.12. COSTATIS	93
2.4.13. POWER-STATIS.....	96
2.4.14. ANISOSTATIS.....	98
2.4.14.1. Criterio 1	98
2.4.14.2. Criterio 2	100
2.4.14.3. Criterio 3	100
2.4.15. STATIS-LDA	101

2.4.16. <i>INTERSTATIS</i>	103
2.4.16.1. Interstatis con AI y ACPC	103
2.4.16.2. Interstatis con AA y ACP de Centros.....	105
2.4.17. <i>SIR-STATIS</i>	106
2.4.18. <i>HiDISTATIS</i>	109
2.4.19. <i>DiDISTATIS</i>	115
2.4.20. <i>CATATIS</i>	122
2.4.21. <i>STATICO-COA</i>	123
2.4.22. <i>CLUSTATIS</i>	124
CAPÍTULO III: SOFTWARE PARA STATIS.....	131
3.1. INTRODUCCIÓN	132
3.2. MULTBILOT.....	134
3.3. LIBRERIAS Y PAQUETES EN R PARA STATIS.....	135
3.3.1. <i>MExPosition</i>	138
3.3.2. <i>MultBiplotR</i>	140
3.3.3. <i>Multigroup</i>	142
3.3.4. <i>ade4</i>	143
3.3.5. <i>SirSTATIS</i>	144
3.3.6. <i>dad</i>	145
3.3.7. <i>ClustBLOCK</i>	148
CAPÍTULO IV: SPARSE STATIS-DUAL.....	152
4.1. INTRODUCCIÓN	153
4.2. SPARSE PCA	154
4.3. SPARSE STATIS-DUAL	156
4.3.1. <i>Statis-dual restringido via elastic net</i>	158
4.4. ALGORITMO EN R: SPARSE STATIS-DUAL.....	163
4.5. ANÁLISIS DE DATOS REALES: APLICACIÓN DE SPARSE STATIS-DUAL	168
4.5.1. <i>Base de datos</i>	168
4.5.2. <i>Análisis</i>	171
CONCLUSIONES	177
ARTICULO PUBLICADO.....	180
ARTÍCULOS SOMETIDOS	197
PARTICIPACIÓN EN CONGRESOS.....	198
ANEXO.....	199
BIBLIOGRAFÍA.....	223

ÍNDICE DE FIGURAS

Figura 1. Esquema de arreglo de tres vías o tensor de tercer orden $\mathcal{X} \in \mathbb{R}^{I \times J \times H}$	4
Figura 2. Esquema de arreglos de tres vías.....	5
Figura 3. Cortes o caras de arreglos de datos de tres vías.	5
Figura 4. Matriciación de Arreglos de Tres Vías	6
Figura 5. Diagrama de Datos de Tres Vías.....	8
Figura 6. Diagrama de Datos de Conjuntos Múltiples.	9
Figura 7. Diagramas de conjuntos múltiples.	10
Figura 8. Clasificación de los modelos para el análisis de datos en tres modos de acuerdo a Kiers (1991)	12
Figura 9. Desarrollo cronológico de algunos modelos de Análisis de datos de Tres Modos de la Escuela Anglosajona.....	16
Figura 10. Desarrollo cronológico de algunos modelos de Análisis de datos de Tres Modos de la Escuela Francesa.....	19
Figura 11. Desarrollos cronológicos de las aportaciones de la Escuela Salmantina al análisis de datos de tres modos.....	20
Figura 12. Distribución de los artículos por área de investigación según la Web of Science....	29
Figura 13. Distribución de las publicaciones por autor, según la Web of Science.....	29
Figura 14. Resultados de búsqueda en Web of Science, por institución.....	30
Figura 15. Distribución de las publicaciones por país de origen, según la Web of Science.....	30
Figura 16. Publicaciones por citas recibidas y año de publicación. Web of Science.	31
Figura 17. Publicaciones por citas recibidas y año de publicación para el período 2013-2020. Web of Science.....	32
Figura 18. Distribución de los artículos por área de investigación para el período 2013-2020 en Web of Science.....	33
Figura 19. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por autor.....	33
Figura 20. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por Institución.	34
Figura 21. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por país.	34

Figura 22. Esquema de las configuraciones de partida en el método STATIS y STATIS dual. Adaptado de Vicente-Galindo (2013).....	53
Figura 23. Estructura de los datos para el método STATIS.....	55
Figura 24. Estructura de datos tipo evaluadores, para la aplicación del método STATIS.....	55
Figura 25. Matriz X correspondiente a la concatenación de las H matrices X_h	56
Figura 26. Descripción general del STATIS.....	64
Figura 27. Estructura de datos en STATIS-dual.....	65
Figura 28. Esquema de los procedimientos del STATIS-dual.....	69
Figura 29. Esquema del Método X-STATIS o PTA, adaptado de Amores et al, 2017	73
Figura 30. Esquema del método STATICO, adaptado de Thioulouse, 2011.....	79
Figura 31. Esquema de la técnica CANOSTATIS	90
Figura 32. Esquema del procedimiento COSTATIS, adaptado de Thioulouse, 2011.	95
Figura 33. Diagrama del método SIR-STATIS (adaptado de Villa-Vialaneix, et al , 2016).....	108
Figura 34. Proceso del método DiDISTATIS, adaptado de Kriegsman et al., 2015.....	117
Figura 35. Programa ade-4, Thioulouse et al. (1997).....	133
Figura 36 Métodos STATIS en el programa MULTBILOT. Vicente-Villardón, J.L, (2014)	134
Figura 37. Diagrama de los pasos del Sparse STATIS-dual.....	162
Figura 38. Estructura de la base de datos	168
Figura 39. Interestructura del análisis STATIS-dual.....	171
Figura 40. Proyección del espacio Compromiso, análisis STATIS-dual.....	172
Figura 41. Interestructura generada por el análisis Sparse STATIS-dual	172
Figura 42. Proyección de las variables sobre el compromiso penalizado, Sparse STATIS-dual.	173
Figura 43. Proyección de las variables y variables sobre el compromiso penalizado, Sparse STATIS-dual.	174

INDICE DE TABLAS

Tabla 1. Revisión bibliográfica de publicaciones con aplicación y desarrollos teóricos sobre los métodos de la Familia STATIS (2012-2020) Aplicaciones a diferentes áreas de investigación	41
Tabla 2. Revisión bibliográfica de publicaciones con aplicación y desarrollos teóricos sobre los métodos de la Familia STATIS (2012-2020). Propuestas de nuevos modelos	51
Tabla 3. Paquetes en el lenguaje R que aplican las técnicas de la familia STATIS	137
Tabla 4. Comparativa de los resultados que arrojan las librerías que realizan Stasis* y/o Stasis-dua**	150
Tabla 5. Descripción de Variables	169
Tabla 6. Cargas de las tres primeras componentes principales obtenidas mediante el STATIS-dual y el método de regularización Elastic Net.....	175
Tabla 7. Trabajos de Investigación en proceso de Publicación	197
Tabla 8. Participación en Congresos Nacionales e Internacionales.....	198

INTRODUCCIÓN

En la actualidad, el aprendizaje estadístico se ha transformado en un campo de vital importancia en muchas áreas científicas, en banca, mercadeo, finanzas y otras disciplinas comerciales, debido a la expansión de los problemas de "Big Data". Produciendo esto, una mayor demanda de especialistas en aprendizaje estadístico, (Hastie, Tibshirani, & Friedman, 2009; Wu & Vapnik, 1999).

El aprendizaje estadístico, es un subcampo de reciente desarrollo en estadística, este nuevo subcampo se asocia con desarrollos semejantes en informática y, específicamente con el aprendizaje automático o machine learning. Este aprendizaje alude al conjunto de herramientas para modelar y comprender conjuntos de datos complejos. De acuerdo a James et al., (2013), abarca muchos métodos, como la regresión LASSO y la regresión dispersa (mejor conocida como SPARSE), los árboles de clasificación y regresión, y los bootsting (para crear reglas de predicción) y máquinas de soporte de vectorial (SVM's).

El aprendizaje automático y los métodos de aprendizaje profundo se están integrando en los procesos de análisis de datos. El aprendizaje automático utiliza técnicas estadísticas para dar a los sistemas informáticos la capacidad de "aprender" con datos, (Dill, 2019; Sagioglu & Sinanc, 2013).

El término 'Big Data' aparece en muchos contextos, comprende desde finanzas, negocios, meteorología, investigación biológica, ambiental, genómica, y simulaciones físicas complejas, incluso atención médica. Big Data se refiere a

flujos de datos de mayor volumen, velocidad, variedad, valor y veracidad. Es así que, la capacidad de procesar efectivamente conjuntos de datos masivos, se ha vuelto en parte integral de una amplia gama de disciplinas científicas y académicas, (Sowmya & Suneetha, 2017).

En la era del Big Data, la necesidad de contar con un conjunto de técnicas utilizadas y desarrolladas por estadísticos e informáticos es esencial y de vital importancia. En los últimos años, el progreso en el aprendizaje estadístico ha estado marcado por la creciente disponibilidad de software potentes como el sistema R.

Aunque el término aprendizaje estadístico es bastante nuevo, muchos de los conceptos que subyacen en el campo se desarrollaron hace mucho tiempo. Inicia, a principios del siglo XIX, con los trabajos de Legendre (1805) y Gauss (1855) sobre el método de los mínimos cuadrados, Fisher (1936) propuso un análisis discriminante lineal. La regresión logística, se presenta como otra alternativa en la década de 1940, con los trabajos de Joseph Berkson (1944), entre otros científicos. A inicios de la década de 1970, Nelder y Wedderburn (1972) acuñaron el término modelos lineales generalizados para toda una clase de métodos de aprendizaje estadístico que incluyen tanto la regresión lineal como la logística. Es así que, a finales de la década de 1970, se contaba con muchas técnicas para “aprender” de los datos. Casi todos los métodos eran

exclusivamente lineales, ya que computacionalmente, el ajuste de relaciones no lineales no era posible en aquella época. Con el desarrollo de la informática, en la década de 1980, fue viable utilizar los métodos no lineales. En esta misma década, Breiman, Friedman, Olshen y Stone (1984) introdujeron árboles de clasificación y regresión. Hastie y Tibshirani (1986), introducen el término modelos aditivos generalizados para una clase de extensiones no lineales a modelos lineales generalizados, y también proporcionaron una implementación práctica de software.

En este contexto, el aprendizaje estadístico conforma un conjunto de herramientas para comprender los datos, las cuales pueden clasificarse como supervisadas o no supervisadas. El aprendizaje estadístico supervisado conlleva la construcción de un modelo estadístico para predecir o estimar un producto basado en una o más entradas; Se aplica en diversas áreas como lo son las finanzas, la medicina y políticas públicas entre otras. Mientras tanto en el aprendizaje estadístico no supervisado, hay entradas, pero no salidas de supervisión; No obstante, permite “aprender” las relaciones y la estructura de dichos datos, (Hastie, Tibshirani, & Friedman, 2009; James et al., 2013; McAfee & Brynjolfsson, 2012).

Para descifrar totalmente la información contenida en los datos, el Big data demanda de técnicas de visualización complejas y avanzadas.

En este contexto, los métodos de análisis de tablas de tres vías son de gran importancia en el análisis estadístico multivariante. Este tipo de tablas presenta una primera entrada, vía o índice, que identifica a los individuos que son objeto de estudio, una segunda entrada para las variables que se han medido sobre esos individuos, y una tercera entrada para las diversas situaciones (condiciones u ocasiones) en las que se realizaron mediciones. La finalidad es analizar las diferencias y semejanzas entre los diferentes escenarios a través de las configuraciones de los individuos y de las relaciones entre los diferentes grupos de variables.

Específicamente estudiaremos el método STATIS (Escoufier, 1980; L'Hermier des Plantes, 1976; Lavit, Escoufier, Sabatier, & Traissac, 1994) y su variante STATIS dual, los cuales son una generalización del análisis de componentes principales (ACP) y por ende métodos de aprendizaje no supervisado. STATIS se usa en diversos campos, como la evaluación sensorial, imágenes moleculares, imágenes cerebrales, economía, sostenibilidad, ecología y quimiometría, entre otras muchas áreas.

En este sentido, en referencia al análisis estadístico multivariante, y con el propósito de contribuir con el desarrollo del análisis de voluminosos y complejos conjuntos de datos; y al mismo tiempo, hacer un nuevo aporte a la familia de los

métodos STATIS, hemos elaborado este trabajo de investigación el cual se divide en cuatro capítulos.

En el Capítulo 1 se presenta un panorama general del análisis de datos en tres modos, con la finalidad de familiar al lector con la terminología y notación de estos tipos de arreglos. De igual manera, se introduce los modelos para el análisis de datos en tres modos.

Es así que, el capítulo 1 se estructura en tres partes: en la primera se realiza una explicación de los arreglos de tres vías, la notación correspondiente, las diferentes formas de arreglos y la matriciación de un arreglo.

La segunda parte del capítulo corresponde a los datos en tres modos, aquí se definen los conceptos y se presentan las formas en que se clasifican los datos en tres modos: Datos en tres vías y Datos de Conjuntos Múltiples. Además, se presenta la notación correspondiente a cada caso.

En la tercera y última parte corresponde a los métodos de análisis de datos en tres modos. Aquí se presentan las diferentes clasificaciones de los métodos y se hace una descripción breve de los métodos de acuerdo a la escuela de procedencia: Anglosajona y Francesa. Adicionalmente se presentan los métodos de la Escuela Salmantina, que ha realizado importantes aportes, formulando técnicas de análisis de datos en tres modos con Biplot.

El Capítulo 2 se hace una revisión de los métodos de la familia STATIS, con la finalidad de presentar los desarrollo iniciales y recientes de estos métodos. Primeramente se presenta una revisión bibliográfica con las publicaciones más recientes de estos métodos, actualizando la realizada por [Vicente-Galindo \(2013\)](#). Encontrándose 176 publicaciones, de las cuales 82 corresponden al período 2013-2019. Dentro de estas se encontraron nuevos aportes a los métodos STATIS, los cuales son el STATIS-LDA, INTERSTATIS , SIR-STATIS, DIDISTATIS y CLUSTATIS.

Seguidamente, en la segunda parte del capítulo, se presenta la notación del método STATIS, para familiarizar al lector con los conceptos utilizados.

En la tercera parte se presentan cada uno de los métodos de la familia STATIS como lo son: STATIS y STATIS-dual, X-STATIS (o PTA), STATIS-CoA COVSTATIS, DISTATIS, CANOSTATIS (Canonical-STATIS), Power-STATIS, ANISOSTATIS, $(k+1)$ STATIS (o STATIS externo), Double-STATIS (Do-ACT), STATICO, COSTATIS, STATIS-4, Kernel-STATIS, STATIS-LDA, INTERSTATIS, HiDiSTATIS, DiDiSTATIS, SIR-STATIS, CATATIS, STATICO-CoA, CLUSTATIS. Se presenta el objetivo, y la metodología de cada uno de ellos.

En el capítulo tres se presenta una primera parte con la revisión de los programas conocidos que contienen alguna técnica STATIS como lo son: SPAD,

ade4 y el MultBiplot desarrollado por la escuela salmantina. Se hace primordial énfasis en los paquetes del lenguaje R que realizan análisis STATIS: MExPosition, DistatisR, MultbiplotR, Multigroup, SIRSTATIS, DiDiSTATIS, ade4, dad y ClustBlock.

En el capítulo 4 se presenta el aporte de nuestra investigación. En la primera sección se presentan una introducción a las SPARSE en datos de tres vías, presentando las técnicas de tres vías a las cuales se le han aplicado restricciones. En la segunda sección se presenta **SPARSE-STATIS dual**, nuestra propuesta basada las restricciones vía Elastic Net, a la que denominamos *STATIS dual restringido vía Elastic Net*. En la tercera sección se presenta nuestro algoritmo en R y la cuarta sección corresponde a las simulaciones hechas aplicando nuestro algoritmo y comparándolos con otros algoritmos.

Por último, se presentan las conclusiones de nuestro trabajo y las referencias bibliográfica utilizadas.

CAPITULO I: MÉTODOS PARA ANÁLISIS DE DATOS DE TRES MODOS

1.1. INTRODUCCIÓN

Antes de presentar los métodos de [Análisis de datos en tres modos](#), haremos una breve introducción de algunos elementos esenciales para la comprensión de los modelos para datos de múltiples modos, estos modelos son la extensión del análisis de datos bidireccional a conjuntos de datos de orden superior.

En muchas ocasiones, las matrices bidireccionales no son suficientes para representar toda la información en los datos y las técnicas estándar de análisis bidireccional que se aplican generalmente a dichas matrices, pueden no encontrar las estructuras subyacentes en conjuntos de datos multimodales. Por esta razón, el análisis de datos de múltiples modos se ha convertido en una herramienta de análisis exploratorio, útil para descubrir estructuras en conjuntos de datos de orden superior, en los casos en donde los datos tienen más de dos modos.

Se ha demostrado en numerosas áreas de investigación, incluidas la minería de textos ([Chew, Bader, Kolda, & Abdelali, 2007](#)), neurociencia ([Estienne, Matthijs, Massart, Ricoux, & Leibovici, 2001](#)), las redes sociales ([Acar, Çamtepe, Krishnamoorthy, & Yener, 2005](#)), y el análisis de [procesos](#) ([Gourvéneq, Stanimirova, Saby, Airiau, & Massart, 2005](#)), que el contenido de información subyacente de los datos puede no ser capturado con precisión o identificado de forma única por el análisis de datos de dos vías. En este sentido, el análisis de datos de múltiples modos es de gran utilidad para extraer estructuras ocultas y capturar correlaciones subyacentes entre variables en una matriz de múltiples modos.

En este sentido, cabe hacer una explicación de las palabras vía (way) y modo (mode). De acuerdo a [Kroonenberg, \(2016\)](#) el término vía (way), se refiere al número de dimensiones de un cubo de datos: una vía se refiere a un vector, dos vías a una matriz y tres vías a un arreglo, los cuales conforman el cubo de datos. Por su parte el término, modo (mode), se refiere al contenido de una vía. En el campo de las ciencias sociales y del comportamiento, una matriz de datos bidireccional se conforma por filas (sujetos) y columnas (variables), de la misma manera muchos cubos de datos de tres vías están formados por filas (sujetos), columnas (variables) y tubos (condiciones o puntos de tiempo). En este caso los

datos son de tres modos y tres vías. Empero, un conjunto de matrices de correlación es de tres vías, pero solo de dos modos, ya que las filas y columnas están formadas por las mismas entidades, en su mayoría variables.

A continuación, procedemos a presentar en forma más amplia los conceptos señalados anteriormente.

1.2. ARREGLOS MULTIVÍA

A los arreglos multivía, se les llama también *tensores*. Este tipo de arreglos son generalizaciones de vectores y matrices de orden superior.

Un tensor captura arreglos de diferentes tamaños, así por ejemplo un vector de orden I es un tensor en \mathbb{R}^I , una matriz $I \times J$ es un tensor $\mathbb{R}^{I \times J}$, mientras que un arreglo de tres vías $I \times J \times H$, es un tensor en $\mathbb{R}^{I \times J \times H}$. Así tenemos que un arreglo de N -vías $I_1 \times I_2 \times \dots \times I_N$, es un tensor en $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, este es un arreglo de orden superior y se representa como $\underline{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, mientras que un vector es un arreglo de orden 1 y una matriz un arreglo de orden 2. Es así que un tensor es un arreglo multidimensional.

La terminología que utilizaremos para la representación de datos en tres modos será la presentada por [Kiers \(1988\)](#). De acuerdo a [Tucker \(1963, 1964\)](#), cada dimensión de un arreglo en multivía se denomina **modo o vía** y el número de variables en cada modo indica la dimensionalidad del modo. El arreglo que mencionamos anteriormente, $\underline{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, es un arreglo multivía con N vías con I_1 dimensiones en el primero modo, I_2 dimensiones en el segundo modo, hasta I_N dimensiones en el n -ésimo modo. Cada elemento de $\underline{\mathcal{A}}$ se denota por a_{i_1, i_2, \dots, i_N} .

1.2.1. ARREGLOS DE TRES VÍAS

Un arreglo de tres vías, un caso especial de arreglo multivía en donde $N=3$, se representa como $\underline{\mathcal{X}} \in \mathbb{R}^{I \times J \times H}$. En este caso, cada elemento de $\underline{\mathcal{X}}$ se denota por x_{ijh} , con $i=1, \dots, I$, modo A, $j=1, \dots, J$, modo B y $h=1, \dots, H$, modo C, la [Figura 1](#) se presenta un diagrama de un arreglo en tres vías.

En ocasiones un arreglo de tres vías se considera como un grupo de vectores. Así, se pueden establecer tres conjuntos de dichos vectores que se denomina *fibras* (“*fibers*”). Las *fibras verticales* se extienden sobre las entidades del modo A (vector \mathbf{JH} de orden $1 \times I$) y representan las columnas de la matriz; las *fibras horizontales* se extienden sobre las entidades del modo B (vector \mathbf{IH} de dimensión $1 \times J$) y representan las filas; y *fibras de profundidad*, que se extienden sobre las entidades del modo C (vector \mathbf{IJ} de dimensión $1 \times H$), ver [Figura 2](#).

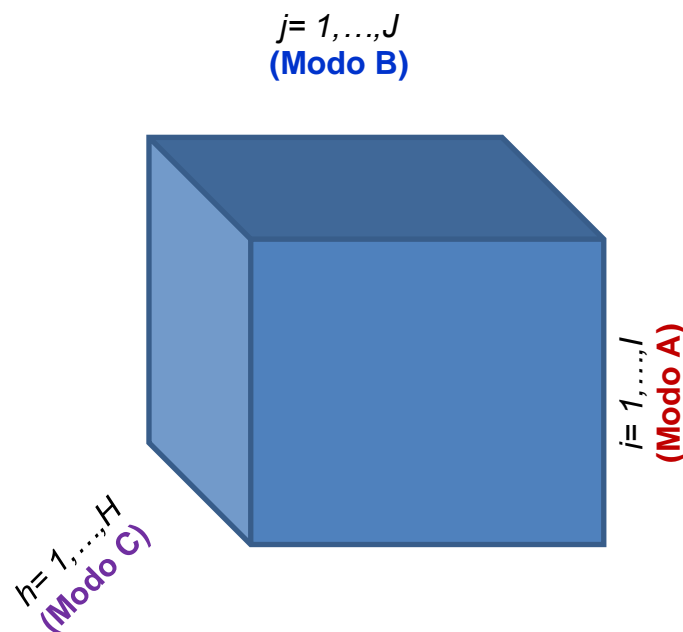


Figura 1. Esquema de arreglo de tres vías o tensor de tercer

$$\text{orden } \underline{\mathcal{X}} \in \mathbb{R}^{I \times J \times H}.$$

Los vectores que hemos mencionado en el párrafo anterior, se conocen como *vectores de modo n*. Se asocian con espacios vectoriales particulares, que se denominarán *espacios de modo n*. Por lo tanto, para una matriz de tres vías, el espacio de modo A es el (sub) espacio comprendido por todas las fibras verticales, el espacio de modo B es el (sub) espacio compendiado por todas las fibras horizontales y el espacio de modo C es el (sub) espacio compendiado por todas las fibras de profundidad. En este caso todas las fibras se consideran vectores columnas (Kolda & Bader, 2009).

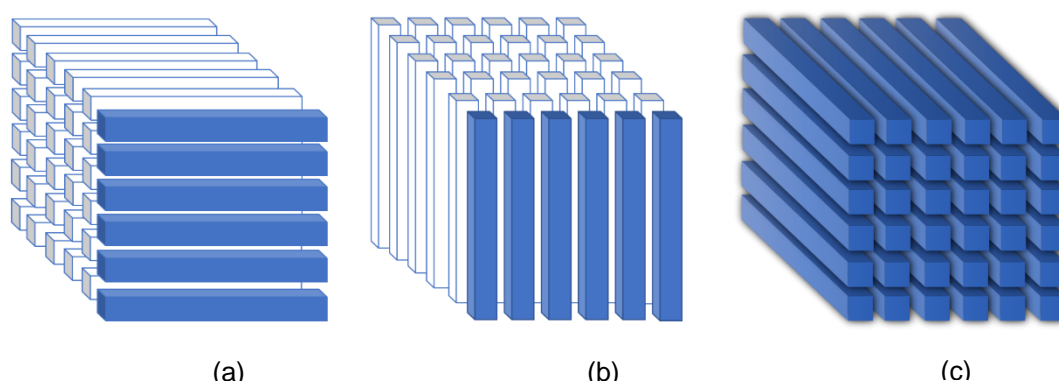


Figura 2. Esquema de arreglos de tres vías.

(a) filas $x_{i:h}$; (b) columnas $x_{:jh}$; (c) profundidad x_{ij} : (Adaptado de Bader & Kolda, 2006)

Por otro lado, un arreglo de tres vías se puede representar en función de un grupo de matrices bidireccionales que se denomina “cortes” o “caras” (“slices”). La partición de los datos consiste en fijar el índice de uno de los modos y variar los índices de los otros dos modos. Los cortes horizontales, laterales y frontales del arreglo de tres vías están constituidos, respectivamente, por las matrices $X_{i::}$, $X_{:j}$ y $X_{::h}$, como se muestra en la Figura 3. Cada uno de estos cortes representan una entrada o modo. Los cortes horizontales pertenecen a las entidades I del modo A (I matrices de dimensión $J \times H$), los cortes laterales pertenecen a las entidades J del modo B (J matrices de dimensión $I \times H$), y los cortes frontales pertenecen a las entidades H del modo C (H matrices de dimensión $I \times J$).

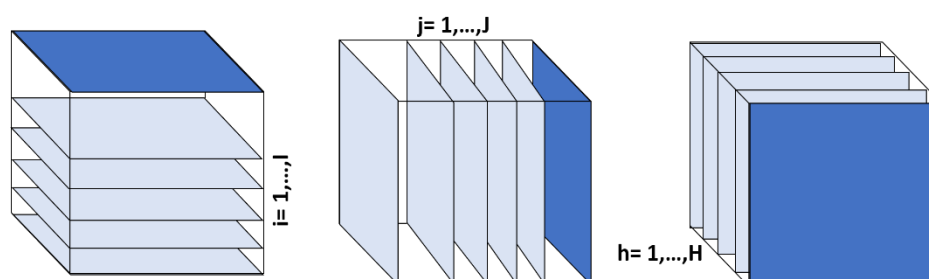


Figura 3. Cortes o caras de arreglos de datos de tres vías.

(a) Corte Horizontal- $X_{i::}$; (b) Corte Lateral $X_{:j}$; (c) Corte Frontal $X_{::h}$ o X_h

1.2.2. MATRICIACIÓN

La matriciación es un proceso por el cual se reordenan los elementos de un tensor de orden N en una matriz, este proceso también se conoce como unfolding (ver Figura 4). Para el caso de un arreglo de tres vías la matriciación puede realizarse en tres formas diferentes:

Matriciación del modo A: se denota como $X_{(a)}$, en este caso se conserva el primer modo (modo A), individuos en las filas, a la vez que el modo B(variables) y C (condiciones u ocasiones) se combinan en las columnas de la matriz.

Matriciación del modo B: se denota como $X_{(b)}$, en este caso se conserva el segundo modo (modo B), las variables en las filas, a la vez que el modo A (individuos) y C (condiciones u ocasiones) se combinan en las columnas de la matriz.

Matriciación del modo C: se denota como $X_{(c)}$, en este caso se conserva el tercer modo (modo C), las condiciones en filas, a la vez que el modo A (individuos) y C (condiciones u ocasiones) se combinan en las columnas de la matriz.

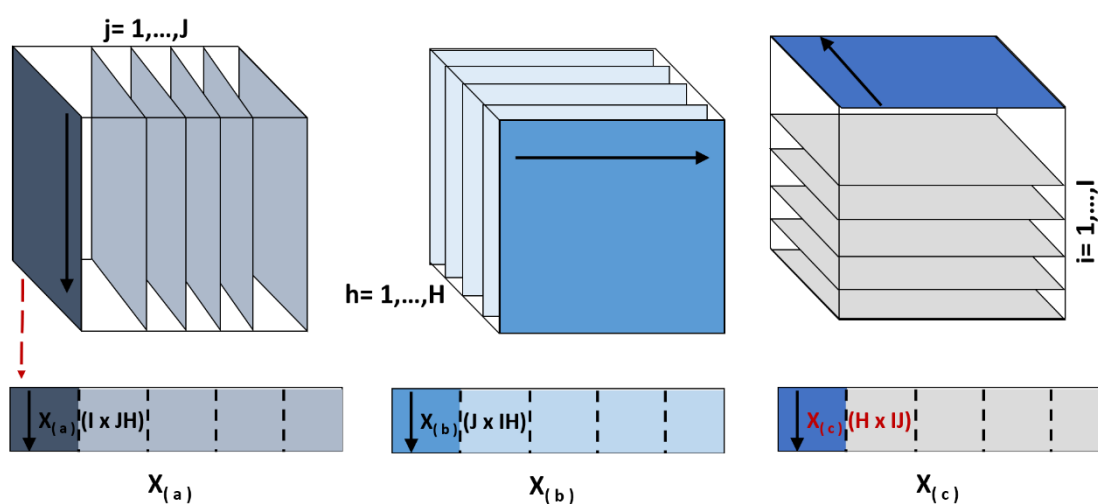


Figura 4. Matriciación de Arreglos de Tres Vías

1.3. DATOS DE TRES MODOS

De acuerdo a [Kiers \(1988\)](#), los datos de tres modos son observaciones sobre elementos que se clasifican de acuerdo con las *categorías* a las que pertenecen de tres **modos** diferentes. Estos modos pueden ser un conjunto de **unidades de observación** (denominadas **objetos**), un conjunto de **variables**, un conjunto de **ocasiones**, un conjunto de evaluadores, etc.

Un arreglo de tres modos, es aquel en donde se tiene una primera entrada, vía u índice que identifica a los **individuos** que son objeto de estudio, una segunda entrada para las **variables** que se han medido sobre esos individuos, y una tercera entrada para las diversas **situaciones** (condiciones u ocasiones) en las que se realizaron mediciones. La finalidad es analizar las diferencias y semejanzas entre los diferentes escenarios a través de las configuraciones de los individuos y de las relaciones entre los diferentes grupos de variables. Es así que el primer modo se refiere a objetos, el segundo modo se refiere a las variables y el tercer modo se refiere a las ocasiones. Los datos longitudinales, es decir, observación repetida de las mismas variables en el mismo conjunto de objetos, son un ejemplo de datos de tres modos.

Siguiendo con el concepto presentado por [Kiers \(1988\)](#), existen diferentes tipos de datos de tres modos, dos de ellos son: datos de conjuntos múltiples y datos de tres vías, los cuales procedemos a presentar.

1.3.1. DATOS DE TRES VÍAS

Se refiere a un conjunto de datos correspondientes a las observaciones de todos los objetos, en todas las variables, y en todas las ocasiones. Este tipo de datos luego pueden ser representados a través de un arreglo de tres vías completo ([ver Figura 5](#)).

De acuerdo a [Carroll y Arabie \(1980\)](#), los datos en arreglos de tres vías se clasifican en:

Datos de tres-vías tres-modos: tres diferentes conjuntos de entradas. Por ejemplo, individuos x variables x ocasiones .

Datos de tres-vías dos-modos: dos de los modos pertenecen al mismo conjunto de entradas. Por ejemplo, matrices de correlación para las mismas variables de

varias muestras diferentes, donde las variables y las muestras son los dos modos, (Kroonenberg, 2008).

Datos de tres-vías un-modo: todos los modos pertenecen al mismo conjunto de entradas.

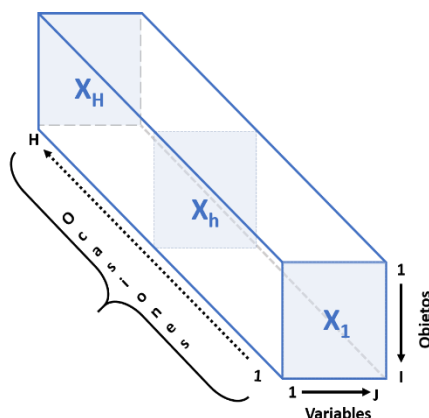


Figura 5. Diagrama de Datos de Tres Vías

Así tenemos que la notación que se utiliza para representar los arreglos de datos de tres vías es la siguiente:

x_{ijh} , elemento del arreglo con $i=1, \dots, I$, (objetos o individuos); $j=1, \dots, J$, (variables); y $h=1, \dots, H$, ocasiones o situaciones.

X_h matriz $I \times J$ que contiene los elementos de los I objetos en las J variables de la ocasión o condiciones H . Es la matriz que representa el corte frontal de un arreglo de tres vías.

1.3.2. DATOS DE CONJUNTOS MÚLTIPLES

Se refiere a las observaciones de diferentes conjuntos de objetos (individuos, lugares, etc.) en el mismo conjunto de variables. Por comodidad, se supone que los diferentes conjuntos de objetos se miden en diferentes ocasiones (escenarios, tiempo, etc.). Los datos de conjuntos múltiples consisten en observaciones en diferentes conjuntos de objetos, por lo que no es posible visualizar datos de conjuntos múltiples mediante un arreglo de tres vías.

Los datos de conjuntos múltiples se representan reuniendo las observaciones de cada conjunto de objetos en una matriz de objetos por variables, y representar

las matrices de datos resultantes en una super-matriz que contiene las matrices para todos los conjuntos de objetos una debajo de la otra, (ver Figura 6).

La notación para representar los datos de conjuntos múltiples es la siguiente:

I_h , denota el número de objetos observados en la ocasión H, ya que no necesariamente las ocasiones tienen el mismo número de objetos.

x_{ijh} , elemento del arreglo con $i=1,\dots,I_h$, (objetos o individuos); $j=1,\dots,J$, (variables); y $h=1,\dots,H$, ocasiones o situaciones.

X_h matriz $I_h \times J$ que contiene los elementos de los I_h objetos en las J variables de la ocasión o condiciones H. Es decir que las matrices, no tienen necesariamente el mismo orden.

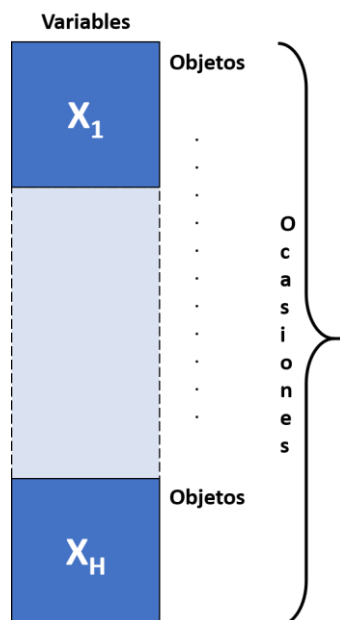


Figura 6. Diagrama de Datos de Conjuntos Múltiples.

Dentro de este tipo de datos también se consideran aquellos en donde se tienen datos de observaciones de un conjunto de objetos en diferentes conjuntos de variables. Su notación será la siguiente:

I , denota el número de objetos observados.

x_{ijh} , elemento del arreglo con $i=1,\dots,I$, (objetos o individuos); $j=1,\dots,J_h$ (variables en cada ocasión); y $h=1,\dots,H$, ocasiones o situaciones.

X_h matriz $I \times J_h$ que contiene los elementos de los I objetos en las J_h variables de la ocasión o condiciones H , (ver Figura 7).

Como se observa en la [Figura 7](#), los datos de conjuntos múltiples no se consideran de tres vías, dado que no conforman un cubo completo de datos, podemos tener el caso (a) con diferente número de objetos en cada ocasión, y (b) diferentes variables en cada ocasión.

Como señalamos anteriormente, Kiers (1991) indica que los datos de conjuntos múltiples no son datos en tres vías, debido a que con configuran un cubo de datos completos, una excepción a esto son los datos longitudinales. A este respecto, Kiers (1988, 1991) señala que, a los datos de conjuntos múltiples, se les pueden aplicar los mismos métodos que a los datos de tres vías, si se utilizan matrices derivadas de las matrices originales, como lo son las matrices de productos cruzado o de productos escalares.

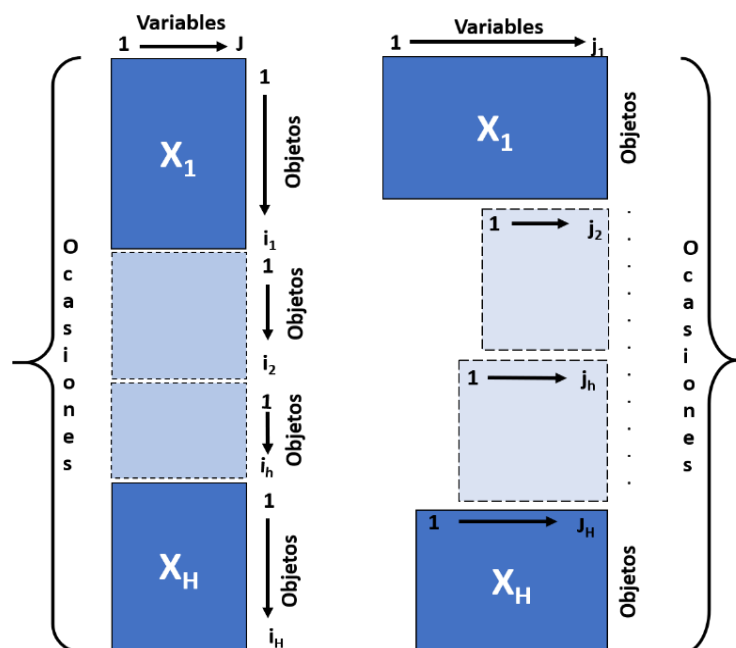


Figura 7. Diagramas de conjuntos múltiples.

1.4. MÉTODOS PARA ANÁLISIS DE DATOS DE TRES MODOS

Los estudios de integración de matrices tiene sus orígenes en el trabajo de [Hotelling, \(1936\)](#) , sobre Análisis Canónico, la técnica tiene como objetivo buscar relaciones entre conjunto de variables a partir de ejes canónicos. Más adelante con los trabajos de [Carroll \(1968\)](#) y [Kattenring \(1971\)](#), sobre Análisis Canónico Generalizado, se hace una extensión a más de dos conjuntos de variables.

A este respecto, en lo referente al estudio de integración de matrices en la literatura se mencionan varios enfoques. Así tenemos, que desde el punto de vista de su origen se clasifican en Escuela Anglosajona y Escuela Francesa. Otra de las formas que plantea [Kiers \(1991\)](#), para diferenciar los métodos, es la del ajuste de las matrices. Así tenemos que hay métodos que ajustan el modelo directamente a los datos de tres vías, y otros ajustan el modelo a matrices derivadas de tres vías. Asimismo, Kiers clasifica los métodos de análisis de datos de tres modos en métodos simétricos y asimétricos (ver [Figura 8](#)). Mientras que [Baccalá \(2004\)](#) clasifica los métodos por un lado los basados en la construcción de una configuración consenso óptima y los que se apoyan en la existencia de modelos subyacentes a la tabla de datos.

Aunque el objetivo de nuestro trabajo no es hacer un análisis detallado de cada uno de los métodos de análisis de tres modos, es importante conocer, en forma breve, como se han desarrollado. A continuación, haremos una explicación concisa de los métodos más conocidos y presentaremos un diagrama resumen de acuerdo a las clasificaciones que hemos señalado anteriormente.

Así tenemos que, de la Escuela Anglosajona se cuenta con los modelos de la familia Tucker y de la familia PARAFAC, mientras que de la Escuela Francesa se tienen los modelos de la familia STATIS, el Análisis Factorial Múltiple (AFM) entre otros muchos.

Desde el punto de vista de las formas de “ajuste” tenemos que cuando se trabaja directamente con las matrices originales, se habla de **Ajuste Directo (Direct Fitting)** y cuando se utilizan matrices derivadas de las originales, se tiene el **Ajuste de Matrices Derivadas (Derived Data Fitting)**, tal es el caso de las matrices de productos cruzados.

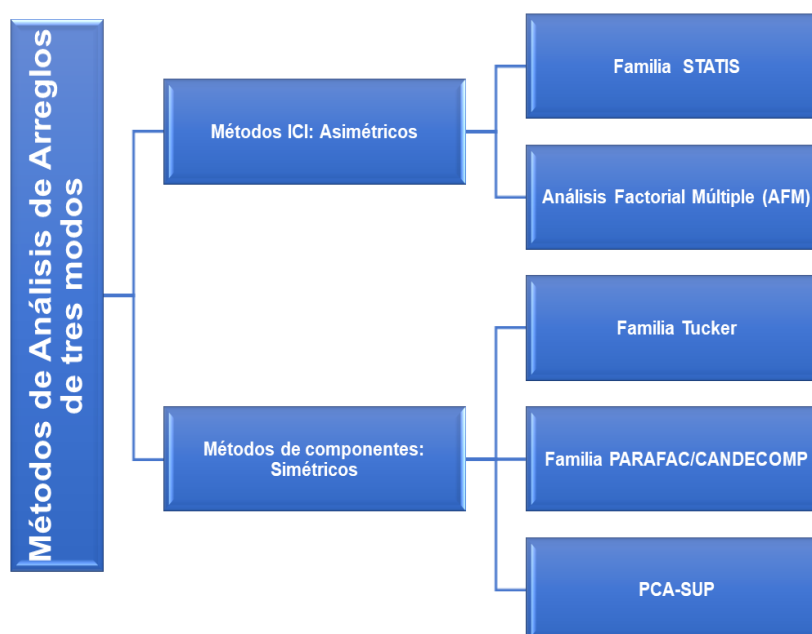


Figura 8. Clasificación de los modelos para el análisis de datos en tres modos de acuerdo a Kiers (1991)

Por otro lado, los métodos asimétricos son aquellos en donde uno de los modos (ocasiones casi siempre) tienen un tratamiento diferente a los otros dos modos considerados (individuos y variables), en el caso de los simétricos todos los modos son tratados de igual forma. Los métodos simétricos se identifican con los llamados **Métodos de componentes**, estos métodos también se caracterizan por ajustar modelos que representan, el dato original, de la forma más confiable posible. Por otro lado, los métodos asimétricos, tienen un enfoque de **Interestructura-Compromiso-Intraestructura (ICI)**, las cuales son las tres etapas que lo conforman.

A continuación, se presenta brevemente una explicación de algunos de los métodos de acuerdo a la escuela a la que pertenece y algunos que, aunque no pertenecen directamente a las escuelas anglosajonas o francesas, se derivan de alguna de ellas.

1.4.1. ESCUELA ANGLOSAJONA

La Escuela Anglosajona se basa en los trabajos de [Tucker \(1966\)](#). Los métodos de esta escuela se centran en las diferencias de las configuraciones de los mismos individuos o las mismas variables en las diferentes ocasiones y son de naturaleza simétrica, se caracterizan por procurar minimizar la función de pérdida de cada uno de ellos, en otras palabras, que los datos reproducidos por el modelo sean lo más semejante a los datos originales.

La terminología que utiliza la Escuela Anglosajona para denominar a las tablas de datos de tres modos es “three-way data arrays”, en el caso de matrices completas, y “three-way o multiway data sets”, para matrices que no se cruzan totalmente en sus tres modos, en forma general, el término “three-mode data” es el utilizado para identificar las tablas de tres modos.

La descomposición Tucker (Three-mode factor analysis), [Tucker \(1966\)](#) (trata los tres modos de una estructura de datos por igual. Introduce una nueva perspectiva basada en la idea de que cada vía está asociada con una estructura factorial subyacente, y que hay tres grupos de factores, no solo uno. Dicho de otra manera, existe una asociación entre las tres entradas de datos (cada una con otra vía no observable), que puede considerarse como un conjunto óptimo de factores o categorías. Los objetos, las variables y las ocasiones se consideran como combinaciones lineales de los objetos idealizados, las variables idealizadas (variables “latentes”) y las ocasiones idealizada (ocasiones prototipo), respectivamente. Las relaciones entre los tres tipos de factores se consideran en una tabla o matriz estimada por el modelo a la cual se le denomina corre array. La descomposición Tucker3 ([Kroonenberg & de Leeuw, 1980](#); [Tucker, 1966](#)), se desarrolló con la finalidad de corregir la desventaja que presentaba el modelo original de Tucker, el cual no proporcionaba aproximaciones de los datos en el sentido de los mínimos cuadrados ordinarios. Dentro de la familia de métodos Tucker, se tiene la descomposición Tucker1 ([Tucker, 1972, 1975](#)), su nombre se refiere al número de dimensiones para las que la matriz de componentes está determinada independientemente de h, es este caso 1 dimensión. Por otro lado, el método Tucker 2 ([Israelsson, 1969](#); [Kroonenberg & de Leeuw, 1980](#); [Tucker, 1975](#)), se denomina así por ser 2 el número de dimensiones para las que se fijan las componentes.

Un caso particular del método Tucker3, en donde se trabaja con las matrices de productos cruzados de variables o matrices de distancias entre individuos, es el modelo de **Escalamiento de Tres Modos (Three Mode Scaling)** (Tucker, 1972), este es un método de ajuste a matrices derivadas, en donde dos de los tres modos coinciden.

Carroll y Chang (1970), y Harshman (1970), en forma independiente, desarrollaron un modelo que descompone las tablas de tres vías en entradas totalmente plegadas de forma más sencilla. Por su lado, Carroll y Chang le dan el nombre de “**CAN**onical **DECOM**position of N-way tables” (CANDECOMP), mientras que Harshman lo llama “**PAR**allel **FACT**or Analysis” (PARAFAC). La diferencia que tienen estos métodos, por un lado el de Carroll y Chang (1970) y por otro el de Harshman (1970), con la idea de Tucker, es que solo muestran la determinación de un único grupo de factores a partir de datos observados, y eso puede ser interpretado por las tres vías, de forma simultánea. El inconveniente reside en el hecho de que los ejes pueden interpretarse en una sola vía, ya que el método no permite la rotación de ejes. Este método puede considerarse como un caso especial del modelo TUCKER2 en el sentido de que, en este, se definen los componentes principales para las variables y los individuos, mientras que en PARAFAC y CANDECOMP, los modelos se definen para ambos simultáneamente. La solución de esto es maximizar la inercia explicada, es decir, minimizar la pérdida de información. Casi paralelamente se introduce el modelo PARAFAC2 (Harshman, 1972), una modificación del PARAFAC. Este nuevo modelo mantiene la mayoría de las características atractivas del modelo PARAFAC, pero es menos restrictivo

Continuando con los métodos de esta escuela tenemos el INDSCAL (Carroll & Chang, 1970), y IDIOSCAL (Carroll & Chang, 1972). El Escalamiento de Diferencias Individuales o INDSCAL por sus siglas en inglés (**IN**dividula **Differences SCAL**ing), es un caso particular del modelo de descomposición CANDECOMP/PARAFAC para tensores de tres vías, los cuales son simétricos en dos de sus modos. Para este método, las dos primeras matrices de la descomposición han de ser iguales.

Por su parte, el método IDIOSCAL (**I**ndividual **D**ifferences In **O**rientation **SCAL**ing), se basa en reducción solamente dos de las tres vías. Es un modelo TUCKER2, pero ligeramente transformado para adaptarse a arreglos de tres

vías con dos modos. Se introduce una restricción consistente en que las matrices de componentes de la primera y segunda vía (de igual modo) son iguales; así que el modelo se ajusta a aquellos casos en donde los datos de la primera y segunda vía (de igual modo) se vinculan a la mismas variables o estímulos, entre otros.

Otra familia de descomposiciones es DEDICOM (**DE**scomposition Into **DI**rectional **COM**ponents), (Harshman, 1978), estos modelos analizan matrices de datos que son intrínsecamente asimétricas. El modelo DEDICOM de tres vías es un caso especial del modelo PARAFAC2. El modelo agrupa **I** objetos en **R** dimensiones latentes y describe sus patrones de interacción.

Kroonenberg y De Leeuw (1980), a partir de la aproximación inicial dada por Tucker (1966), proponen dos algoritmos, a los que llamaron **Tuckals-3** y **Tuckals-2**. Estos surgen debido a que el modelo de Tucker presenta la desventaja de que no proporciona ajuste de mínimos cuadrados.

El modelo **CANDELINC** (**CAN**onical **DE**composition with **LIN**ear Constraints), fue desarrollado para ajustar modelos de múltiples vías bajo restricciones lineales (Carroll, Pruzansky, & Kruskal, 1980). Con este modelo, el arreglo original se comprime al expresar el arreglo en las coordenadas de un subespacio de baja dimensión definido por un conjunto de bases en cada modo. Solo se permiten las bases ortogonales, pero cualquier base no ortogonal puede pasar por un proceso de ortogonalización, antes de la compresión sin pérdida de información.

Los modelos presentados son los más conocidos (ver Figura 9), existen otros modelos de la escuela anglosajona que se basan en estos y que integran ajustes a los mismos, entre los que podemos mencionar la introducción de restricciones en los modelos de tensor como son los modelos PARAFAC restringidos. Así tenemos el modelo PARATUCK2 (Harshman & Lundy, 1996) el cual es una generalización del modelo PARAFAC, que añade algo de la flexibilidad de los modelos de tres modos de Tucker, y al mismo tiempo recoge algunas de las propiedades de singularidad de PARAFAC. Su nombre, PARATUCK2, indica su similitud tanto con el modelo PARAFAC como con el modelo Tucker2. Este modelo es adecuado para una cierta clase de problemas de múltiples vías que involucran interacciones entre factores. Entre otros modelos PARAFAC restringido, se tiene el modelo PARALIND (**PARA**llel profiles with **LIN**ear

Dependencias), (Bahram & Bro, 2007), que conduce a dos interpretaciones diferentes, como un modelo de Tucker restringido cuyo tensor central admite una descomposición PARAFAC con matrices de factores $\Phi^{(n)}$, llamadas "matrices de interacción", y como un modelo PARAFAC restringido con matrices de factores restringidos. Otro modelo es el CONFAC (**CON**strained **FA**ctor decomposition), (de Almeida, Favier, & Mota, 2008), consiste en descomponer un tensor de tercer orden en una suma triple de factores tensoriales de rango uno, donde se permiten las interacciones que involucran los componentes de diferentes factores tensoriales. El patrón de interacción está controlado por tres matrices de restricción cuyas columnas son vectores canónicos. Cada matriz de restricción está asociada con una dimensión dada, o modo, del tensor. El uso explícito de estas matrices de restricción proporciona grados de libertad a la descomposición de CONFAC para modelar señales tensoras con estructuras restringidas que no pueden manejarse con la descomposición de factor paralelo estándar (PARAFAC) de igual manera están los modelos PARAFAC-3, Tucker-3, Tucker-(2,3), PARALIND/CONFAC-3 y PARATUCK-(2,4), para más detalles de estos modelos ver, (Favier & de Almeida, 2014).

Otros modelos con restricciones y de reciente desarrollo son CP-Ridge (Candecomp/Parafac with ridge regularization, (Giordani & Rocci, 2013a) y el CP-Lasso (Candecomp/Parafac with Lasso constraints), (Giordani & Rocci, 2013b).

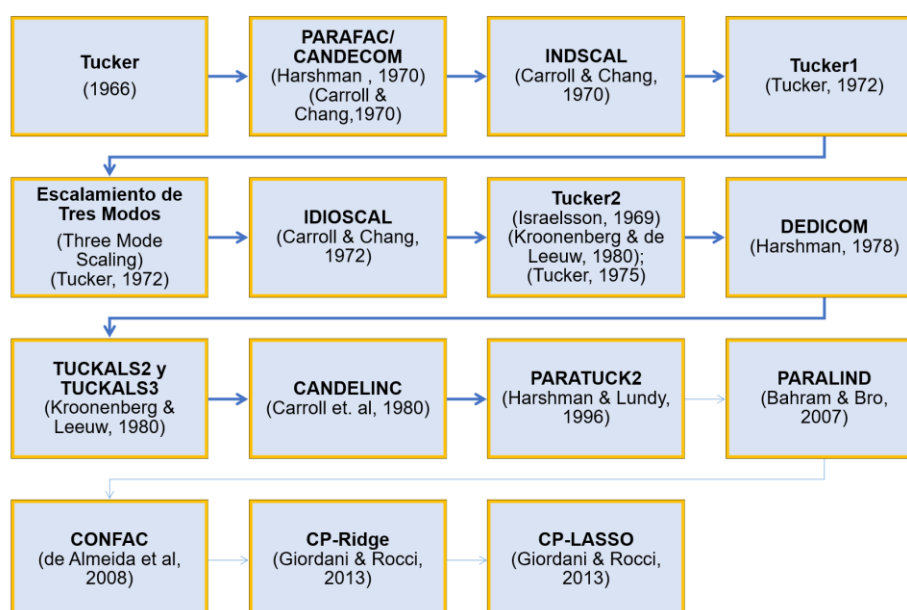


Figura 9. Desarrollo cronológico de algunos modelos de Análisis de datos de Tres Modos de la Escuela Anglosajona

Los métodos de Análisis Procrustes (Gower, 1975; Gower & Hand, 1996); Meta-Componentes Principales (Krzanowski, 1979, 1982) y Análisis de Componentes Principales Comunes (Flury, 1984, 1988); trabajan el problema de la integración de matrices. Estos se fundamentan en encontrar una configuración consenso “óptima”, en el sentido de aproximar, las distintas configuraciones asociadas a cada matriz, lo máximo posible, (Galindo-Villardón, 2007).

1.4.2. ESCUELA FRANCESA

En esta sección haremos una breve introducción de los modelos de la Escuela Francesa. En el capítulo dos de este trabajo se presentarán los métodos de manera más detallada.

La Escuela Francesa identifica los datos de tres modos, generalmente, como Tablas de Tres Índices (*Tableaux à Trois Indices*). Los métodos de la Escuela Francesa más conocidos son los de la familia STATIS, el Análisis Triádico (*Analyse triadique*), y el Análisis Factorial Múltiple (AFM). Estos métodos varían en la forma en que se obtiene un referencial llamado compromiso que permite ubicar en un mismo subespacio los individuos de las diferentes situaciones, así como las variables. También se incluye el método LONGI, para datos longitudinales.

Estos métodos contemplan tres fases o etapas: el análisis de **la interestructura**, el cual es una comparación global de las matrices originales.; la siguiente fase corresponde a la búsqueda de una matriz consenso o **compromiso**, esta matriz se elabora a partir de la concatenación de las matrices originales ponderadas (la diferencia entre los distintos métodos se da, generalmente, en la ponderación seleccionada); la tercera y última corresponde al estudio minucioso de cada elemento, es decir individuos y variables, de las matrices originales, sobre un subespacio común creado en la fase dos, a esta última fase se le denomina análisis de la **infraestructura**

Los métodos de la Familia STATIS, se origina de los trabajos de Escoufier (1980), fueron descritos por primera vez por L'Hermier des Plantes (1976) y desarrollados ampliamente por Lavit, Escoufier, Sabatier y Traissac (1994), se utiliza en el análisis de datos de conjuntos múltiples de naturaleza cuantitativa. Estos métodos se presentan en distintas versiones, basadas en conceptos de

geometría euclídea, entre las que podemos mencionar: *Analyse triadique*, *STATIS* y *STATIS Dual*. En el capítulo 2 presentaremos con más detalle algunos de los más recientes métodos de la Familia STATIS.

El acrónimo STATIS proviene de la expresión en francés "*Structuration des Tableaux à Trois Indices de la Statistique*", es decir, "Estructuración de Tablas Estadísticas de Tres Índices". También se utiliza el acrónimo "ACT", refiriéndose a la expresión "*Analyse Conjointe de Tableaux*" (Análisis Conjunto de Tablas). STATIS (L'Hermier des Plantes, 1976), es esencialmente una generalización del análisis de componentes principales (PCA) y, cuyo objetivo es analizar, varios conjuntos de variables en **el mismo conjunto de observaciones**, o bien, varios conjuntos de observaciones en el **mismo conjunto de variables**, para el caso de STATIS Dual (L'Hermier des Plantes, 1976). Pertenece al grupo de métodos de la familia de PCA de multi tablas, también conocidas como Análisis de bloque múltiple o de Consenso.

El Análisis Triádico o X-STATIS (Jaffrenou, 1978), nos permite el análisis simultáneo de matrices de datos relacionadas a los mismos individuos (filas) y a las mismas variables (columnas). Es necesario el cálculo de las distancias euclídeas entre las tablas de datos, es decir que se debe considerar una tabla como un vector (yuxtaposición de columnas) y utilizar el producto escalar usual entre vectores. Se construye una nube en la que cada punto corresponde a una tabla, a partir de la matriz de productos escalares entre tablas.

El Análisis Factorial Múltiple (AFM), es un método que permite analizar, simultáneamente, varios grupos de variables, medidas sobre el mismo conjunto de individuos, fue desarrollado por Escofier y Pagès (1984). Su característica principal es explorar lo más posible las relaciones intrínsecas entre objetos aparentemente muy diferentes. Además de la obtención de una tipología de los objetos definida a través del conjunto de variables, también se tiene como objetivo la búsqueda de posibles relaciones entre las estructuras obtenidas en el seno de cada uno de los grupos. La *Figura 10* muestra algunos métodos para análisis de datos en tres vías de la escuela francesa.

El método LONGI fue propuesto por Pernin (1986). Se aplica al análisis de datos longitudinales, tomando en cuenta la naturaleza de tres-vías de los datos. Se basa en encontrar combinaciones lineales de las variables, a los cuales se les llama "índices de situación". Estos, maximizan las diferencias entre los individuos

y varían mínimamente, entre los individuos en las diferentes ocasiones. Sin embargo, no puede aplicarse a datos de conjuntos múltiples.

De igual manera se han desarrollado modelos en tres vías para datos binarios como lo son los de la familia de modelos de clases jerárquicas (HICLAS), (De Boeck & Rosenberg, 1988), entre ellos tenemos: el modelo INDCLAS (**IND**ividual **D**ifferences **H**ierarchical **CLAS**ses) (Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999), el modelo TUCKER3-HICLAS (E Ceulemans, Mechelen, & Leenen, 2003), el modelo TUCKER2-HICLAS (Eva Ceulemans & Van Mechelen, 2004) y el modelo TUCKER1-HICLAS (Eva Ceulemans & Van Mechelen, 2005) y para datos acoplados el modelo CHIC (**C**ombined **H**iclas–**I**ndclas model for **C**oupled binary data) (Wilderjans, Ceulemans, & Van Mechelen, 2008).

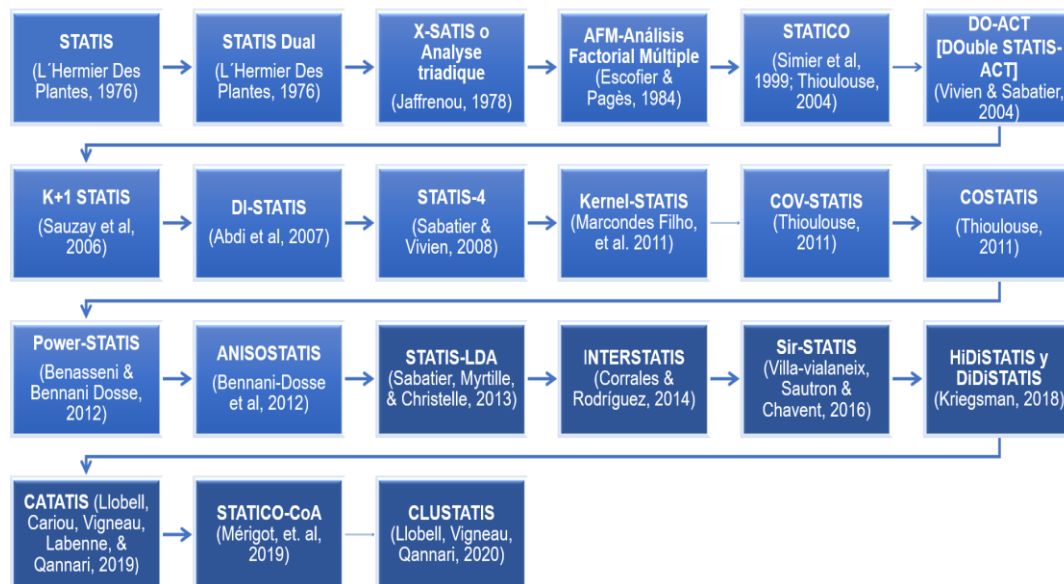


Figura 10. Desarrollo cronológico de algunos modelos de Análisis de datos de Tres Modos de la Escuela Francesa

1.4.3. ESCUELA SALMANTINA

Otra familia de métodos para el análisis de datos en tablas de tres modos son las formulaciones BIPLLOT.

Las factorizaciones BIPLLOT (Gabriel, 1971; Galindo-Villardón, 1986) permiten representar conjuntamente individuos y variables. Estas han dado lugar a nuevos métodos de análisis multivariante de datos, al ser combinado con otras técnicas. El BIPLLOT, ha sido generalizado al caso de varias matrices de datos, podemos mencionar el Biplot Conjunto y Biplot Interactivo (Carlier & Kroonenberg, 1996). En este contexto, el Departamento de Estadística de la Universidad de Salamanca ha contribuido desarrollando nuevos métodos para el análisis de datos en tres modos con formulaciones biplot.

Así tenemos que, con relación a la de búsqueda de una configuración consenso, Martín-Rodríguez, (1996) y Martín-Rodríguez, Galindo Villardón y Vicente-Villardón (2002), hacen una generalización para el caso en el que las configuraciones son el resultado de aplicar un análisis biplot a cada matriz inicial de datos, al que denominan **METABIPLLOT**.

Baccalá (2004) formula una alternativa Biplot al Análisis Factorial Múltiple, al cual denomina **BIPLLOT Múltiple**. El cual permite representar los individuos y los distintos grupos de variables en el mismo espacio, y obtener un patrón de

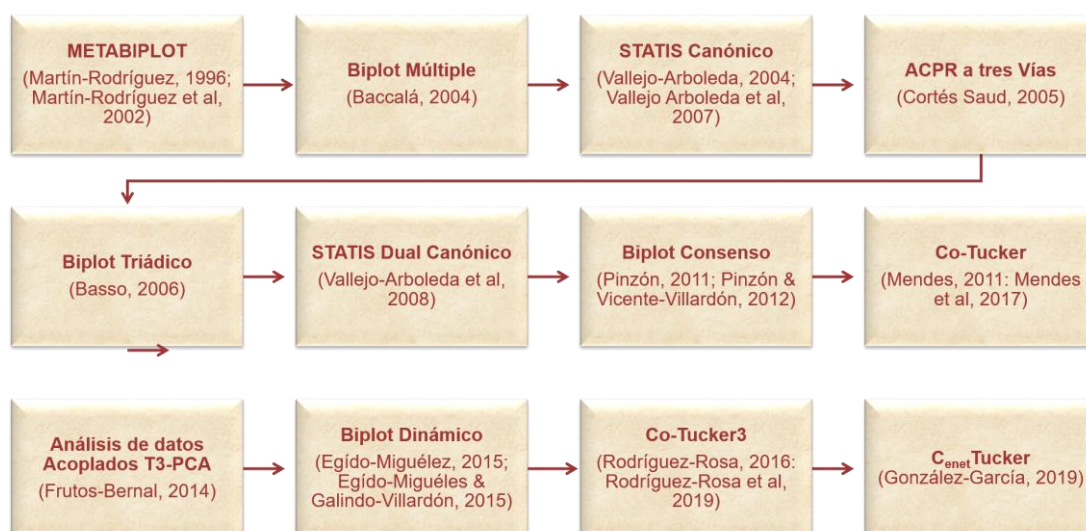


Figura 11. Desarrollos cronológicos de las aportaciones de la Escuela Salmantina al análisis de datos de tres modos

comparación común para todos los grupos, esto hace factible calcular medidas de la calidad de representación para cada uno de los elementos representados. Los elementos de las columnas en la matriz original yuxtapuesta, se reproducen, aproximadamente, con las proyecciones de los marcadores filas sobre los marcadores columnas, lo cual permite el ordenamiento aproximado de los individuos respecto a cada una de las variables.

En esta línea, [Vallejo-Arboleda \(2004\)](#) y [Vallejo-Arboleda, Vicente-Villardón y Galindo-Villardón \(Vallejo-Arboleda, Vicente-Villardón, & Galindo-Villardón, 2007\)](#), plantean un método, al cual llaman **STATIS CANÓNICO**, que se fundamenta en la metodología STATIS, este método posibilita recoger la estructura de grupos de los datos y la evolución de la misma en diferentes ocasiones. El método se complementa por medio de una representación simultánea de grupos, variables y ocasiones (**BIPLOT**), que incrementa las capacidades de los métodos clásicos. Las limitaciones que muestran, por un lado, el MANOVA y/o el Análisis Canónico, que, pese a que recogen la estructura de grupos, no capturan la evolución temporal; y por otro lado el STATIS que si bien, es sensible a la evolución temporal, confunde la variabilidad “entre” y “dentro” de los grupos, son superadas con el **STATIS Canónico**.

La versión dual del **STATIS canónico** es propuesta por [Vallejo-Arboleda et al.\(2008\)](#), se le denomina **STATIS Dual Canónico** y tiene como objetivo buscar un subespacio de referencia común para el análisis de variables canónicas de todas las ocasiones. Tiene la misma estructura del STATIS-ACT dual, en este caso, se parte de las matrices de medias de los grupos en todas las variables para cada ocasión y se toman la inversa de la matriz de covarianza común y la matriz diagonal de tamaños de los grupos como métricas para los grupos y las variables, respectivamente. Por consiguiente, se supone que las matrices de covarianza son homogéneas en todas las ocasiones dentro de los grupos y es preciso obtener un estimador de la matriz de covarianza común dentro de los grupos. El propósito es determinar las combinaciones de variables que maximizan la dispersión entre grupos, relativa a la dispersión dentro de los grupos.

Igualmente, [Basso, \(2006\)](#) hace una aportación al Análisis Triádico, a través de lo que denomina **BIPLOT Tríadico**. Basso formula las representaciones Biplot del Análisis Triádico, proyectando en el sistema de referencia común que

proporcionan las variables o los individuos consenso. Esta nueva formulación, amplía la aplicabilidad del análisis triádico, ya que proporciona, para datos de tres vías, la definición adecuada de un conjunto de trayectorias en el espacio consenso y la posibilidad de definir índices que dejen evaluar la calidad de las representaciones.

El **Biplot Consenso** (Pinzón, 2011; Pinzón & Vicente-Villardón, 2012), propone la obtención de un subespacio de referencia común para todas las tablas, a partir de la proyección sobre subespacios de dimensión reducida, y estudia sus propiedades, generalizando a esta situación los índices de bondad de ajuste de los biplots clásicos.

Otra de las últimas, contribuciones al análisis de tres modos con formulaciones biplot, es el **Biplot Dinámico** (Egido Miguélez, 2015; Egido Miguélez & Galindo-Villardón, 2015), este método cuenta con dos etapas: Primero se realiza un análisis de la situación de referencia (análisis estático), a través de un análisis biplot clásico o bien un HJ-Biplot, ya que este permite una mejor representación simultánea de filas y columnas; en la segunda etapa, se realiza el análisis dinámico, en el cual se proyecta sobre el gráfico biplot (el cual se obtiene en la etapa anterior), el resto de las situaciones que se desean analizar. Como resultado se tiene un conjunto de trayectorias de los individuos y de las variables sobre el biplot de la situación de referencia.

A continuación, presentamos las aportaciones a los métodos de análisis de datos de tres modos, realizados por investigadores españoles, pero basado en los modelos TUCKER.

Frutos-Bernal (2014) propone en su tesis doctoral, un modelo para **Análisis de datos acoplados T3-PCA** el cual analiza un bloque de datos de tres vías y un bloque de datos de dos vías, ambos de tipo continuo, que tienen un modo en común, el modelo tiene como restricción que la matriz de componentes del modo común ha de ser igual en ambos submodelos.

Co-Tucker (Mendes, 2011; Mendes et al., 2017), es otra de las aportaciones para datos en tres modos, con este método primero se construye un cubo de datos formado por tablas de productos cruzados a partir de cada par de matrices para cada repetición, y en segundo lugar se realiza un análisis con un método Tucker. Creado para analizar datos de tipo ecológico, es un enfoque combinado de las técnicas STATICO y Tucker3, aborda el problema de describir no solo la

parte estable de la dinámica de las relaciones estructura-función entre las comunidades y su entorno (en diferentes ubicaciones y / o en diferentes tiempos), pero también las interacciones y cambios asociados con la dinámica de los ecosistemas. Una de las principales ventajas de modelar conjuntos de datos con el modelo **CO-TUCKER** es la ganancia en interpretabilidad.

Co-Tucker3 (Rodríguez-Rosa, Gallego-Álvarez, & Galindo-Villardón, 2019; Rodríguez Rosa, 2016), basado en el método Tucker3 y en el Análisis de la Co-Inercia, resuelve el problema de describir no solo la parte estable de una estructura de datos, sino también la posibilidad de extraer la estructura latente, así como las interacciones entre las tres dimensiones; tomando en cuenta un par de cubos de datos. Su objetivo es descomponer los dos cubos de datos en matrices ortogonales y dos cubos de datos (los core arrays) de orden $P \times Q \times R$ más simple que los originales; y después descubrir las relaciones entre estas filas, columnas y repeticiones reducidas a un número más pequeño de componentes.

Estos métodos han sido desarrollados por investigadores del departamento de Estadística de la Universidad de Salamanca, y aunque se derivan del grupo de métodos de la familia anglosajona y francesa, los agrupamos como de la **Escuela Salmantina**. Otra de las contribuciones de la escuela salmantina al análisis de datos de tres modos, ha sido el **ACPR a tres vías** (Cortés Saud, 2005), que extiende la metodología del ACP restringido (ACPR), a ACPR a tres vías, realizando una aportación metodológica que incorpora información externa de dos vías (información dada en matrices) al análisis de un arreglo de datos de tres vías. En su desarrollo no sólo se obtiene una perspectiva más general del ACPR, sino que se alcanzan resultados interesantes, tales como el Análisis Canónico de Correspondencias (ACC) de tres vías con información adicional en dos vías.

El más reciente aporte, en este contexto, de la escuela salmantina, es el método denominado **C_{enet}Tucker** (González-García, 2019), cuyo objetivo fundamental es el de integrar la penalización Elastic Net (Zou & Hastie, 2005), sobre los vectores de cargas de las matrices de cada modo, de tal manera que se obtienen coeficientes exactamente nulos. Como señala su autora, este es el primer método de descomposición para datos tensoriales que hace la restricción Elastic Net. Con este método no es necesario penalizar todas las matrices de

componentes en forma simultánea. Dependiendo del objetivo de la investigación, se determinará cuales matrices de componentes serán restringidas, (ver [Figura 11](#)).

CAPÍTULO II: METODOS STATIS

2.1. INTRODUCCIÓN

A partir de la información que proporciona la revisión bibliográfica realizada por [Vicente-Galindo, \(2013\)](#), y de la reciente información existente sobre el tema, se realiza una revisión exhaustiva de las técnicas de la familia STATIS para el análisis de tablas de tres vías o modos.

Los métodos STATIS, se atribuyen a los trabajos de [Escoufier \(1980\)](#), y fueron descritos por primera vez por [L'Hermier des Plantes \(1976\)](#) y desarrollados ampliamente por [Lavit et al., \(1994\)](#), se utiliza en el análisis de tablas múltiples de naturaleza cuantitativa. Este método se presenta en distintas versiones, basadas en conceptos de geometría euclídea, entre las que podemos mencionar: STATIS y STATIS-dual ([L'Hermier des Plantes, 1976](#)), X-STATIS o PTA ([Jaffrenou, 1978](#)), STATICO ([Thioulouse, Simier, & Chessel, 2004](#)), Double-STATIS (Do-ACT) ([Vivien & Sabatier, 2004](#)), (k+1)STATIS o STATIS externo ([Sauzay, Hanafi, Qannari, & Schlich, 2006](#)), DISTATIS ([Abdi, Valentin, Chollet, & Chrea, 2007](#)), CANOSTATIS o Canonical-STATIS ([Vallejo-Arboleda et al., 2007](#)), STATIS-4 ([Sabatier & Vivien, 2008](#)), Kernel-STATIS ([Marcondes Filho, Fogliatto, & de Oliveira, 2011](#)), COVSTATIS ([Thioulouse, 2011](#)), COSTATIS ([Thioulouse, 2011](#)), Power-STATIS ([Bénasséni & Bennani-Dosse, 2012](#)), ANISOSTATIS ([Bennani D., Groenen, P., and Abdi, H. en Abdi, Williams, Valentin, & Bennani-Dosse, 2012](#)), STATIS-LDA ([Sabatier, Myrtille, & Christelle, 2013](#)), INTERSTATIS ([Corrales & Rodríguez, 2014](#)), Sir-STATIS ([Villa-Vialaneix, Sautron, & Chavent, 2016](#)), HiDiSTATIS y DiDiSTATIS ([Kriegsman, 2018a](#)), CATATIS ([Llobell, Cariou, Vigneau, Labenne, & Qannari, 2019](#)), STATICO-CoA ([Mérigot et al., 2019](#)), y CLUSTATIS ([Llobell, Cariou, Vigneau, Labenne, & Qannari, 2020](#)).

La metodología ha sido aplicada en estudios de muchas áreas del conocimiento. Los primeros trabajos en el marco de esta metodología son los realizados por [L'Hermier des Plantes y Thiébaud \(1977\)](#) con el estudio de la precipitación en la región mediterránea; además [Lavit \(1988\)](#) describe la evolución de las provincias españolas entre 1960 y 1979 a través de análisis de datos económicos.

[Figueiredo et al., \(Figueiredo, Figueiredo, Monteiro, & Straume, 2012\)](#) estudiaron el proceso de reestructuración de los bancos Portugueses después de la

privatización; Thioulouse (2011) presenta un análisis simultáneo de parejas de tablas ecológicas; Gudmundsson, Tallaksen & Stahl (2011) estudiaron los patrones de caudales bajos, medios y más alto de Europa; Niang, Fogliatto & Saporta (2009) estudiaron el control multivariante de procesos por lotes; Fournier et al. (2009) investigan el transporte acuífero dentro del Karst; Oliveira y Mexia (2007b) proponen una solución inferencial y aplican la metodología a los datos sobre VIH en Portugal; Stanimirova et al. (2004), utilizan la metodología para el análisis de datos ambientales; Caballero-Juliá, Galindo-Villardón y García (2017) lo aplican a datos textuales múltiples; Ramos-Barberán et al., (2018) aplica la metodología en el área de control estadístico de calidad; entre otros muchos.

2.2. ESTADO DEL ARTE DE LOS MÉTODOS STATIS

El método ha sido aplicado en múltiples áreas del conocimiento, Vicente-Galindo (2013) realizó una revisión bibliográfica y a partir de dicha revisión, presentamos la actualización de la misma. La revisión que hemos hecho, se limitó a las publicaciones realizadas entre 1990 hasta agosto de 2020, con la finalidad de revisar la literatura más actual. Para llevar a cabo dicha revisión se seleccionó como instrumento de búsqueda la base de datos Web Of Science y se utilizaron los siguientes criterios de inclusión:

Palabras clave: STATIS AND MULTIVARIATE ANALYSIS, X-STATIS; partial triadic analysis, DISTATIS, ACT (STATIS method), STATICO, COSTATIS, DIDISTATIS, Power-STATIS, ANISOSTATIS, (k+1)-STATIS, double-STATIS, STATIS-4, INTERSTATIS, COVSTATIS, CLUSTATIS, CATATIS.

Años: 1900 a 2020

Tipos de documentos: artículos, conferencias, resúmenes o libros.

Como resultado se obtuvieron 176 registros en la plataforma Web of Science. De los 176 registros, 169 pertenecen a la colección principal de Web of Science, 6 a SieELO Citation Index y 1 a MEDLINE. La Figura 12 muestra la distribución de los registros por área de investigación, observamos que la mayor parte de las publicaciones se dan en las áreas de matemáticas, tecnología de los alimentos, biología marina y ciencias ambientales y ecología.



Figura 12. Distribución de los artículos por área de investigación según la Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020.

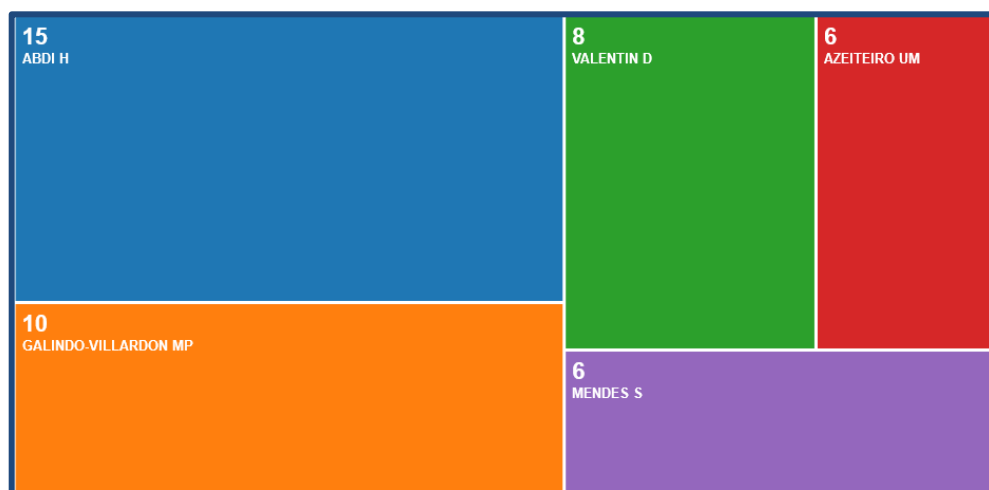


Figura 13. Distribución de las publicaciones por autor, según la Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020

En la Figura 13 se muestran la distribución de las publicaciones de acuerdo a su autor, observamos que Hervé Abdi presenta la mayor cantidad de publicaciones, seguido por Purificación Galindo-Villardón. Es de anotar que Purificación Galindo-Villardón en realidad tiene 16 publicaciones, pero al estar los nombres escritos de forma diferente, el sistema los diferencia (Villardón MPG, Villardon Ma Purificacion Galindo, Galindo-Villardón M y Galindo-Villardón PM).



Figura 14. Resultados de búsqueda en Web of Science, por institución.

Fuente: Web of Science, 14 de septiembre de 2020.

Por otro lado, en cuanto a las instituciones que han publicado sobre la temática, observamos en la Figura 14, que el Centro Nacional de investigación Científica de Francia (CNRS) y el Instituto Nacional de Investigación Agronómica de Francia (INRAE), son los que más publicaciones tienen, les siguen el Instituto de investigación francés para la explotación del mar (IFREMER), el Instituto de investigación para el desarrollo (IRD) de Francia y la Universidad de Salamanca en España. En el caso de la Universidad de Salamanca, la cantidad total de publicaciones es de 18, pero al estar el nombre de la institución en inglés y español, el sistema las considera como diferentes. Los países que más han publicado sobre el tema son: Francia, España y Estados Unidos, como podemos observar en la Figura 15.

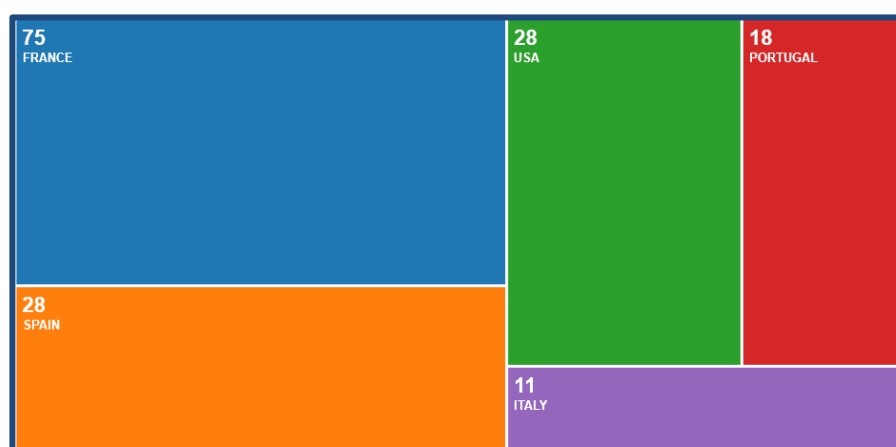


Figura 15. Distribución de las publicaciones por país de origen, según la Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020.

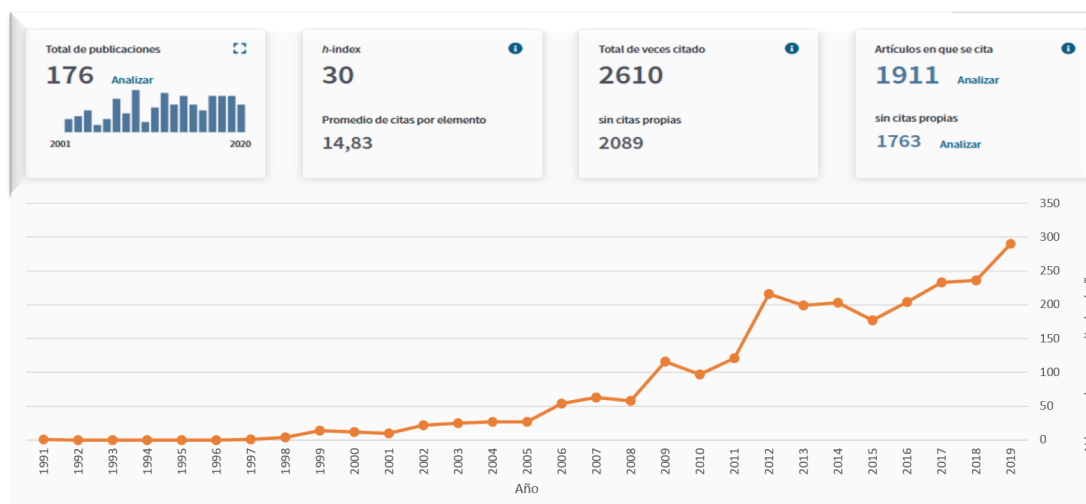


Figura 16. Publicaciones por citas recibidas y año de publicación. Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020

En cuanto a las citas recibidas, podemos ver en la [Figura 16](#) que el año 2019 fue donde se realizó la mayor cantidad de citas para la totalidad de artículos, 290 citas en total para ese año. En el año 2009 se realizaron la mayor cantidad de publicaciones, 14 en total. En general los artículos han sido citados un total de 2572 veces, en 1879 artículos. Presentan un promedio de citas por elemento de 14.7, un promedio de 85.73 citas por año, y un índice h de 30. El artículo con mayor cantidad de citas es el de Lavit *et al.*, (1994), con un total de 172 citas; le siguen Abdi *et al.*, (2007) con 105 citas; y Rudnitskaya *et al.*, (2009) con 97 citas. Dado que el objetivo de esta revisión bibliográfica es actualizar la ya realizada por [Vicente-Galindo \(2013\)](#), presentamos a continuación la información recabada de 2013 a la fecha de la plataforma Web of Science.



Figura 17. Publicaciones por citas recibidas y año de publicación para el período 2013-2020. Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020

La Figura 17 nos muestra los resultados en el período 2013-2020. En dicho período y según la Web of Science, se hicieron 82 publicaciones. En cuanto a las citas recibidas, el año 2019 fue donde se realizó la mayor cantidad de citas, 118 para ese año

En general los artículos en este período, han sido citados un total de 427 veces, en 387 artículos. Presentan un promedio de citas por elemento de 5.21, un promedio de 61 citas por año y un índice h de 12.

El artículo con mayor cantidad de citas es Hernández-Fariñas *et al.*, (2014) con 40 citas, le siguen Beaton *et al.*, (2014) con 38 citas, Klie *et al.*, (2014) con 33 citas, y Mielby *et al.*, Shinkareva *et al.*, (2014) con 25 citas.

Si observamos cual es el comportamiento desde al año 2013, cuando se realizó revisión por parte de Vicente-Galindo, vemos que las áreas donde más se han aplicado los métodos STATIS, durante este período son Matemáticas y en Ciencias Ambientales y Ecología, ver Figura 18.

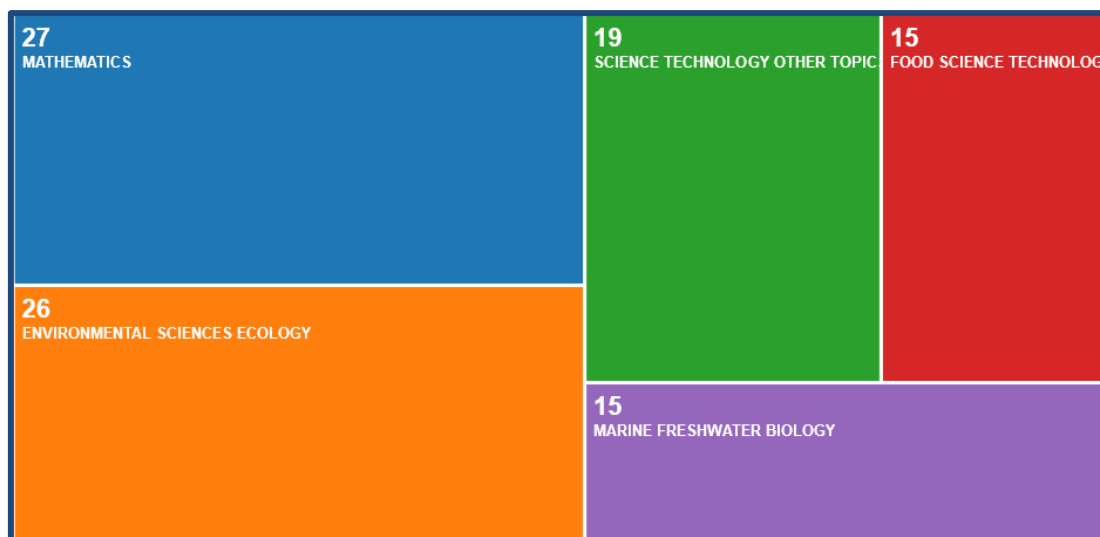


Figura 18. Distribución de los artículos por área de investigación para el período 2013-2020 en Web of Science.

Fuente: Web of Science, 14 de septiembre de 2020

En la Figura 19 se muestran la distribución de las publicaciones de acuerdo a su autor, observamos que la mayor cantidad de publicaciones corresponden a Purificación Galindo-Villardón, con 10 en total para el período 2013-2020, la figura solo muestra 6 por lo que ya explicamos anteriormente.

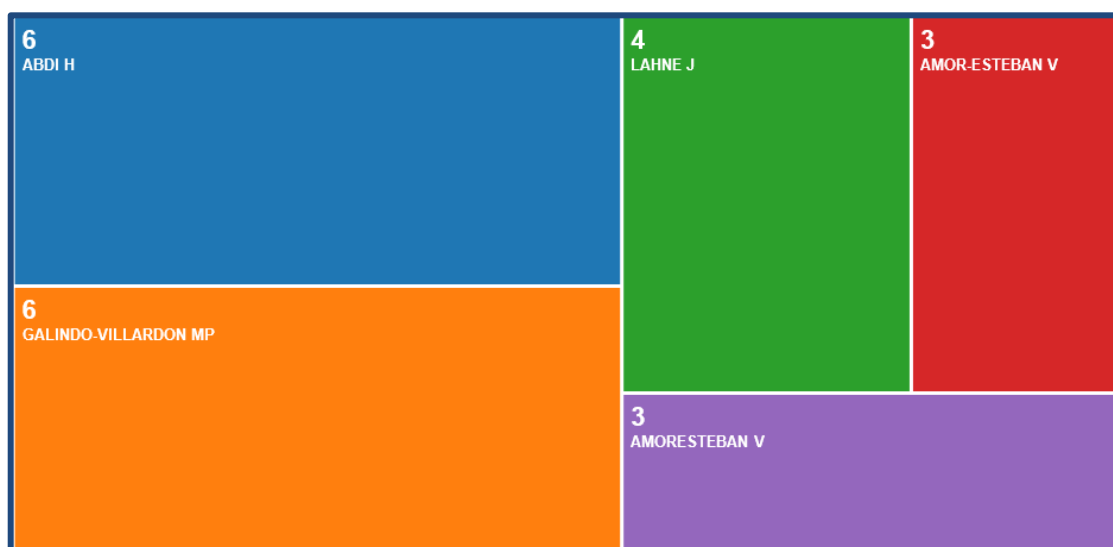


Figura 19. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por autor.

Fuente: Web of Science, 14 de septiembre de 2020

La Universidad de Salamanca la que cuenta con la mayor cantidad de publicaciones, 17 en total durante el período 2013-2020, ver [Figura 20](#).

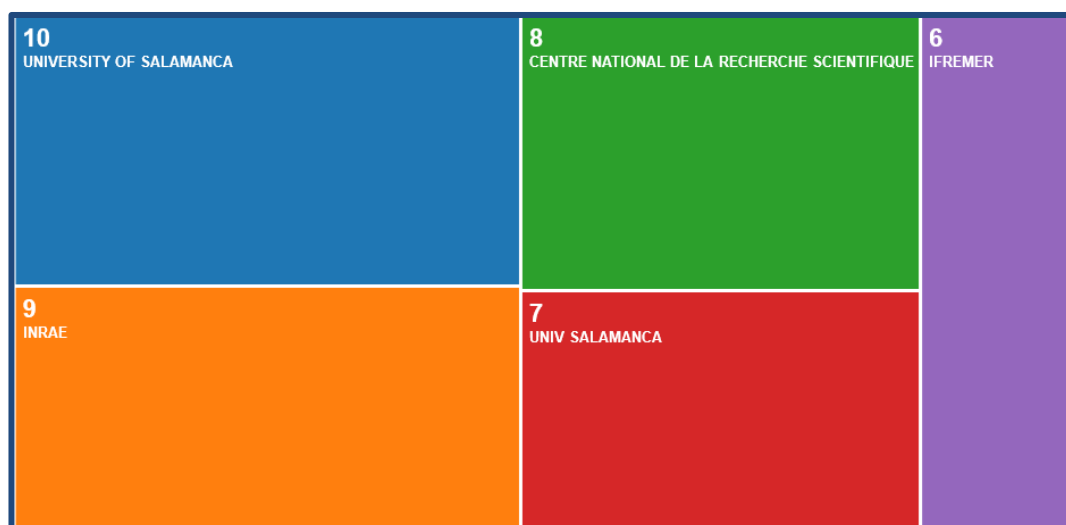


Figura 20. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por Institución.

Fuente: Web of Science, 14 de septiembre de 2020

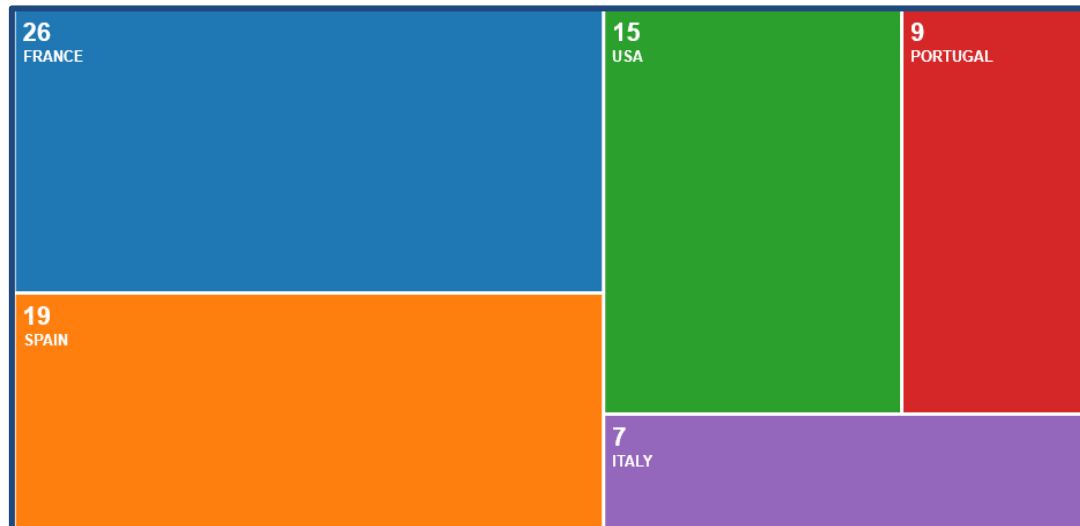


Figura 21. Resultados de la búsqueda bibliográfica para el período 2013-2020 en Web of Science, por país.

Fuente: Web of Science, 14 de septiembre de 2020.

Los países que más han publicado sobre el tema, durante el período 2013-2020 son: Francia, España y Estados Unidos, como podemos observar en la [Figura 21](#).

Algo que queremos señalar, que aunque no es fundamental en nuestro trabajo, es importante comentarlo, y es el hecho de que, la información que genera el sistema, en este caso el informe de citas y el análisis de los resultados de Web of Science, muchas veces presenta errores, que no necesariamente se deben al sistema como tal, sino al hecho que, al momento de registrar la publicación, los autores varían en la forma en que escriben sus nombres o el de las instituciones a las que pertenecen, por lo cual el sistema los registra como autores o instituciones diferentes. Por ejemplo, en nuestra búsqueda uno de los gráficos arroja dos nombres diferentes de la Universidad de Salamanca (University of Salamanca y Univ Salamanca), lo que el sistema registra como instituciones distintas y no calcula las estadísticas de forma correcta, (ver [Figura 20](#)). Lo mismo sucede con los nombres de los autores, vemos los casos de Amor-Esteban (AmorEsteban V) y Gallego-Álvarez I (GallegoAlvarez I) entre otros, que se consideran autores diferentes y por tanto bajan sus estadísticas (ver [Figura 19](#)).

Además de la búsqueda en el repositorio Web of Science, donde se encontraron 176 publicaciones diferentes, se hizo una búsqueda en el repositorio SCOPUS, encontrándose 2 registros diferentes a los encontrados en Web of Science, además se hizo una búsqueda en plataformas de instituciones académicas, dicha búsqueda arrojó 5 registros adicionales.

Otra plataforma de búsqueda utilizada fue NUMDAM (<http://www.numdam.org/>), la biblioteca digital francesa de referencia de matemáticas, esta fue incluida por la importancia que tiene al ser el repositorio en donde se encuentra el artículo base de los estudios sobre STATIS. La búsqueda en este repositorio generó 10 artículos más.

Se obtuvieron en total 193 publicaciones que corresponden al período comprendido entre 1976 a 2020, de los cuales, 82 se publicaron entre 2013 y 2020.

Dentro de las 82 publicaciones hechas el período 2013 a 2020, la mayor parte corresponden a aplicaciones de los métodos STATIS en diferentes áreas de investigación como lo son Ciencias de la Vida y Biomedicina, Ciencias físicas, Ciencias Sociales, Tecnología y Matemáticas (ver [Tabla 1](#)); y el resto son propuestas de modelos basados en métodos STATIS (ver

Tabla 2).

Así tenemos que [Krishnan, Kriegeskorte, y Abdi, \(2013\)](#), introduce un nuevo método estadístico llamado Regresión de Mínimos Cuadrados Parciales basada en la distancia (DISPLSR), que predice una matriz de distancia de otra. Dando una alternativa a los modelos GPA (Análisis generalizado de Procrustes), INDSCAL y DISTATIS que extraen similitudes entre varias matrices de distancia, pero no predicen.

El método, llamado STATIS-LDA ([Sabatier et al., 2013](#)). Propone un nuevo enfoque y su algoritmo asociado para resolver la tarea de clasificación, cuando los grupos de individuos son los mismos en todas las tablas. Utiliza el análisis discriminante lineal habitual (LDA-discriminante lineal de Fisher) y el enfoque STATIS. utiliza la descomposición entre variaciones de cada bloque y permite cuantificar la importancia de cada bloque con respecto a la clasificación.

[Corrales y Rodríguez \(2014\)](#), desarrollan el método INTERSTATIS, que aplica el método STATIS, teniendo como entrada datos de tipo intervalo. Para esto se utilizan la aritmética de intervalos de Moore ([Moore, 1979](#)) y la aritmética afín ([Stolfi & De Figueiredo, 2003](#)), así como el Método de los Centros para Análisis en Componentes Principales con datos de tipo intervalo ([Cazes, Chouakria, Diday, & Schektman, 1997](#)).

De igual manera [da Silva y Ramos \(2014b\)](#), presentan una técnica que consiste en aproximar uniformemente la secuencia de distribución $\{g(\mathbf{V}_n + \boldsymbol{\mu}_n), n \geq 1\}$, donde g es una función suavizamiento de varias variables, $\{\mathbf{V}_n, n \geq 1\}$ es una secuencia de vectores aleatorios distribuidos idénticamente de tipo continuo y $\boldsymbol{\mu}_n$ es un vector no aleatorio.

La matriz $\mathbf{S} = [\text{tr}(\mathbf{W}_i \mathbf{Q} \mathbf{W}_j \mathbf{Q})]$ con $i, j = 1, \dots, h$; donde \mathbf{Q} es una matriz simétrica definida positiva y $\mathbf{W}_i = \mathbf{X}_i^T \mathbf{D}_i \mathbf{X}_i$, con $i = 1, \dots, h$ que está formada por tablas de datos \mathbf{X}_i , y matrices diagonales de pesos \mathbf{D}_i , juegan un papel central en el método STATIS dual. Su propuesta aproxima la función de distribución de las entradas de \mathbf{S} , asumiendo las tablas de datos \mathbf{X}_i dadas por $\mathbf{U}_i + \mathbf{E}_i$, para $i = 1, \dots, h$ con matrices aleatorias independientes \mathbf{E}_i que representan errores, para obtener (aproximadamente) la distribución de $\mathbf{S}\mathbf{v}$, donde \mathbf{v} es el vector propio ortonormal de \mathbf{S} asociado al valor propio más grande. Para lograr este objetivo, aproximan uniformemente la distribución de cada entrada de \mathbf{S} .

Nuevamente [da Silva y Ramos \(2014a\)](#), desarrollan límites uniformes para la secuencia de funciones de distribución de $g(\mathbf{V}_n + \mu_n)$, donde g es una función de suavizamiento, $\{\mathbf{V}_n, n \geq 1\}$ es una secuencia de variables aleatorias idénticamente distribuidas con distribución común que tiene una derivada acotada $\{\mu_n\}$ y son constantes de tal manera que $\mu_n \rightarrow \infty$. Estos límites permiten identificar una secuencia adecuada de variables aleatorias que son asintóticamente del mismo tipo de $g(\mathbf{V}_n + \mu_n)$, mostrando que la tasa de convergencia para estas aproximaciones uniformes depende de la relación de la razón entre la segunda y la primera derivada de g . Analizan la generalización correspondiente al caso multivariante y presentan una aplicación al método STATIS-ACT.

[Marcondes Filho, Fogliatto, y de Oliveira \(2014\)](#), presentan un enfoque que utiliza gráficos de control basados en el método STATIS. Dado que STATIS solo evalúa las similitudes en un contexto lineal, es decir, investiga estructuras de correlación lineal en los datos, los autores proponen gráficos de control basados en el método STATIS, junto con un núcleo para monitorear procesos en presencia de no linealidades. A través de los núcleos definen funciones no lineales de datos que dan una mejor representación de la estructura que se caracteriza por el método STATIS, ilustran su propuesta utilizando datos simulados.

La propuesta de [Boumazza, Yousfi, y Demotes-Mainard \(2015\)](#), utiliza el análisis de componentes principales funcionales (FPCA), su propuesta se basa en que al ser el FPCA una técnica de reducción de datos de un número T finito de funciones puede usarse para identificar los modos dominantes de variación de datos numéricos de tres vías. Aplican el FPCA sobre funciones de densidad de probabilidad multidimensional, relacionan este método con otros métodos estándar y definen sus versiones centradas o estandarizadas. Basado en la relación entre FPCA de densidades, FPCA de sus funciones características correspondientes, PCA de las expansiones de MacLaurin de estas funciones características y el método STATIS dual aplicado a sus matrices de varianza, proponen un método para interpretar los resultados del FPCA de densidades. Este método se basa en las investigaciones de las relaciones entre los puntajes de la FPCA y los momentos asociados a las densidades.

Amerise y Tarsitano (2016), proponen utilizar un coeficiente distinto al RV de Escoufier (1973), ya que este tiene el inconveniente de depender de la correlación de Pearson, que es muy propensa a los efectos de los valores atípicos. La solución que proponen es la sustitución del coeficiente de Pearson con correlaciones de rango que se ven menos afectadas por errores de medición, falta de linealidad o valores atípicos. Los resultados que obtienen con datos reales y simulados confirman que la aplicación de correlaciones de rango de vectores atenúa los efectos adversos de las anomalías y, en el caso de datos limpios y sin defectos, produce pesos que básicamente se ajustan a los obtenidos utilizando el coeficiente de Escoufier.

Marcondes Filho, Oliveira, y Fogliatto, (Marcondes Filho, de Oliveira, & Fogliatto, 2017) proponen gráficas de control estadístico de calidad (CC) basados en el método STATIS. Los datos se organizan de tal manera que se prioriza el monitoreo a lo largo del tiempo. Señalan que la metodología permite fácilmente un monitoreo en línea no paramétrico de procesos por lotes complejos a tiempo, en situaciones donde hay una gran cantidad de variables presentes; ilustran el nuevo enfoque utilizando datos simulados.

En 2017, Mendes *et al.* (2017), publican su propuesta “CO-tucker: un nuevo método para el análisis simultáneo de una secuencia de tablas emparejadas”- (*CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables*), en la misma presenta un nuevo método que es un enfoque combinado de las técnicas STATICO y Tucker3; aborda el problema de describir no solo la parte estable de la dinámica de las relaciones estructura-función entre las comunidades y su entorno (en diferentes ubicaciones y / o en diferentes veces), pero también las interacciones y cambios asociados con la dinámica de los ecosistemas. Presenta una comparación con el método STATICO en el mismo conjunto de datos (real), donde se exploran y discuten las ventajas y desventajas, e indican que el método proporciona ganancias en la capacidad de interpretación.

Llobell *et al.* (2020), publican el artículo titulado “Análisis y agrupamiento de conjuntos de datos multibloque mediante los métodos STATIS y CLUSTATIS. Aplicación a la sensometría”- (*Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics*).

En este artículo, discuten el uso y beneficios del método STATIS, y comparan sus resultados con métodos alternativos para el análisis de datos multibloque que surgen en situaciones como el mapeo proyectivo y los experimentos de clasificación libre. Introducen un método para agrupar una colección de conjuntos de datos medidos en los mismos individuos, llamado **CLUSTATIS**, el cual se basa en la optimización de un criterio y consiste en un análisis de agrupamiento jerárquico y un algoritmo de partición similar al algoritmo K-means. Señalan que el procedimiento de análisis puede verse como una extensión del análisis de conglomerados de variables alrededor de componentes latentes-(CLV), (Vigneau & Qannari, 2003) al caso de bloques de variables. Junto con la determinación de los grupos, el método STATIS determina una configuración latente. Se discute e ilustra el interés de CLUSTATIS en la sensometría, sobre la base de dos estudios de caso relacionados con el mapeo proyectivo, también llamado “Napping” y las tareas de clasificación libre, respectivamente. Podemos ver aplicaciones de este método en (Llobell, Cariou, et al., 2019; Llobell, Vigneau, & Qannari, 2019).

Ramos-Barberán et al.(2018), proponen en su artículo “Control y monitoreo de procesos por lotes: un enfoque de STATIS dual y coordenadas paralelas (DS-PC)”-(Batch process control and monitoring: a Dual STATIS and Parallel Coordinates (DS-PC) approach), una estrategia de control de calidad no paramétrica para el monitoreo fuera de línea de lotes y variables, además de la agrupación visual de observaciones dentro de lotes, ya que las técnicas para datos multivariados centradas en el análisis de variables no se han desarrollado ampliamente,. En su enfoque, los gráficos de control (CC) se crean basados en STATIS dual (L’Hermier des Plantes, 1976), utilizando bagplot robustos (Rousseeuw, Ruts, & Tukey, 1999), para mejorar la detección de señal en análisis por lotes y variables, mientras que los diagramas de coordenadas paralelas (Inselberg, 1985), se usan para identificar el comportamiento de las observaciones inusuales por variable, independientemente de los supuestos de distribución. La propuesta presenta la principal ventaja de detectar diferentes tipos de cambios a través de herramientas de visualización significativas, lo que permite una interpretación más fácil de los resultados en entornos industriales

Como hemos visto, de las 15 propuestas, tres de ellas desarrollan nuevas aportaciones a los métodos de la familia STATIS, estos son el artículo de [Sabatier, Myrtille y Christelle \(2013\)](#), con el título *Une nouvelle proposition, l'Analyse Discriminante Multitableaux* : STATIS-LDA método para clasificar en el caso de que los individuos sean los mismos en todas las tablas; el de [Corrales y Rodríguez \(2014\)](#) titulado *Interstatis: el método statis para datos de tipo intervalo* y; el de [Llobell, \(2020\)](#), titulado *Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics*, donde se aporta una nueva técnica de clusters para matrices en tres vías, denomina CLUSTATIS. Además incluimos un nuevo aporte denominado DiDISTATIS, presentado en la tesis doctoral *Discriminant DiSTATIS: A Multi-Way Discriminant Analysis for Distance Matrices, Illustrations with the Sorting Task* por [Kriegsman \(2018a\)](#) en donde propone una metodología basada en DISTATIS y análisis discriminante; y otra denominada HiDiSTATIS para resolver el problema de múltiples tablas, en donde en lugar de mapear a los participantes, mapea los grupos y cómo los participantes varían sobre los grupos. HiDiSTATIS luego cuantifica la variabilidad entre grupos y dentro del grupo para evaluar si los grupos percibieron de manera diferente los estímulos. Dentro de la búsqueda realizada, se encontró una propuesta, que aunque no es una publicación, si fue presentada en un congreso y hemos querido incluirla dentro de los nuevos aportes, dicha técnica se denomina SIR-STATIS ([Villa-Vialaneix et al., 2016](#)).

Estos métodos los presentaremos con más detalle en el capítulo III, es de señalar que en este período [Mendes et al., \(2017\)](#), desarrolla la técnica Co-Tucker, sin embargo al ser más orientadas a los métodos Tucker la excluimos de nuestro trabajo.

Tabla 1. Revisión bibliográfica de publicaciones con aplicación y desarrollos teóricos sobre los métodos de la Familia STATIS (2012-2020) Aplicaciones a diferentes áreas de investigación

No.	Repositorio	Autores	Título	Area	Año	Método
1	WOS	(Baumberger et al., 2012)	Plant community changes as ecological indicator of seabird colonies' impacts on Mediterranean Islands	Biodiversity & Conservation	2012	STATICO
2	Scopus	(Figueiredo et al., 2012)	Restructuring in privatised firms: A Statis approach	Business & Economics	2012	STATIS
3	WOS	(Eros, Sály, Takács, Specziár, & Biró, 2012)	Temporal variability in the spatial and environmental determinants of functional metacommunity organization - stream fish in a human-modified landscape	Environmental Sciences & Ecology	2012	X-STATIS (PTA)
4	WOS	(Blancher, Clavier, Egoroff, Duineveld, & Parcon, 2012)	A method to investigate the stability of a sorting map	Food Science & Technology	2012	DiSTATIS
5	WOS	(Napoléon et al., 2012)	Spatiotemporal dynamics of physicochemical and photosynthetic parameters in the central English Channel	Marine & Freshwater Biology	2012	X-STATIS (PTA)
6	WOS	(Abdi et al., 2012)	STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling	Mathematics	2012	STATIS; STATIS-dual; DiSTATIS; X-STATIS(PTA); CovSTATIS; CanoSTATIS; (k+1)STATIS; STATIS4; ANISOSTATIS; Power STATIS; Double STATIS
7	WOS	(Areia & Oliveira, 2012)	Transversal Analysis in Matched Series of Studies	Mathematics	2012	STATIS

No.	Repositorio	Autores	Título	Area	Año	Método
8	WOS	(Mendes, Fernández-Gómez, Pereira, Azeiteiro, & Galindo-Villardón, 2012)	An empirical comparison of Canonical Correspondence Analysis and STATICO in the identification of spatio-temporal ecological relationships	Mathematics	2012	STATICO
9	WOS	(Churchill et al., 2012)W.	Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods	Neurosciences & Neurology	2012	DiSTATIS
10	WOS	(Shinkareva, Malave, Just, & Mitchell, 2012)	Exploring commonalities across participants in the neural representation of objects	Neurosciences & Neurology	2012	STATIS
11	WOS	(Mazzocchi, Dubroca, García-Comas, Capua, & Ribera d'Alcalà, 2012)	Stability and resilience in coastal copepod assemblages: The case of the Mediterranean long-term ecological research at Station MC (LTER-MC)	Oceanography	2012	STATICO
12	WOS	(Klie, Caldana, & Nikoloski, 2012)	Compromise of multiple time-resolved transcriptomics experiments identifies tightly regulated functions	Plant Sciences	2012	STATIS; STATIS-dual
13	WOS	(Gonçalves et al., 2012)	Responses of Copepoda Life-History Stages to Climatic Variability in a Southern-European Temperate Estuary	Zoology	2012	STATICO
14	WOS	(Godefroid, Delaville, Marie-Luce, & Quénéhervé, 2013)	Spatial stability of a plant-feeding nematode community in relation to macro-scale soil properties	Agriculture	2013	STATIS-CoA
15	WOS	(Pizarro, Esteban-Díez, Rodríguez-Tecedor, & González-Sáiz, 2013)	A sensory approach for the monitoring of accelerated red wine aging processes using multi-block methods	Food Science & Technology	2013	STATIS
16	WOS	(Gourdol, Hissler, Hoffmann, & Pfister, 2013)	On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry	Geochemistry & Geophysics	2013	X-STATIS (PTA)
17	WOS	(Areia & Oliveira, 2013)	Longitudinal Analysis for Matched Series of Studies	Mathematics	2013	STATIS

No.	Repositorio	Autores	Título	Area	Año	Método
18	WOS	(Kissita, Niere, Makany, & Mizere, 2013)	Canonical correlation analysis between two vertical multi-block tables and applications to qualitative variables	Mathematics	2013	STATICO
19	WOS	(Krishnan et al., 2013)	Distance-Based Partial Least Squares Analysis	Mathematics	2013	DiSTATIS
20	WOS	(Niang, Fogliatto, & Saporta, 2013)	Non parametric on-line control of batch processes based on STATIS and clustering	Mathematics	2013	STATIS
21	WOS	(Ferraro et al., 2013)	Multi-marker network in ST-elevation myocardial infarction patients undergoing primary percutaneous coronary intervention: When and what to measure	Medical Laboratory Technology	2013	STATIS
22	WOS	(Derek Beaton et al., 2014)	An ExPosition of multivariate analysis with the singular value decomposition in R	Computer Science	2014	STATIS; STATIS-dual; DiSTATIS; X-STATIS(PTA); CovSTATIS; CanoSTATIS; (k+1)STATIS; STATIS4; ANISOSTATIS; Power STATIS; Double STATIS
23	WOS	(Pagès, 2014)	Multiple Factor Analysis: General Presentation and Comparison with STATIS	Computer Science	2014	STATIS
24	WOS	(Marcondes Filho et al., 2014)	Monitoring Nonlinear Batch Process Using Statis-based Method	Engineering; Mechanics	2014	STATIS
25	WOS	(Hernández-Fariñas et al., 2014)	Temporal changes in the phytoplankton community along the French coast of the eastern English Channel and the southern Bight of the North Sea	Fisheries	2014	X-STATIS (PTA)

No.	Repositorio	Autores	Título	Area	Año	Método
26	WOS	(Mielby, Hopfer, Jensen, Thybo, & Heymann, 2014)	Comparison of descriptive analysis, projective mapping and sorting performed on pictures of fruit and vegetable mixes	Food Science & Technology	2014	DiSTATIS
27	WOS	(da Silva & Ramos, 2014a)	On the rate of convergence of uniform approximations for sequences of distribution functions	Mathematics	2014	STATIS
28	WOS	(da Silva & Ramos, 2014b)	Uniform approximations for distributions of continuous random variables with application in dual STATIS method	Mathematics	2014	STATIS-dual
29	WOS	(Shinkareva et al., 2014)	Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data	Neurosciences & Neurology	2014	STATIS
30	WOS	(Viswanathan, Lee, Lees, Cheong, & Sloot, 2014)	Quantitative comparison between crowd models for evacuation planning and evaluation	Physics	2014	DiSTATIS
31	WOS	(Klie et al., 2014)	Conserved Changes in the Dynamics of Metabolic Processes during Fruit Development and Ripening across Species	Plant Sciences	2014	STATIS
32	WOS	(Rossi et al., 2014)	Dissecting the Space-Time Structure of Tree-Ring Datasets Using the Partial Triadic Analysis	Science & Technology	2014	X-STATIS (PTA)
33	WOS	(Schwarz et al., 2015)	A Read-Aloud Storybook Selection System for Prereaders at the Preschool Language Level: A Pilot Study	Audiology & Speech-Language Pathology	2015	DiSTATIS
34	WOS	(Gallego-Álvarez, Galindo-Villardón, & Rodríguez-Rosa, 2015)	Evolution of sustainability indicator worldwide: A study from the economic perspective based on the X-STATICO method	Biodiversity & Conservation	2015	X-STATIS (PTA); X-STATICO
35	WOS	(Darwiche-Criado, Jiménez, Comín, Sorando, & Sánchez-Pérez, 2015)	Identifying spatial and seasonal patterns of river water quality in a semiarid irrigated agricultural Mediterranean basin	Environmental Sciences & Ecology	2015	X-STATIS (PTA)

No.	Repositorio	Autores	Título	Area	Año	Método
36	WOS	(Jiménez, Darwiche-Criado, Sorando, Comín, & Sánchez-Pérez, 2015)	A Methodological Approach for Spatiotemporally Analyzing Water-Polluting Effluents in Agricultural Landscapes Using Partial Triadic Analysis	Environmental Sciences & Ecology	2015	X-STATIS (PTA)
37	WOS	(Lemesle, Mussio, Rusig, Menet-Nédélec, & Claquin, 2015)	Impact of seaweed beachings on dynamics of delta N-15 isotopic signatures in marine macroalgae	Environmental Sciences & Ecology	2015	X-STATIS (PTA)
38	WOS	(Masmoudi et al., 2015)	Salinity gradient and nutrients as major structuring factors of the phytoplankton communities in salt marshes	Environmental Sciences & Ecology	2015	STATICO
39	WOS	(Ladhar, Tastard, Casse, Denis, & Ayadi, 2015)	Strong and stable environmental structuring of the zooplankton communities in interconnected salt ponds	Marine & Freshwater Biology	2015	CoSTATIS
40	WOS	(Boumaza et al., 2015)	Interpreting the Principal Component Analysis of Multivariate Density Functions	Mathematics	2015	STATIS-dual
41	WOS	(Kidé, Manté, Dubroca, Demarcq, & Mérigot, 2015)	Spatio-Temporal Dynamics of Exploited Groundfish Species Assemblages Faced to Environmental and Fishing Forcings: Insights from the Mauritanian Exclusive Economic Zone	Science & Technology	2015	STATICO
42	WOS	(Bono & Giacomarra, 2016)	The photovoltaic growth in the European Union requires stronger RES support	Business & Economics	2016	STATIS
43	WOS	(Lahne, Collins, & Heymann, 2016)	Replication Improves Sorting-Task Results Analyzed by DISTATIS in a Consumer Study of American Bourbon and Rye Whiskeys	Food Science & Technology	2016	DiSTATIS
44	WOS	(Ren et al., 2016)	Spatial and temporal assessment of the initial pattern of phytoplankton population in a newly built coastal reservoir	Geology	2016	X-STATIS (PTA)
45	WOS	(Guesdon et al., 2016)	Effect of local hydroclimate on phytoplankton groups in the Charente estuary	Marine & Freshwater Biology	2016	X-STATIS;STATICO

No.	Repositorio	Autores	Título	Area	Año	Método
46	WOS	(Amerise & Tarsitano, 2016)	Combining dissimilarity matrices by using rank correlations	Mathematics	2016	DiSTATIS
47	WOS	(Connolly et al., 2016)	How the Human Brain Represents Perceived Dangerousness or "Predacity" of Animals	Neurosciences & Neurology	2016	STATIS
48	WOS	(Rivadeneira, Figueiredo, Figueiredo, Carvajal, & Rivadeneira, 2016)	Analysis of well-being in OECD countries through STATIS methodology	Science & Technology	2016	STATIS
49	WOS	(Marcondes Filho et al., 2017)	Multivariate quality control of batch processes using STATIS	Automation & Control Systems	2017	STATIS
50	WOS	(Nguyen & Holmes, 2017)	Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations	Biochemistry & Molecular Biology	2017	DiSTATIS
51	WOS	(Amor-Esteban, Galindo-Villardón, & Sánchez, 2017)	Cultural values on CSR patterns and evolution: A study from the biplot representation	Biodiversity & Conservation	2017	X-STATIS
52	WOS	(Antonio Duarte Santos, Ribeiro, Castela, & Tavares da Silva, 2017)	The Dynamics between Economic Growth and Living Standards in EU Countries: A STATICO Approach for the Period 2006-2014	Business & Economics	2017	STATICO
53	WOS-SciELO	(Caballero-Juliá et al., 2017)	JK-Meta-Biplot and STATIS-dual as multiple textual tables tools	Computer Science	2017	STATIS-dual
54	WOS	(Lopes, Silva, Castela, & Rebelo, 2017)	Knowledge transfer through journals	Education & Educational Research	2017	STATIS
55	WOS-SciELO	(Olivares, Zingaretti, Demey Zambrano, & Demey, 2017)	The STATIS-ACT method applied to rainfall in the Venezuelan Eastern Region	Education & Educational Research	2017	STATIS
56	WOS	(Slimani et al., 2017)	The use of STATICO and COSTATIS, two exploratory three-ways analysis methods: an	Environmental Sciences & Ecology	2017	STATICO; CoSTATIS

No.	Repositorio	Autores	Título	Area	Año	Método
			application to the ecology of aquatic heteroptera in the Medjerda watershed (Tunisia)			
57	WOS	(Gómez-Corona, Valentin, Escalona-Buendía, & Chollet, 2017)	The role of gender and product consumption in the mental representation of industrial and craft beers: An exploratory study with Mexican consumers	Food Science & Technology	2017	DiSTATIS
58	WOS	(Castillo-Escrivà, Valls, Rochera, Camacho, & Mesquita-Joanes, 2017)	Metacommunity dynamics of Ostracoda in temporary lakes: Overall strong niche effects except at the onset of the flooding period	Marine & Freshwater Biology	2017	X-STATIS (PTA)
59	WOS	(Mendes et al., 2017)	CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables	Mathematics	2017	STATICO
60	WOS	(Rodríguez-Rosa, Gallego-Álvarez, Vicente-Galindo, & Galindo-Villardón, 2017)	Are Social, Economic and Environmental Well-Being Equally Important in all Countries Around the World? A Study by Income Levels	Social Sciences	2017	X-STATIS (PTA); CoSTATIS
61	WOS	(Rundle, Coch, Connolly, & Granger, 2018)	Dissociating frequency and animacy effects in visual word processing: An fMRI study	Audiology & Speech-Language Pathology	2018	STATIS
62	WOS	(Amor-Esteban, Galindo-Villardón, & David, 2018)	Study of the Importance of National Identity in the Development of Corporate Social Responsibility Practices: A Multivariate Vision	Business & Economics	2018	X-STATICO
63	WOS	(Ramos-Barberán et al., 2018)	Batch process control and monitoring: a Dual STATIS and Parallel Coordinates (DS-PC) approach	Engineering	2018	STATIS-dual
64	WOS	(Amor-Esteban, Galindo-Villardón, & García-Sánchez, 2018)	Industry mimetic isomorphism and sustainable development based on the X-STATIS and HJ-biplot methods	Environmental Sciences & Ecology	2018	X-STATIS (PTA)
65	WOS	(Brand et al., 2018)	Sorting in Combination with Quality Scoring: A Tool for Industry Professionals to Identify Drivers of Wine Quality Rapidly	Food Science & Technology	2018	DiSTATIS

No.	Repositorio	Autores	Título	Area	Año	Método
66	WOS	(Lahne, Abdi, & Heymann, 2018)	Rapid sensory profiles with DISTATIS and Barycentric Text Projection: An example with amari, bitter herbal liqueurs	Food Science & Technology	2018	DiSTATIS
67	WOS	(Feki-Sahnoun et al., 2018)	Multi-table approach to assess the biogeography of phytoplankton: ecological and management implications	Marine & Freshwater Biology	2018	STATIS
68	WOS	(Rodrigues et al., 2018)	Use of phytoplankton functional groups as a model of spatial and temporal patterns in reservoirs: a case study in a reservoir of central Brazil	Marine & Freshwater Biology	2018	STATICO
69	WOS	(Corduas & Ragozini, 2018)	Comparing Multistep Ahead Forecasting Functions for Time Series Clustering	Mathematics	2018	DiSTATIS
70	WOS	(Thabet, Leignel, Ayadi, & Tastard, 2018)	Interannual and seasonal effects of environmental factors on the zooplankton distribution in the solar saltern of Sfax (south-western Mediterranean sea)	Oceanography	2018	STATICO
71	WOS	(Gallego-Álvarez, Lozano, & Rodríguez-Rosa, 2018)	An analysis of the environmental information in international companies according to the new GRI standards	Science & Technology	2018	X-STATIS (PTA)
72	WOS	(Amor-Esteban, García-Sánchez, & Galindo-Villardón, 2018)	Analysing the Effect of Legal System on Corporate Social Responsibility (CSR) at the Country Level, from a Multivariate Perspective	Social Sciences	2018	X-STATIS (PTA)
73	WOS	(Giordano, Ragozini, & Vitale, 2019)	Analyzing multiplex networks using factorial methods	Anthropology	2019	DiSTATIS
74	WOS	(Nguyen & Holmes, 2019)	Ten quick tips for effective dimensionality reduction	Biochemistry & Molecular Biology	2019	STATIS; DiSTATIS
75	U (Salamanca University)	(Amor-Esteban, Galindo-Villardón, & David, 2019)	A Multivariate vision of the industry Specificity as an explanatory factor of companies' social performance	Business & Economics	2019	X-STATIS (PTA)

No.	Repositorio	Autores	Título	Area	Año	Método
76	WOS	(Volkova, 2019)	Russian and European Population's Quality of Life Analysis with the Instruments of Common Principal Components (CPC)	Business & Economics	2019	STATIS
77	WOS	(Fu, Jiang, Zhang, & Zhang, 2019)	Comparison of Perceptual Differences Between Users and Designers in Mobile Shopping App Interface Design: Implications for Evaluation Practice	Computer Science	2019	DiSTATIS
78	WOS	(António Duarte Santos, da Silva, & Castela, 2019)	A COSTATIS approach to business sustainability in turbulent environments from 2008 to 2014	Engineering	2019	CoSTATIS
79	WOS	(Mabrouki, Taybi, El Alami, & Berrahou, 2019)	Biotypology of stream macroinvertebrates from North African and semi arid catchment: Oued Za (Morocco)	Fisheries	2019	X-STATIS (PTA)
80	WOS	(Llobell, Cariou, et al., 2019)	A new approach for the analysis of data and the clustering of subjects in a CATA experiment	Food Science & Technology	2019	STATIS
81	WOS	(Liu, Wang, Li, & Wu, 2019)	Quantifying the Spatio-Temporal Process of Township Urbanization: A Large-Scale Data-Driven Approach	Physical Geography	2019	X-STATIS (PTA)
82	WOS	(Rodríguez-Martínez, García-Sánchez, Vicente-Galindo, & Galindo-Villardón, 2019)	Exploring Relationships between Environmental Performance, E-Government and Corruption: A Multivariate Perspective	Science & Technology	2019	STATIS
83	WOS	(Sáinz et al., 2019)	Influence of cultivar, irrigation, ripening stage, and annual variability on the oxidant/antioxidant systems of olives as determined by MDS-PTA	Science & Technology	2019	MDS-X-STATIS (MDS-PTA)
84	WOS-MedLine	(Zingaretti, Renand, Morgavi, & Ramayo-Caldas, 2020)	Link-HD: a versatile framework to explore and integrate heterogeneous microbial communities.	Bioinformática	2020	Link-HD (STATIS Generalización)
85	WOS	(Ceron, Santana, Lucas, Zocche, & Provete, 2020)	Climatic variables influence the temporal dynamics of an anuran metacommunity in a nonstationary way	Environmental Sciences & Ecology	2020	STATICO

No.	Repositorio	Autores	Título	Area	Año	Método
86	WOS	(Honoré-Chedozeau, Chollet, Lelièvre-Desmas, Ballester, & Valentin, 2020)	From perceptual to conceptual categorization of wines: What is the effect of expertise?	Food Science & Technology	2020	DiSTATIS
87	WOS	(Lahne, 2020)	Sorting backbone analysis: A network-based method of extracting key actionable information from free-sorting task results	Food Science & Technology	2020	DiSTATIS
88	WOS	(Raman, Kriegsman, Abdi, Tillmann, & Dowling, 2020)	Bach, Mozart, and Beethoven: Sorting piano excerpts based on perceived similarity using DiSTATIS	Psychology	2020	DiSTATIS
89	WOS	(Mérillet, Kopp, Robert, Mouchet, & Pavoine, 2020)	Environment outweighs the effects of fishing in regulating demersal community structure in an exploited marine ecosystem	Environmental Sciences & Ecology	2020	STATICO
90	WOS	(Schwarz et al., 2020)	Storybook selection criteria used by teachers of d/Deaf and hard-of-hearing prereaders communicating in English	Education & Educational Research	2020	DISTATIS

Tabla 2. Revisión bibliográfica de publicaciones con aplicación y desarrollos teóricos sobre los métodos de la Familia STATIS (2012-2020). Propuestas de nuevos modelos

No.	Repositorio	Autores	Título	Area	Año	Método
1	WOS	(Bénasséni & Bennani-Dosse, 2012)	Analyzing multiset data by the Power STATIS-ACT method	Mathematics	2012	Power-STATIS
2	WOS	(Sabatier et al., 2013)	A new proposal, the Multitableaux Discriminating analysis: STATIS-LDA	Mathematics	2013	STATIS-LDA
3	WOS-SciELO	(Corrales & Rodríguez, 2014)	Interstatis: the stasis method for interval valued data	Mathematics	2014	InterSTATIS
4	U (Toulouse University)	(Villa-Vialaneix et al., 2016)	Multiway-SIR for Longitudinal Multi-table Data Integration Sommaire	Mathematics	2016	SIR-STATIS
5	U (University of Texas at Dallas)	(Kriegsman, 2018a)	Discriminant DISTATIS. A Multi.Way Discriminant Analysis for Distance Matrices.	Mathematics	2018	DiDiSTATIS HiDiSTATIS
6	WOS	(Llobell, Cariou, et al., 2019)	A new approach for the analysis of data and the clustering of subjects in a CATA experiment	Food Science & Technology	2019	CATATIS
7	WOS	(Llobell, Vigneau, & Qannari, 2019)	Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics	Food Science & Technology	2019	STATIS; CluSTATIS
8	WOS	(Mérigot et al., 2019)	Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern mediterranean sea	Marine & Freshwater Biology	2019	STATICO-CoA
9	WOS	(Llobell et al., 2020)	Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics	Food Science & Technology	2020	STATIS; CluSTATIS

2.3. PRIMEROS DESARROLLOS DE LA FAMILIA DE MÉTODOS STATIS

Los métodos de la familia STATIS fueron introducidos por Escoufier y L'Hermier des Plantes (Escoufier, 1976; L'Hermier des Plantes, 1976) posteriormente Lavit (1988), los desarrolla ampliamente. Es una generalización de las componentes principales-APC-, (Abdi et al., 2012). Estos métodos permiten la exploración simultánea de varias tablas de datos. Se aplica a datos cuantitativos, que se hayan recopilados de acuerdo a una de las siguientes situaciones:

Se tienen H tablas de datos con información recogida en diferentes ocasiones y en los mismos individuos; las variables pueden o no ser diferentes en cada tabla; se analizan en este caso H estudios que se representan como $(\mathbf{X}_h, \mathbf{M}_h, \mathbf{D})$, con $h=1, \dots, H$. El objetivo es estudiar las posiciones relativas de los individuos, y esto se logra a través del método STATIS.

Se tienen H grupos de individuos, posiblemente diferentes, a los que se les miden las mismas variables; luego se analizan los H estudios $(\mathbf{X}_h, \mathbf{M}, \mathbf{D}_h)$, con $h=1, \dots, H$. El objetivo es estudiar las relaciones entre las variables, lo cual se logra con el método STATIS dual.

En el caso de que se quiera estudiar tablas que cruzan los mismos individuos con las mismas variables, es recomendable aplicar los dos métodos.

Aunque los dos métodos son parecidos, parten de diferentes matrices. En el STATIS, inicialmente se trabaja con la matriz de similitud entre individuos \mathbf{XX}^T ; mientras que en el STATIS DUAL se utiliza la matriz que recoge la estructura de covariación, $\mathbf{X}^T\mathbf{X}$. En la [Figura 22](#) se muestra un esquema con esta configuración.

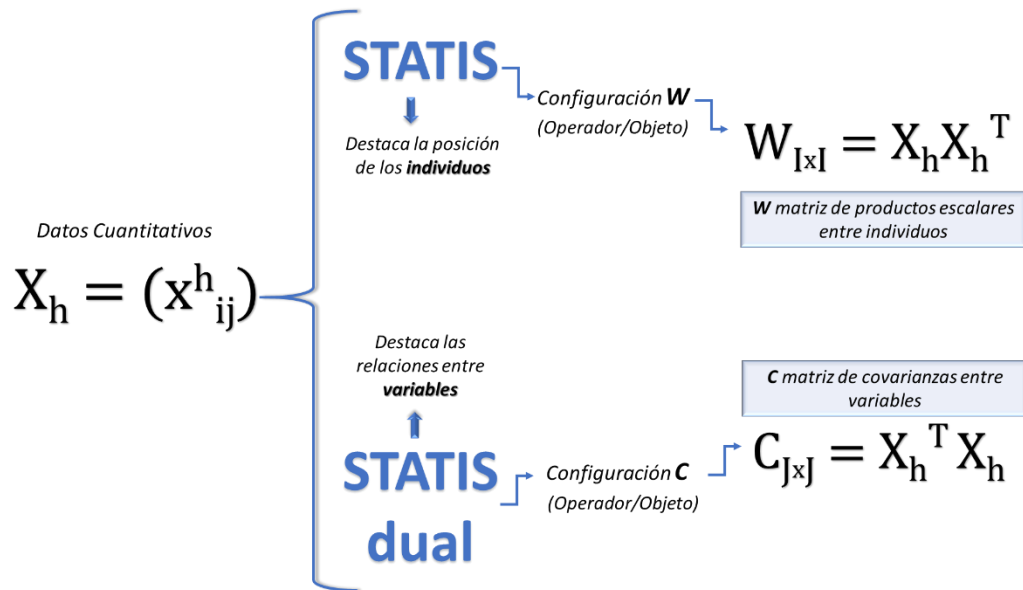


Figura 22. Esquema de las configuraciones de partida en el método STATIS y STATIS dual. Adaptado de Vicente-Galindo (2013).

Tanto en el STATIS como en el STATIS dual, se llevan a cabo en cuatro etapas:
Etapa 1: Estudio de la inter-estructura, cuyo objetivo es proceder a una comparación global de la estructura de las R matrices de datos.

Etapa 2: Construcción de la matriz compromiso, es la búsqueda de una estructura común a las matrices en estudio.

Etapa 3: Estudio de la intra-estructura, que consiste en representar la estructura de cada matriz de datos en un espacio de baja dimensión.

Etapa 4: La representación de las trayectorias de los individuos. Se trazan las trayectorias de los individuos en la imagen euclídea compromiso de los individuos.

La inter-estructura destaca las diferencias y similitudes entre las tablas. La observación de las trayectorias de los individuos ayuda a explicar estas diferencias a nivel individual.

Los objetivos de los métodos STATIS se pueden resumir de la siguiente manera:
Comparar y analizar la relación entre los diferentes conjuntos de datos;
Combinarlos en una estructura común llamada compromiso que luego se analiza mediante ACP para revelar la estructura común entre las observaciones y;
Proyectar cada uno de los conjuntos de datos originales en el compromiso para analizar las comunalidades y las discrepancias

2.3.1. STATIS O STATIS-ACT

En el método STATIS, la cantidad y/o las variables utilizadas para describir las observaciones pueden variar de una tabla de datos a otra, pero los individuos deben ser los mismos en todas las tablas de datos. Por ejemplo, también se pueden tener los mismos grupos de variables medidas en mismo conjunto de individuos en diferentes condiciones (temporales, espaciales, experimentales, etc.).

Este método analiza la estructura de los conjuntos de datos individuales, y deriva de esta estructura un conjunto óptimo de pesos para calcular una representación común de las observaciones llamada **compromiso**. Los pesos utilizados para calcular el compromiso se eligen para que sea lo más representativo posible de todas las tablas de datos.

Un ACP del compromiso proporciona la posición de las observaciones en el espacio compromiso. La posición de las observaciones para cada tabla de datos se representa en el espacio compromiso como puntos complementarios. Los conjuntos de datos se pueden representar como puntos en un espacio multidimensional.

Dependiendo de la forma en que se hayan hecho las mediciones de las tablas de datos, podemos tener diferentes objetivos, por ejemplo:

Evaluar si la posición de las observaciones es estable en el tiempo: es necesario que los conjuntos de datos sean medidas tomadas en las mismas observaciones (individuos u objetos) en diferentes momentos u ocasiones. Es decir, tendremos un primer conjunto de datos que corresponde a los datos recogidos en el tiempo u ocasión $h = 1$, el segundo conjunto corresponde a los datos recogidos en el tiempo u ocasión $h = 2$, el tercer conjunto corresponderá a los datos recogidos en el tiempo u ocasión $h = 3$, y así sucesivamente, (ver [Figura 23](#)).

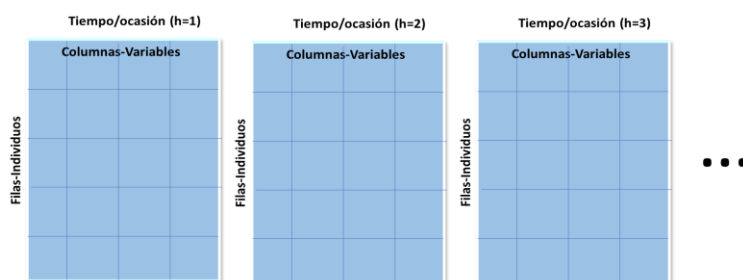


Figura 23. Estructura de los datos para el método STATIS.

Evaluar si hay un acuerdo entre los sujetos o grupos de sujetos: los conjuntos de datos deben ser medidas tomadas en las mismas observaciones (individuos u objetos) por diferentes sujetos o evaluadores o grupo de evaluadores. Es decir, el primer conjunto de datos corresponde al primer evaluador, el segundo al segundo evaluador, el tercer conjunto al tercer evaluador y así sucesivamente, (ver Figura 24).

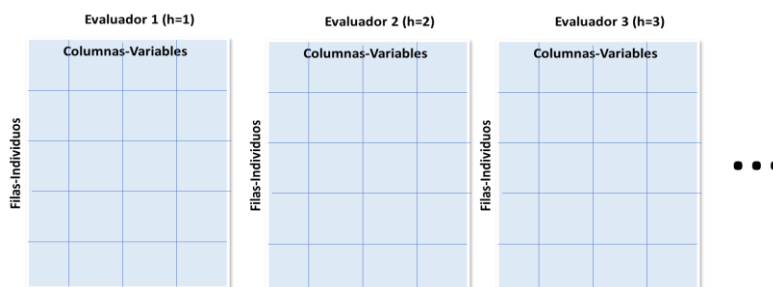


Figura 24. Estructura de datos tipo evaluadores, para la aplicación del método STATIS.

2.3.1.1. Notación

En los métodos STATIS, los datos corresponden a H conjuntos de datos. A cada conjunto de datos se le denomina “estudio” o “escenario”. Un estudio o escenario corresponde a una matriz $I \times J_h$, esta matriz es rectangular y se denota como,

$$Y_h$$

donde

I = número de observaciones

J_h = el número de variables observadas sobre los individuos o muestras para en el h -ésimo estudio u ocasión

Cada matriz generalmente es preprocesada, este preproceso puede consistir en un centrado, una normalización, entre otros. La matriz procesada se denota como,

$$\mathbf{X}_h$$

Las H matrices de datos \mathbf{X}_h de dimensión $\mathbf{I} \times \mathbf{J}_h$, son concatenadas dentro de una matriz completa \mathbf{X} de dimensión $\mathbf{I} \times \mathbf{J}$: $\mathbf{X} = [\mathbf{X}_1 | \dots | \mathbf{X}_h | \dots | \mathbf{X}_H]$ y es la que se utiliza para el análisis, (ver [Figura 25](#))

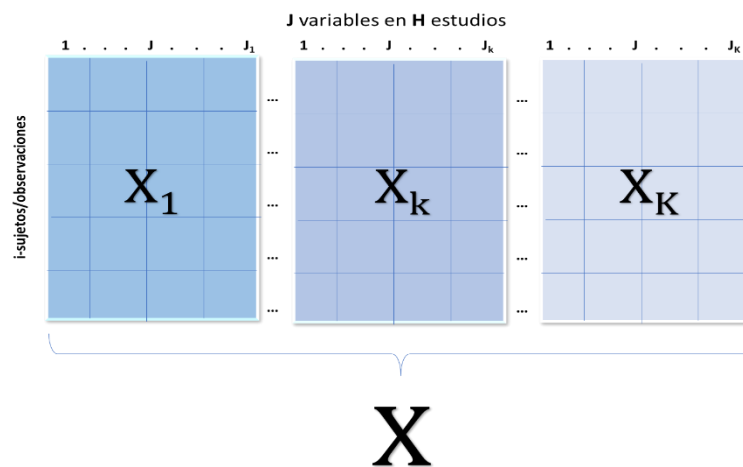


Figura 25. Matriz \mathbf{X} correspondiente a la concatenación de las H matrices \mathbf{X}_h .

Luego las matrices \mathbf{X}_h , son transformadas en matrices de productos cruzados $\mathbf{I} \times \mathbf{I}$, que se denota como,

$$\mathbf{W}_h = \mathbf{X}_h \mathbf{X}_h^T \quad (1)$$

La matriz \mathbf{W}_h de un estudio, expresa como se ven los patrones de relación entre las observaciones de ese estudio.

Al presentar la matriz \mathbf{X} una estructura de bloques, el producto cruzado de \mathbf{X} se puede expresar también de la siguiente manera,

$$\begin{aligned} \mathbf{X}\mathbf{X}^T &= [\mathbf{X}_1 | \dots | \mathbf{X}_k | \dots | \mathbf{X}_H] \mathbf{X} [\mathbf{X}_1 | \dots | \mathbf{X}_k | \dots | \mathbf{X}_H]^T \\ &= \sum_h \mathbf{X}_h \mathbf{X}_h^T \\ &= \sum_h \mathbf{W}_h \end{aligned} \quad (2)$$

Como vimos anteriormente es posible caracterizar un estudio (\mathbf{X}_h , \mathbf{M}_h , \mathbf{D}_h), por un objeto o configuración \mathbf{W} ,

donde:

\mathbf{X}_h : representa la matriz de datos con I individuos y J variables,

\mathbf{M}_h : es la métrica en el hiperespacio de los individuos y

\mathbf{D}_h : es la métrica en el hiperespacio de las variables.

Generalmente las métricas corresponden con la matriz identidad. (L'Hermier des Plantes, 1976)

Este objeto se denota por \mathbf{W}_h : es la matriz de productos cruzados entre individuos. Así tenemos que,

$$\mathbf{W}_h = \mathbf{X}_h \mathbf{M}_h \mathbf{X}_h^T \quad (3)$$

2.3.1.2. Análisis de la Inter-estructura

Como indicamos anteriormente el primer paso del método STATIS consiste en analizar la similitud de las estructuras del conjunto de las \mathbf{H} tablas, lo que se denomina *análisis de la inter-estructura* (Lavit et al., 1994).

La similaridad entre dos matrices de datos, por ejemplo \mathbf{X}_h y $\mathbf{X}_{h'}$, se denota $s_{h,h'}$. Esta similitud se evalúa transformando primero las matrices en matrices de productos cruzados de $\mathbf{I} \times \mathbf{I}$ denotadas como \mathbf{W}_h y $\mathbf{W}_{h'}$, y se obtienen de la siguiente manera,

$$\mathbf{W}_h = \mathbf{X}_h \mathbf{X}_h^T \quad (4)$$

y

$$\mathbf{W}_{h'} = \mathbf{X}_{h'} \mathbf{X}_{h'}^T \quad (5)$$

La medida de similaridad $s_{h,h'}$ la cual es muy similar a un coeficiente de correlación entre matrices cuadradas, se denomina producto interno de Hilbert-Schmidt. Este producto interno entre las matrices \mathbf{W}_h y $\mathbf{W}_{h'}$ se denota como,

$$\langle \mathbf{W}_h | \mathbf{W}_{h'} \rangle_{HS}$$

Los coeficiente $s_{h,h'}$ se obtienen a partir de la siguiente ecuación,

$$\begin{aligned} s_{h,h'} &= \langle \mathbf{W}_h | \mathbf{W}_{h'} \rangle_{HS} \\ &= \text{tr} [\mathbf{X}_h \mathbf{X}_h^T \times \mathbf{X}_{h'} \mathbf{X}_{h'}^T] \\ &= \text{tr} [\mathbf{W}_h \times \mathbf{W}_{h'}] \end{aligned}$$

$$\begin{aligned}
&= \text{vec}(\mathbf{W}_h)^T \times \text{vec}(\mathbf{W}_{h'}) \\
&= \sum_i^I \sum_j^I \mathbf{w}_{i,j,h} \mathbf{w}_{i,j,h'}
\end{aligned} \tag{6}$$

El producto interno puede ser interpretado geoméricamente como un producto escalar entre dos matrices semi-definidas positivas \mathbf{W}_h y $\mathbf{W}_{h'}$. Por lo cual este es proporcional al coseno entre las matrices, dado que el conjunto de matrices semi-definidas positivas es un espacio vectorial.

El producto interno también se utiliza para definir la norma de una matriz como la raíz cuadrada del producto interno de esta matriz consigo misma. Así, la norma de la matriz de productos cruzados \mathbf{W}_h se denota $\|\mathbf{W}_h\|$ y se define como,

$$\|\mathbf{W}_h\|^2 = \langle \mathbf{W}_h, \mathbf{W}_h \rangle \tag{7}$$

Dado que las matrices \mathbf{W}_h son semi-definidas positivas, el producto interno entre dos matrices siempre es mayor o igual a cero.

Es importante considerar que cuando las matrices \mathbf{W}_h y $\mathbf{W}_{h'}$ se normalizan de modo que la suma de los cuadrados de sus elementos es uno, el producto interno entre estas dos matrices es igual al coseno entre ellas. Este coseno también se conoce como el coeficiente RV (Escoufier, 1973; Robert & Escoufier, 1976), el cual fue desarrollado como una medida de similitud entre matrices cuadradas simétricas, específicamente para matrices semi-definidas positivas.

Los productos internos entre todas las \mathbf{H} matrices se recogen en una matriz de productos internos denotada como $\mathbf{S}_{H \times H}$, donde $s_h s_{h'}$ proporciona el valor del producto interno entre las tablas h y h' . La matriz \mathbf{S} también se puede calcular a partir de la matriz $\mathbf{Z}_{H \times I^2}$ definida como,

$$\mathbf{Z} = [\text{vec}(\mathbf{W}_1) | \dots | \text{vec}(\mathbf{W}_h) | \dots | \text{vec}(\mathbf{W}_H)]^T \tag{8}$$

Dada la matriz \mathbf{Z} , se puede calcular \mathbf{S} como,

$$\mathbf{S} = \mathbf{Z}\mathbf{Z}^T \tag{9}$$

Así tenemos que \mathbf{S} es una matriz semi-definida positiva, por lo cual, sus valores propios son todos positivos o nulos y sus vectores propios son reales y ortogonales entre sí. Dando como resultado que \mathbf{S} se pueda descomponer en valores propios,

$$\mathbf{S} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^T \quad (10)$$

en donde

$$\mathbf{V}^T\mathbf{V} = \mathbf{I} \quad (11)$$

Por lo tanto, la descomposición en valores propios de \mathbf{S} proporciona el ACP de la estructura de similitud entre las tablas, que se representarán como puntos en un ACP, lo cual se logra utilizando los puntajes de factores como las coordenadas, calculándolos como,

$$\mathbf{L} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}} \quad (12)$$

Los vectores propios de \mathbf{S} , pueden ser obtenidos igualmente de la descomposición en valores singulares de \mathbf{Z} como,

$$\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T \quad (13)$$

Esta perspectiva, llamado RV-ACP, es utilizada en ocasiones para analizar correlaciones múltiples o matrices de covarianza. Proporciona la estructura de similitud de las tablas, e información sobre los valores de qué pares de observaciones, contribuyen al patrón de similitud medido por $(\mathbf{W}_h, \mathbf{W}_h')$, (Abdi et al., 2012).

Además de proporcionar una representación visual de la estructura de similitud, la descomposición en valores propios de la matriz \mathbf{S} , también proporciona pesos óptimos para combinar las tablas en un “*compromiso*”.

La descomposición en valores propios de la matriz \mathbf{S} siempre proporciona un primer vector propio cuyos elementos tienen el mismo signo, los cuales convenientemente se consideran positivos. Esta particularidad se debe al teorema de Perron-Frobenius, que determina que las matrices semi-definidas

positivas, cuyos elementos son todos positivos, siempre tienen un primer vector propio con todos los elementos del mismo signo.

El valor de una tabla para el primer vector propio refleja su similitud general con todas las otras tablas. Con esto se pueden utilizar los valores de este primer vector propio para ponderar las tablas con el fin de dar más importancia a las tablas que representan bien el grupo. Entonces, si el primer vector propio de \mathbf{S} se denota como \mathbf{a}_1 , el conjunto de pesos óptimos para las tablas se almacena en un vector $\boldsymbol{\gamma}_{H \times 1}$ y se calcula re-escalando \mathbf{a}_1 de modo que la suma de los elementos de $\boldsymbol{\gamma}$ sea igual a uno:

$$\boldsymbol{\gamma} = \mathbf{a}_1 \times (\mathbf{a}_1^T \mathbf{1})^{-1} \quad (14)$$

Los pesos $\boldsymbol{\gamma}$ se pueden reunir en un vector $\mathbf{c}_{J \times 1}$ donde a cada variable se le asigna el peso $\boldsymbol{\gamma}$ de la matriz a la que pertenece. Específicamente, \mathbf{c} se construye como:

$$\mathbf{c} = [\boldsymbol{\gamma}_1 \mathbf{1}_1^T, \dots, \boldsymbol{\gamma}_h \mathbf{1}_h^T, \dots, \boldsymbol{\gamma}_H \mathbf{1}_H^T] \quad (15)$$

donde $\mathbf{1}_h$ representa un vector \mathbf{J}_h de unos. Alternativamente, los pesos se pueden almacenar como elementos de la diagonal principal de una matriz diagonal denotada como \mathbf{C} obtenida como,

$$\begin{aligned} \mathbf{C} &= \text{diag}\{\mathbf{c}\} \\ &= \text{diag}\{[\boldsymbol{\gamma}_1 \mathbf{1}_1^T, \dots, \boldsymbol{\gamma}_h \mathbf{1}_h^T, \dots, \boldsymbol{\gamma}_H \mathbf{1}_H^T]\} \end{aligned} \quad (16)$$

2.3.1.3. Análisis del compromiso

Una vez obtenidos los pesos, se utilizan para calcular la descomposición generalizada de valores singulares de \mathbf{X} bajo las restricciones proporcionadas por \mathbf{M} , una matriz de masas la cual se construye con las masas \mathbf{m}_i asignadas a cada observación de la matriz \mathbf{X} ; estas masas se recogen en el vector de masa, denotado \mathbf{m} , que conforman la matriz \mathbf{M} ; todas las masas son positivas o nulas y su suma es igual a uno.

$$\mathbf{M} = \text{diag}(\mathbf{m}) \quad (17)$$

y \mathbf{C} , que corresponde a los pesos óptimos para las \mathbf{H} tablas. La descomposición generalizada de valores singulares se expresa como:

$$\mathbf{X} = \mathbf{G} \mathbf{O} \mathbf{P}^T \quad (18)$$

donde

$$\mathbf{G}^T \mathbf{M} \mathbf{G} = \mathbf{P}^T \mathbf{C} \mathbf{P} = \mathbf{I} \quad (19)$$

La descomposición generalizada de valores singulares corresponde a un análisis de componentes principales generalizado de la matriz \mathbf{X} y, en consecuencia, proporcionará puntajes de factores para describir las observaciones y las cargas factoriales para describir las variables. Cada columna de \mathbf{G} y \mathbf{P} se refiere a una componente principal también llamada dimensión. En ACP, la ecuación 19 se puede reescribir como.

$$\mathbf{X} = \mathbf{F} \mathbf{P}^T \quad (20)$$

con

$$\mathbf{F} = \mathbf{G} \mathbf{\Theta} \quad (21)$$

donde \mathbf{F} contiene los puntajes factoriales (que describen las observaciones) y \mathbf{P} los puntajes de las cargas (que describe las variables).

Debido a que la matriz \mathbf{X} consta de \mathbf{H} tablas, cada una de las cuales contiene \mathbf{J}_h variables, la matriz \mathbf{P} de vectores singulares izquierdos puede dividirse de la misma manera que \mathbf{X} . Específicamente, \mathbf{P} puede expresarse como una matriz columna como:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_h \\ \vdots \\ \mathbf{P}_H \end{bmatrix} = [\mathbf{P}_1^T \mid \dots \mid \mathbf{P}_h^T \mid \dots \mid \mathbf{P}_H^T]^T \quad (22)$$

siendo \mathbf{P}_h es una matriz de dimensión $\mathbf{J}_h \times \mathbf{R}$, donde \mathbf{R} es el rango de \mathbf{X} . La matriz \mathbf{P} contiene los vectores singulares correspondientes a las variables de la matriz \mathbf{X}_h , así la ecuación 18 puede expresarse como:

$$\begin{aligned} \mathbf{X} &= (\mathbf{X}_1 \mid \dots \mid \mathbf{X}_h \mid \dots \mid \mathbf{X}_H) = \mathbf{G} \mathbf{\Theta} \mathbf{P}^T \\ &= \mathbf{G} \mathbf{\Theta} \left[(\mathbf{P}_1^T \mid \dots \mid \mathbf{P}_h^T \mid \dots \mid \mathbf{P}_H^T)^T \right]^T \\ &= \mathbf{G} \mathbf{\Theta} (\mathbf{P}_1^T \mid \dots \mid \mathbf{P}_h^T \mid \dots \mid \mathbf{P}_H^T) \\ &= (\mathbf{G} \mathbf{\Theta} \mathbf{P}_1^T \mid \dots \mid \mathbf{G} \mathbf{\Theta} \mathbf{P}_h^T \mid \dots \mid \mathbf{G} \mathbf{\Theta} \mathbf{P}_H^T) \end{aligned} \quad (23)$$

la ecuación 18 no se generaliza completamente en la ecuación 23 porque, si definimos \mathbf{C}_h como,

$$\mathbf{C}_h = \gamma_h \mathbf{I} \quad (24)$$

tenemos, en general que, $\mathbf{P}_h^T \mathbf{C}_h \mathbf{P}_h \neq \mathbf{I}$.

El puntaje factorial para \mathbf{X} representa el mejor compromiso, es decir la mejor representación común, para el conjunto de las \mathbf{H} matrices. Los puntajes factoriales, llamados puntajes factoriales compromiso, se calculan en 18 y 20 como,

$$\mathbf{F} = \mathbf{G}\mathbf{\Theta} \quad (25)$$

Los puntajes factoriales se pueden usar para trazar las observaciones como se hace en el ACP para el cual cada columna de \mathbf{F} representa una dimensión. La varianza de las puntuaciones de los factores de las observaciones se calcula utilizando sus masas, las cuales están contenidas en la matriz \mathbf{M} y se puede encontrar como la diagonal de la matriz $\mathbf{F}^T \mathbf{M} \mathbf{F}$. Para cada dimensión, esta varianza es igual al cuadrado del valor singular de esta dimensión,

$$\mathbf{F}^T \mathbf{M} \mathbf{F} = \mathbf{\Theta} \mathbf{G}^T \mathbf{M} \mathbf{G} \mathbf{\Theta} = \mathbf{\Theta} \quad (26)$$

Como en **ACP**, \mathbf{F} se puede obtener de \mathbf{X} combinando las ecuaciones 18 y 25 para obtener:

$$\mathbf{F} = \mathbf{G}\mathbf{\Theta} = \mathbf{XCP} \quad (27)$$

Teniendo en cuenta la estructura de bloques de \mathbf{X} , \mathbf{C} y \mathbf{P} , la ecuación 18 también puede reescribirse como:

$$\begin{aligned} \mathbf{F} &= \mathbf{XCP} = (\mathbf{X}_1 | \dots | \mathbf{X}_h | \dots | \mathbf{X}_H) \mathbf{X} \mathbf{C} \mathbf{X} \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_h \\ \vdots \\ \mathbf{P}_H \end{bmatrix} \\ &= \sum_h \mathbf{X}_h \mathbf{C}_h \mathbf{P}_h \\ &= \sum_h \gamma_h \mathbf{X}_h \mathbf{P}_h \end{aligned} \quad (28)$$

Esta ecuación sugiere que los puntajes factoriales parciales para una tabla pueden definirse como las proyecciones de esta tabla en sus vectores singulares

(\mathbf{P}_k). Específicamente, los puntajes parciales del factor para el estudio h-ésimo están contenidos en una matriz denominada \mathbf{F}_h calculada como

$$\mathbf{F}_h = \mathbf{X}_h \mathbf{P}_h \quad (29)$$

La matriz de puntajes factoriales del compromiso es el baricentro de los puntajes factoriales parciales porque es el promedio ponderado de los puntajes de los factores parciales, en donde los pesos están dados por los γ .

$$\sum_h \gamma_h \mathbf{F}_h = \sum_h \gamma_h \mathbf{X}_h \mathbf{P}_h = \mathbf{F} \quad (30)$$

Además, como en el ACP estándar, los elementos de \mathbf{P} son de cargas y se pueden trazar solos o junto con los puntajes factoriales como un biplot.

Como las cargas, que corresponden a las variables de una tabla, viene en bloques, es práctico crear un biplot con los puntajes factoriales parciales, \mathbf{F}_h para un bloque y las cargas \mathbf{P}_h para el otro. Al hacerlo, a menudo es práctico normalizar las cargas de modo que su varianza sea calculable con la varianza de los puntajes factoriales, por ejemplo, normalizando, para cada dimensión, las cargas de un bloque de modo que su varianza sea igual al cuadrado del valor singular de la dimensión o incluso al valor singular en sí mismo. Estos biplots son útiles para comprender la estructura estadística de cada bloque, a pesar de que las posiciones relativas de los puntajes factoriales y las cargas no son directamente interpretables, ya que solamente las proyecciones de las observaciones en los vectores de carga se pueden interpretar de manera significativa en un biplot.

Una representación gráfica alternativa de las variables y los componentes traza las correlaciones entre las variables originales de \mathbf{X} y las puntuaciones factoriales. Estas correlaciones se trazan como gráficos bidimensionales en los que también se traza un círculo de radio uno, denominado círculo de correlación. Cuanto más cerca del círculo estén las variables, mejor será "explicada" esta variable por los componentes utilizados para crear la gráfica. La [Figura 26](#) muestra es esquema de los diferentes pasos del método STATIS.

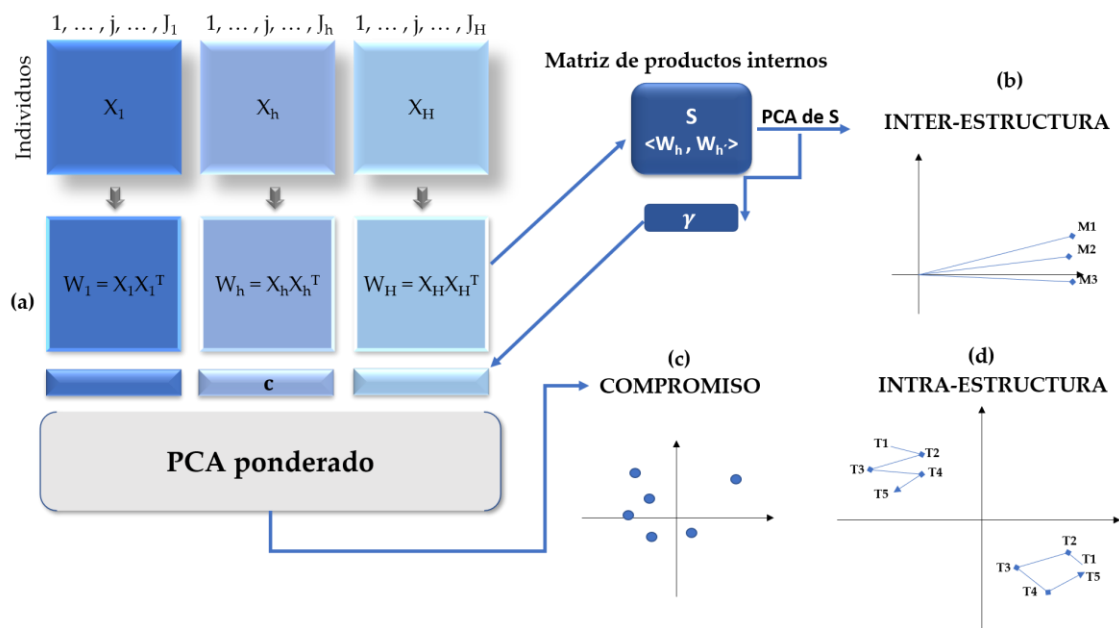


Figura 26. Descripción general del STATIS.

El producto escalar captura las covarianzas de las matrices (a); A partir de la matriz de correlaciones vectoriales y aplicando PCA, se construye una imagen en un subespacio euclídeo (b); Se construye la configuración compromiso, que es una media ponderada de las matrices W_h (c); Se proyecta la información original para analizar las trayectorias, la cual es el cambio de posición de una variable o individuo a lo largo del tiempo, representado sobre la matriz compromiso (d).

2.3.2. STATIS-DUAL

En el STATIS-dual, los datos consisten en conjuntos de H observaciones medidas en el mismo conjunto de variables (ver Figura 27). Aquí, en lugar de calcular H matrices de productos cruzados entre las observaciones, se calculan H matrices de covarianza entre las variables, una por cada conjunto de observaciones. El enfoque del STATIS-dual sigue los mismos pasos que el STATIS y proporciona un mapa del compromiso para las variables, en lugar de las observaciones como en el STATIS, y cargas parciales para cada tabla.

Al igual que en el STATIS, permite el análisis simultáneo de distintas matrices numéricas, pero en este caso la información se recoge sobre las mismas variables (columnas), medidas en diferentes conjuntos de individuos (filas). Se comparan las nubes de variables normadas por medio de las matrices de correlación.

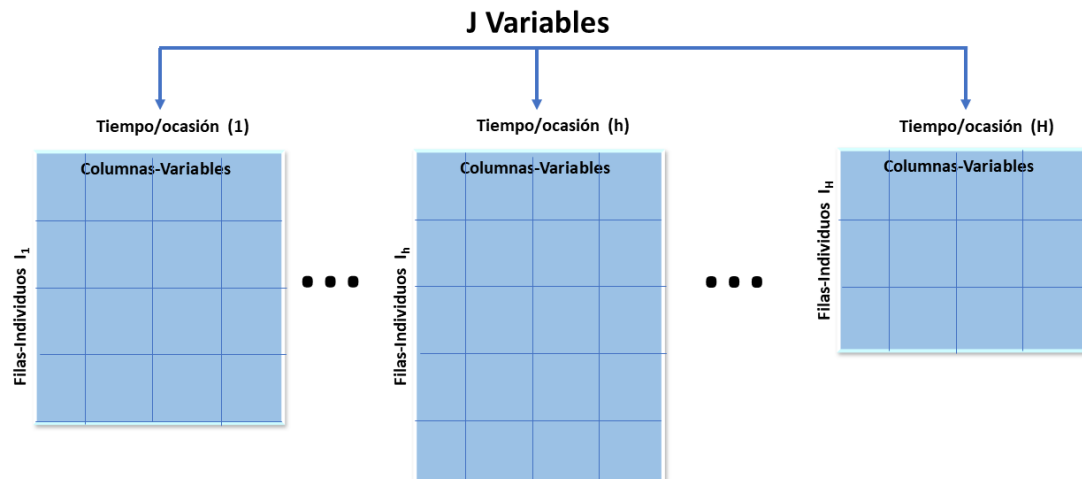


Figura 27. Estructura de datos en STATIS-dual

Los productos escalares entre matrices de correlación definen una configuración de diversos puntos, en la que cada uno de ellos representa a una de las matrices (nubes de puntos). Al igual que en el STATIS, la idea es encontrar una matriz compromiso, próxima a todas las matrices de correlación. Esta se define como una media ponderada de estas matrices, siendo, por tanto, una matriz de correlación.

A partir de esta matriz compromiso se construye la nube compromiso de variables. La metodología STATIS proporciona, además, en todas las versiones, y a partir de los matrices compromiso, una serie de resultados bajo la forma de nubes de puntos que serán explotadas de manera gráfica, a través de planos factoriales que (a diferencia de lo que ocurre en análisis en componentes principales y en análisis de correspondencias) no pasan forzosamente por el centro de gravedad de la nube. La ponderación que, en ocasiones, utiliza este método no equilibra la influencia de las diferentes tablas, sino que asigna mayor peso a aquellas que presentan una estructura similar a la estructura común, penalizando, en cierto sentido, al resto.

El método STATIS-dual considera las H tablas $Y_1, \dots, Y_h, \dots, Y_H$ con $I_1; \dots, I_H$ observaciones (individuos) y J variables; cada uno de estos representa diferentes estudios, escenarios o momentos. Al igual que STATIS estándar, las tablas deben pre-procesarse utilizando métodos como centrado, escalado y/o normalización de datos como indican [Abdi et al. \(2012\)](#) y [Filho \(Marcondes Filho et al., 2017\)](#) entre otros.

El método STATIS-dual permite representar las matrices de datos correspondientes a las diferentes ocasiones como puntos en un espacio vectorial de baja dimensión, lo cual se hace utilizando las matrices de covarianzas.

En la imagen euclídea resultante, la distancia entre puntos se interpreta en términos de similitud y, por ende, en semejanza entre estructura de varianza-covarianza y congruencia entre estructuras factoriales. Las estructuras serán semejantes si los ángulos formados por los vectores de la imagen euclídea se aproximan a cero.

2.3.2.1. Interestructura

En el proceso de la Interestructura, se estudia la relación entre las distintas matrices, el propósito es hallar una matriz de correlaciones vectoriales entre matrices; en otras palabras, la diferenciación global entre tablas de datos. La finalidad es analizar configuraciones de los H puntos que corresponden a las H matrices en la representación gráfica de una o más imágenes euclídeas en el plano de la proyección de los H puntos.

Con tal fin, la interestructura se representa en un subespacio de dimensión reducida, descomponiendo espectralmente la matriz de correlaciones vectoriales y proyectándola.

Se define al *objeto* que representa cada matriz, se elige una *métrica* en el espacio de los objetos y se determina una *imagen euclídea* de dichas matrices, asociada a los productos escalares introducidos en la etapa anterior.

La proximidad entre dos puntos corresponde a la similitud (con respecto a la distancia considerada) entre las matrices correspondientes a esos puntos.

De esta forma se obtienen las H matrices preprocesadas $X_1 \dots X_H$ cada una con dimensión $I_h \times J$. Estas matrices se apilan verticalmente para estructurar la matriz $X_{I \times J}$, donde $I = I_1 + \dots + I_H$. Para cada tabla X_h , se obtiene una matriz $J \times J$ de productos cruzados, de la siguiente manera:

$$\mathbf{C}_h = \mathbf{X}_h^T \mathbf{X}_h \quad (31)$$

Cada matriz simétrica de productos cruzados \mathbf{C}_h se vectoriza. Los vectores obtenidos se almacenan en una nueva matriz $\mathbf{Z}_{H \times J^2}$.

La matriz $\mathbf{A}_{H \times H}$ se calcula a partir de la matriz $\mathbf{Z}\mathbf{Z}^T$. Esta matriz semi-definida positiva permite representar en dos dimensiones, cada una de las H tablas usando la descomposición de valores propios, considerando que los valores propios se ordenan de mayor a menor, tenemos:

$$\mathbf{A} = \mathbf{U}\mathbf{\Theta}\mathbf{U}^T \quad (32)$$

donde \mathbf{U} es una matriz que contiene los vectores propios de \mathbf{A} y $\mathbf{\Theta}$ es una matriz diagonal que recoge los correspondientes valores propios.

Una forma opcional de calcular \mathbf{U} es realizar una Descomposición de valor singular (DVS) de la matriz \mathbf{Z} :

$$\mathbf{Z} = \mathbf{U}\mathbf{T}\mathbf{V}^T \quad (33)$$

con

$$\mathbf{\Gamma}^2 = \mathbf{\Theta} \quad (34)$$

Sea L el rango de la matriz \mathbf{Z} , $\mathbf{U}_{H \times L}$ la matriz que contiene los vectores singulares izquierdos de \mathbf{Z} , $\mathbf{\Gamma}$ una matriz diagonal de L x L que recoge los valores singulares de \mathbf{Z} , y \mathbf{V} contiene los vectores singulares derechos de \mathbf{Z} en una matriz $J^2 \times L$.

La descomposición en valores singulares (DVS) de \mathbf{A} permite representar tablas como puntos en un diagrama de ACP bidimensional llamado espacio interestructura, utilizando la primera y segunda columna de la matrix \mathbf{U} como coordenadas:

$$\mathbf{G} = \mathbf{U}\mathbf{\Gamma} = \mathbf{Z}\mathbf{V} \quad (35)$$

Esta representación particular, también llamada biplot, fue presentada por primera vez por [Gabriel \(1971\)](#) y luego modificada por [Galindo-Villardón \(1986\)](#). Un biplot permite que la información sobre individuos y variables de una matriz de datos se muestre gráficamente.

Del espacio interestructura, se deriva un conjunto óptimo de pesos para las tablas, para calcular la mejor representación común de las J variables ([Abdi et al., 2012](#)). Estos pesos se determinan reescalando el primer vector propio u_1 de \mathbf{A} y dividiéndolo por la suma \hat{u}_1 de sus elementos, definiendo el vector m como,

$$m = \frac{u_1}{\hat{u}_1} \quad (36)$$

Luego, la matriz diagonal **M** de pesos óptimos se calcula de la siguiente manera. Los valores m_i de m se apilan y cada valor se repite I_h veces de acuerdo con el número de observaciones en cada tabla. **M** es una matriz cuadrada de orden $I = I_1 + \dots + I_H$.

2.3.2.2. Intraestructura

En la Intraestructura se asigna una masa, denominada α_j , a cada variable. Las masas son elementos no negativos cuya suma es igual a uno. Aunque se pueden calcular diferentes masas para cada variable, a menudo se eligen masas iguales para asegurar que todas las variables tengan la misma importancia para el análisis (Abdi et al., 2012).. Se obtiene una matriz diagonal **D'**, para las masas de las variables, dividiendo la matriz de identidad $I_{J \times J}$ por las **J** de variables:

$$\mathbf{D}' = \frac{I_{J \times J}}{J} \quad (37)$$

Ahora la tripleta (**X**; **M**; **D'**) está conformado por tablas preprocesadas, los pesos y masas.

Usando la descomposición generalizada de valores singulares (DGVS) de la tripleta, se calcula la siguiente descomposición:

$$\mathbf{X} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^T \quad (38)$$

con

$$\mathbf{P}^T\mathbf{M}\mathbf{P} = \mathbf{Q}^T\mathbf{D}'\mathbf{Q} = \mathbf{I} \quad (39)$$

Sea T el rango de la matriz **X**, $\mathbf{P}_{I \times T}$ una matriz que contiene los vectores singulares izquierdos de **X**, $\mathbf{\Delta}$ una matriz diagonal $T \times T$ con los valores singulares, y $\mathbf{Q}_{J \times T}$ contiene los vectores singulares derechos.

Esta descomposición proporcionará un espacio de compromiso para la representación de variables en un diagrama de dimensiones reducidas. Las puntuaciones del factor de compromiso permiten trazar las J variables en el compromiso. Estos puntajes se calculan como la primera y segunda columnas de la matriz $J \times T$:

$$\mathbf{F} = \mathbf{Q}\mathbf{\Delta} = \mathbf{X}^T\mathbf{M}\mathbf{P} = \sum_h \mathbf{m}_k \mathbf{X}_h^T \mathbf{P}_h \quad (40)$$

Tenga en cuenta que los puntajes factoriales parciales se pueden definir como la proyección de la tabla h-ésima en sus vectores singulares izquierdos (Abdi et al., 2012).. Luego, los puntajes de factores parciales para esta tabla se calculan como la primera y segunda columna de la matriz,

$$F_h = X_h^T P_h \quad (41)$$

Como se sugiere en la ecuación (40), la matriz **F** de puntajes factoriales del compromiso, también es el promedio ponderado de los puntajes factoriales parciales y se considera el baricentro (Abdi et al., 2012).

Por lo tanto, para calcular las puntuaciones factoriales parciales, la matriz **P_{IxT}** está horizontalmente particionada en H matrices de dimensión $I_h \times T$, denotadas como **P_h** con el mismo número de filas que **X_h**. La Matriz **P_h** se calcula de la ecuación (38),

$$P = X D' Q \Delta^{-1} \quad (42)$$

$$P_h = X_h D' Q \Delta^{-1} \quad (43)$$

La figura 28 nos muestra un esquema de los pasos del STATIS dual.

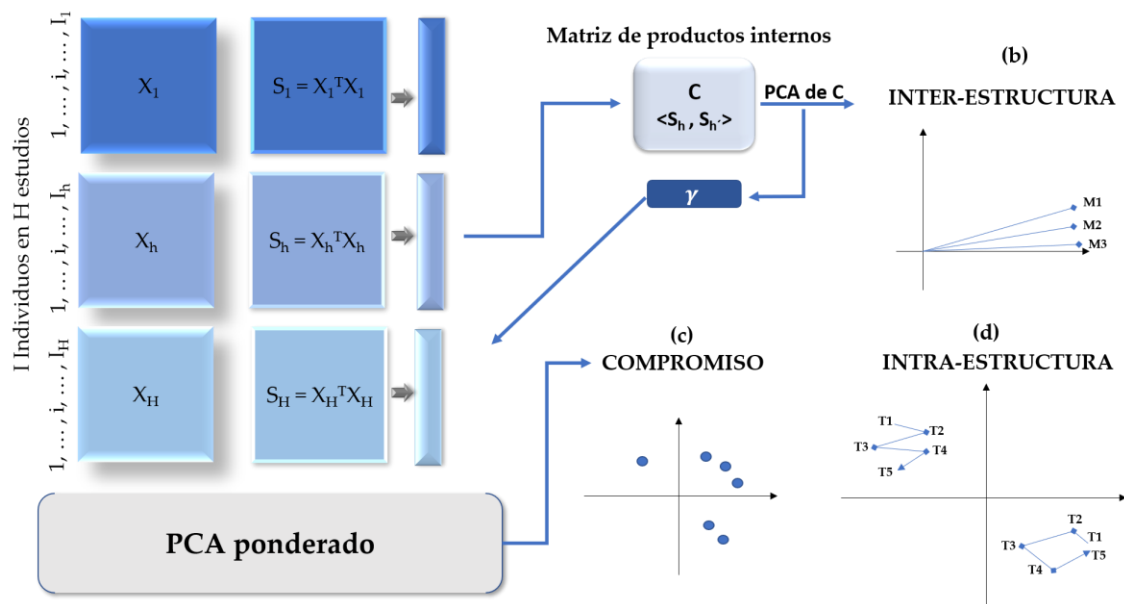


Figura 28. Esquema de los procedimientos del STATIS-dual

2.4. NUEVOS APORTES A LOS MÉTODOS STATIS

De acuerdo a [Vicente-Galindo \(2013\)](#), los nuevos aportes a los métodos de la familia STATIS se pueden clasificar de la siguiente manera: según los datos de partida, a los pesos asignados al crear la matriz consenso, a si se toma en cuenta información externa, o a si se dispone de pares de tablas en H situaciones o tiempos diferentes.

Así tenemos que de acuerdo a los datos de partida se pueden presentar tres formas:

el caso en que todas las tablas de datos incluyan información sobre las mismas variables observadas sobre los mismos individuos, en momentos o escenarios diferentes. Para estos casos se emplea el X-STATIS o Análisis Parcial Triádico (PTA) ([Jaffrenou, 1978](#)), que trabaja con las matrices de datos y con operadores. otro caso es cuando se trabaja con los mismos sujetos y las mismas variables, partiendo de una matriz de operadores. Las técnicas que utilizan esta metodología son el COVSTATIS ([Thioulouse, 2011](#)), y el DISTATIS ([Abdi, O'Toole, Valentin, & Edelman, 2005](#); [Abdi et al., 2007](#)).

el caso en que los individuos tengan estructura de grupos, se aplica la técnica CANOSTATIS, ([Vallejo-Arboleda et al., 2007](#)).

Por otro lado y según [Vicente-Galindo\(2013\)](#), de acuerdo a los pesos asignados al crear la matriz consenso se tiene una versión de STATIS en la cual el peso que se da a cada matriz es diferente a la propuesta de L'HERMIER, esta es el Power-STATIS ([Bénasséni & Bennani-Dosse, 2012](#)), en la cual el peso está relacionado con las componentes del primer vector propio de la matriz de correlaciones vectoriales.

Otra de las clasificaciones señaladas por [Vicente-Galindo \(2013\)](#), es cuando se tiene en cuenta información externa, las cuales se pueden presentar de dos formas:

- cuando se consideran las H matrices que se integran en el STATIS y una matriz más con información externa sobre los individuos, para estos caso [Sauzay et al. \(2006\)](#), proponen el $t+1$ STATIS. Más adelante se propone

el STATIS-4 (Sabatier & Vivien, 2008), el cual es una extensión al t+1 STATIS.

- cuando los individuos tengan estructura de grupos, se tiene el CANOSTATIS, (Vallejo-Arboleda et al., 2007).

En cuarto caso corresponde a la situación en donde se dispone de pares de tablas en H situaciones o tiempos diferentes. Dentro de las técnicas de este tipo se tienen el STATICO (Simier, Blanc, Pellegrin, & Nandris, 1999; Thioulouse et al., 2004) que aplica COINERCIA y STATIS, y el COSTATIS, (Thioulouse, 2011), que aplica STATIS y luego COINERCIA.

2.4.1. X-STATIS O PTA

El X-STATIS (Abdi et al., 2012) fue propuesto por Jaffrenou (1978). Inicialmente, Thioulouse y Chessel (1987) lo llamaron Análisis triádico, más tarde Análisis triádico completo de Kroonenberg (1989). A diferencia de otros métodos de la familia STATIS, el X-STATIS no usa operadores, ya que trabaja directamente con las matrices originales, es más restrictivo porque supone las mismas J variables medidas en los mismos I individuos, pero tiene la ventaja de no perder la información original al no trabajar con operadores, permitiendo más representaciones (Vicente-Galindo, 2013). Esencialmente, es una herramienta exploratoria para el análisis de datos de tres vías y contempla tres fases esenciales de STATIS estándar: la inter-estructura, el análisis del compromiso y la intra-estructura, (Lavit et al., 1994), ver Figura 29.

Primero, se calcula una matriz de productos escalares entre tablas. La diagonalización de esta matriz proporciona vectores propios. Los H coeficientes α_h de los primeros vectores propios se utilizan para ponderar las H tablas en el cálculo de la matriz de compromiso. A continuación, se realiza un ACP para establecer la ordenación de las diferentes matrices. Del mismo modo, se puede utilizar una matriz de correlaciones vectoriales (RV) (Robert & Escoufier, 1976), para re-escalar la importancia de las tablas. Cada tabla se proyecta en el plan factorial obtenido del análisis y se representa con una flecha, para establecer la ordenación de las tablas, que resume la estructura global y las relaciones entre las tablas. Esta configuración, basada en la matriz de covarianza, permite una

comparación gráfica general de las tablas y muestra las proximidades entre las configuraciones de las mismas observaciones.

A continuación, se construye el compromiso, que consiste en una matriz media de máxima inercia (la matriz compromiso). Esta matriz, se deriva de las sub-matrices iniciales en proporción a su peso. El \cos^2 , un valor adimensional, representa un indicador de cuánto estado de compromiso contiene la información contenida en cada sub-matriz. Es decir, la matriz que representa las correlaciones vectoriales entre las diferentes sub-matrices (coeficientes RV), proporciona una indicación de la fuerza de las relaciones entre las diferentes sub-matrices de las distintas tablas, (Robert & Escoufier, 1976).

Finalmente se determinan las trayectorias, se resumen la variabilidad de la serie de tablas en torno a la estructura común definida por el compromiso. Las filas y columnas de todas las tablas se proyectan en el mapa de factores del ACP del compromiso como elementos adicionales. Los puntos se pueden vincular por líneas para subrayar sus trayectorias (Thioulouse et al., 2004).

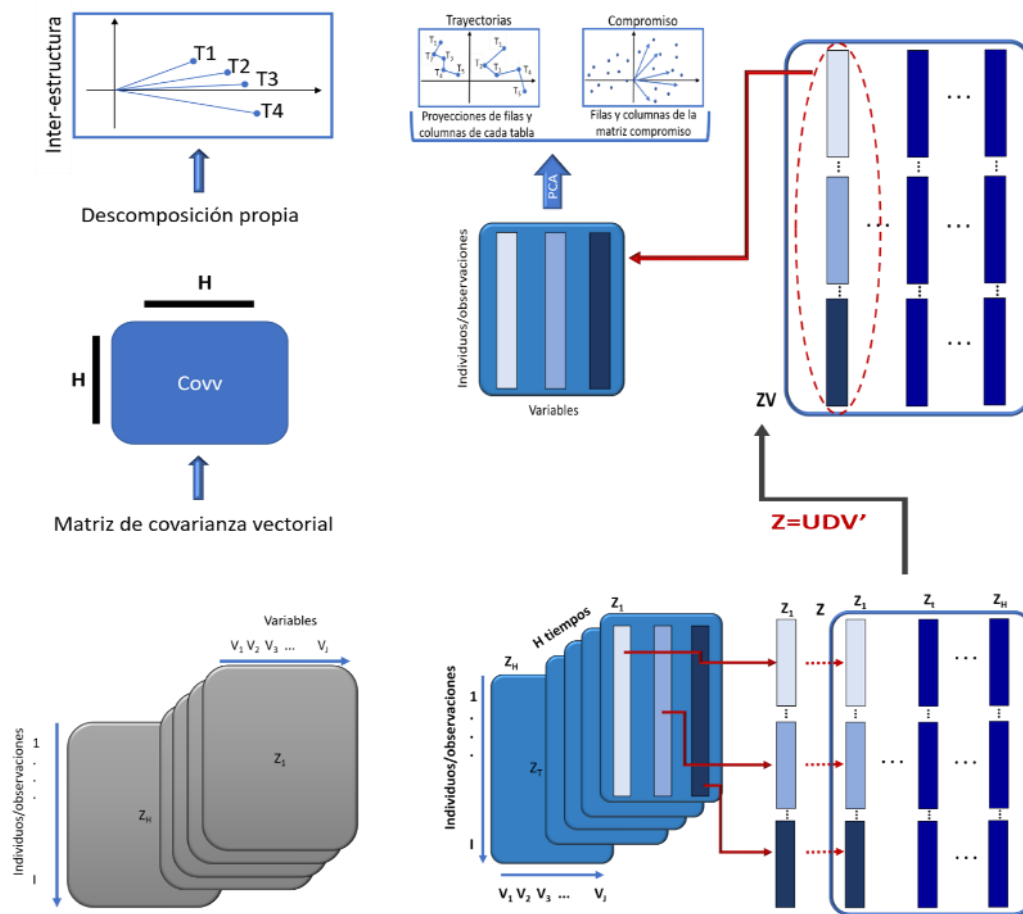


Figura 29. Esquema del Método X-STATIS o PTA, adaptado de Amores et al, 2017

Analíticamente, para la inter-estructura, se construye una matriz de productos escalares entre tablas, donde el elemento en la fila h y la columna ℓ es,

$$\text{Covv}(\mathbf{Z}_h, \mathbf{Z}_\ell) = \text{traza} (\mathbf{Z}_h^T \mathbf{M}_I \mathbf{X}_\ell \mathbf{M}_J) \quad (44)$$

donde \mathbf{Z}_h es la h -ésima tabla de la secuencia y \mathbf{M}_I , \mathbf{M}_J son las dos métricas para las filas y columnas, respectivamente. Alternativamente, se puede usar una matriz de correlación vectorial que reescala la el vector de covarianza:

$$\text{RV}(\mathbf{Z}_h, \mathbf{Z}_\ell) = \frac{\text{Covv}(\mathbf{Z}_h, \mathbf{Z}_\ell)}{\sqrt{(\text{Covv}(\mathbf{Z}_h, \mathbf{Z}_h) \text{Covv}(\mathbf{Z}_\ell, \mathbf{Z}_\ell))}} \quad (45)$$

Para trazar la inter-estructura, se utilizan los vectores que inician en el origen y terminan en los puntos dados por las filas de $\mathbf{V}_2 \mathbf{D}_2$, siendo \mathbf{V}_2 los primeros dos

vectores propios de la matriz de covarianza del vector y D_2 la matriz diagonal con los dos valores propios asociados.

El compromiso Z_c es una combinación lineal de Z_h ponderada por γ_h , las coordenadas del primer vector propio de la inter-estructura, que se suponen todas positivas, ya que la matriz de covarianza vectorial es simétrica y todos sus elementos son positivo, así

$$Z_c = \sum_h \gamma_h Z_h \quad (46)$$

También, con el análisis de compromiso se pueden obtener las representaciones bidimensionales, gráficos de ejes principales, para interpretar su estructura.

Este proceso para calcular la matriz de compromiso es equivalente a realizar una descomposición de valores singulares de la matriz $Z = UDV^T$ obtenida colocando las columnas como vectores, concatenando las columnas de cada una de las matrices de la secuencia. Luego, la primera columna de ZV se tomará como compromiso, mostrándose como una matriz.

La intraestructura se obtiene proyectando las filas y las columnas de cada tabla de la secuencia en el análisis de compromiso. Sea que V_r la primera r matriz de vectores propios del análisis de compromiso. Las coordenadas de las filas de la tabla Z_h son las filas de $Z_h M_j V_r$, y las columnas son las filas de $Z_h^T M_1 U_r$ donde U_r contiene los primeros vectores propios de $Z_c M_j Z_c^T M_1$.

2.4.2. ANÁLISIS FACTORIAL MÚLTIPLE (AFM)

El AFM, es un método que permite analizar simultáneamente varios grupos de variables medidas sobre el mismo conjunto de individuos nivelando la influencia de cada grupo, fue desarrollado por [Escofier y Pagès \(1990\)](#). Las agrupaciones de variables pueden nacer del uso conjunto de variables de naturaleza heterogénea, cuantitativa y cualitativa, de la utilización de tablas que se derivan de otras de tres dimensiones o de un mismo conjunto de variables medidas en distintos periodos de tiempo. La posible estructuración en grupos de variables de los datos originales enriquece su estudio. Además de obtención de una tipología de los individuos definida a través del conjunto de variables, también

se tiene como objetivo la búsqueda de posibles relaciones entre las estructuras obtenidas en el seno de cada uno de los grupos.

Esta técnica se encuentra dentro de los denominados métodos asimétricos, los cuales tienen como propósito es definir tipologías de objetos, o bien el análisis de los tres modos o entrada. La primera fase de análisis en este tipo de métodos es el análisis de la inter-estructura con el fin de estudiar la relación entre las tablas que conforman la tabla global, y esto se hace a través de los operadores que las representan. La segunda fase corresponde al análisis del compromiso, en la cual quiere obtener una nube media (ponderada) que represente la estructura común de esas tablas Z_r . La tercera etapa corresponden al análisis de la intra-estructura cuyo propósito es analizar las relaciones entre individuos y variables en la matriz compromiso (Glacon, 1981).

La terminología del AFM es propia de la escuela francesa, se definen diferentes elementos al considerar múltiples tablas (Abascal & Landaluce, 2000), los cuales son:

- ⊕ Nube parcial: Es la nube de individuos caracterizados por un grupo de variables. Se obtienen tantas nubes parciales como grupos de variables se consideren.
- ⊕ Nube global: Es la nube de individuos caracterizados por todas las variables de la tabla global, siendo la unión de todas las nubes parciales.
- ⊕ Puntos parciales: El AFM proporciona una representación de todas las nubes parciales superpuestas a la representación de los puntos medios. De este modo, se obtienen de forma gráfica cada una de las trayectorias de los individuos.
- ⊕ Puntos compromiso: el AFM también representa los elementos de la nube global en los planos factoriales definidos en el análisis global. Estos puntos hacen posible determinar la posición compromiso (media) del individuo, teniendo en cuenta todos los grupos de variables considerados.
- ⊕ Inercia Intra-estructura: Se refiere a la inercia de los puntos parciales respecto a su punto medio en la nube global, en otras palabras, es la inercia de las nubes correspondientes a la imagen de cada individuo desde los diferentes puntos de vista, es decir a través de los distintos grupos, estudiados con respecto a los centros de gravedad de cada individuo.

- ⊕ Inercia Inter-estructura: Se refiere a la inercia de los centros de gravedad de cada nube parcial respecto al centro de gravedad de la global.
- ⊕ Inercia Total: Es la variabilidad de la nube global respecto a su centro de gravedad. Se descompone, en forma aditiva, en inercia Intra e inercia Inter-estructura.

La finalidad de Análisis Factorial Múltiple (AFM) es la de inspeccionar la presencia de estructuras comunes a todas, o parte de las nubes parciales, lo que se refleja en una inercia intra mínima o pequeña, o en una inercia inter máxima o elevada. Los individuos cuyos puntos parciales son próximos, reflejan una inercia intra débil y muestran la estructura común que poseen los datos analizados. Por otro lado, los individuos con puntos parciales más distantes reflejan una inercia intra alta y constituyen las excepciones a tal estructura común.

Supongamos una tabla de datos denotada por \mathbf{X} , la cual cruza el conjunto I de individuos con el conjunto J de todas las variables sin diferenciar las \mathbf{J}_k variables de cada tabla y \mathbf{X}_k la tabla que cruza I con todas las variables del grupo \mathbf{J}_k ; por tanto,

$$\mathbf{J} = \bigcup_k \mathbf{J}_k$$

Se supone que I, J, K, \mathbf{J}_k designan a la vez el conjunto y su cardinal.

La matriz de pesos para los individuos se denota por \mathbf{D} , si todos los individuos cumplen el mismo papel, todos tendrán el mismo peso $1/I$. No obstante, se dan ocasiones en las que se necesita dar diferentes pesos a los individuos. Los pesos toman parte en el cálculo de las medias de cada variable y en la medida de asociación entre ellas. La matriz de pesos para las variables la podemos denotar por \mathbf{W} para la tabla global \mathbf{X} y \mathbf{W}_k para las variables de la tabla \mathbf{X}_k . Generalmente, en casi todos los casos, las variables intervienen de la misma forma, dándoles un peso igual a 1, pero pese a esto si se desea dar mayor o menor relevancia la influencia que tenga alguna de las variables, se puede cambiar su peso. Dichos pesos afectan las distancias entre los individuos dado que los pesos de las variables ponderan la relevancia en cada columna de las tablas.

Con el propósito de analizar la tabla \mathbf{X}_k se considerara el estudio $(\mathbf{X}_k, \mathbf{W}_k, \mathbf{D})$. En el análisis de la tabla \mathbf{X} , la introducción simultánea de varios conjuntos requiere

nivelar la influencia de éstos, lo cual se consigue al multiplicar los pesos iniciales de todas las variables del grupo J_k por el coeficiente α_k .

2.4.3. STATIS-COA

Gaertner, Chessel y Bertrand (1998), proponen un método que combina el STATIS con la lógica del análisis de correspondencia (CoA), para estudiar la estabilidad de la organización espacial de ensamblajes demersales en términos de composición de especies. El método combina el método STATIS (Lavit et al., 1994), y el análisis de correspondencias (Benzécri, 1973).

La primera etapa de STATIS-CoA consiste en calcular una matriz de productos escalares entre variables para cada tabla con el fin de estandarizar las dimensiones de las tablas de contingencia. Este paso permite comparar todas las muestras mediante el cálculo de una matriz de productos escalares (Robert & Escoufier, 1976) entre tablas de contingencia. La diagonalización de esta matriz proporciona vectores propios. Los k coeficientes del primer vector propio se utilizan para ponderar las k tablas de contingencia en el cálculo de una tabla de promedios (matriz compromiso). Esta ponderación permite la construcción de una matriz compromiso que contiene la parte común de las estructuras estudiadas. Es decir, la matriz compromiso es una combinación lineal de las k tablas iniciales calculadas con el objetivo de construir una tabla media de inercia máxima. Así, en la construcción de la matriz compromiso se le da una mayor importancia a las matrices que tienen estructuras similares y una limitada a las otras matrices. El análisis CoA de la tabla de compromiso define ejes y componentes que expresan la parte común de las estructuras estudiadas. Independientemente, se lleva a cabo un Análisis de Correspondencia por separado de cada una de las matrices. Este paso permite la proyección de los primeros ejes de los análisis separados de cada matriz sobre los primeros ejes de la matriz compromiso. Esta representación indica cuánto se tienen en cuenta las estructuras principales de cada matriz inicial en el compromiso (Lavit et al., 1994).

2.4.4. STATICO

El método STATICO (Simier et al., 1999; Thioulouse et al., 2004), analiza dos cubos de datos (ver Figura 30), para evaluar una estructura común entre pares de tablas (uno de cada cubo de datos) y la estabilidad de esa estructura, durante el tiempo o momentos.

El STATICO se ejecuta en tres pasos: el primer paso consiste en analizar cada tabla mediante un método básico, tradicionalmente un ACP, debidamente normalizado o centrado; luego, cada par de tablas está vinculado por un análisis de COINERCIA (Doledec & Chessel, 1994) que proporciona una imagen promedio de la coestructura existente; finalmente, se utiliza un Análisis Triádico Parcial (Thioulouse & Chessel, 1987), para analizar esta secuencia.

Las variables ambientales deben ser las mismas en todas las tablas, y las especies también (recordemos que esta técnica inicialmente se desarrolló para estudios ecológicos). Los sitios de muestreo (filas de las tablas) deben ser los mismos para las dos tablas de un par, pero pueden ser diferentes entre las series.

El paso entre estructuras de STATICO, al igual que el método entre estructuras del método STATIS, proporciona pesos óptimos que se utilizan para construir un compromiso. El compromiso de STATICO es una media ponderada de las tablas de covarianza cruzada. En análisis de este compromiso ofrece una visualización gráfica de las variables. Finalmente, el paso de Infraestructura proyecta las filas y columnas de cada tabla de la secuencia en el análisis del compromiso. Esto proporciona una visualización de las variables ambientales y las especies, y de los sitios de muestreo, tanto desde el punto de vista de las variables ambientales como de las especies.

Dadas \mathbf{X} una matriz $I \times J$, y \mathbf{Y} otra de $I' \times J'$. Sean $(\mathbf{X}_h, \mathbf{D}_j, \mathbf{D}_{I_h})$ y $(\mathbf{Y}_h, \mathbf{D}_{j'}, \mathbf{D}_{I_h})$ el par de tripletas correspondientes a la tabla h , donde $h = 1, \dots, H$.

\mathbf{X}_h es la tabla de variables medidas en la ocasión h , e \mathbf{Y}_h es la tabla de otro grupo de variables observadas en la misma ocasión. \mathbf{D}_j y $\mathbf{D}_{j'}$, son las mismas para todas las ocasiones y $\mathbf{D}_{I_h} = \text{diag}\left(\frac{1}{I_h}\right)$, el mismo para cada par de tablas.

Sea $\mathbf{Z}_h = \mathbf{Y}_h^T \mathbf{D}_{I_h} \mathbf{X}_h$, la h -ésimo matriz de productos cruzados, la tripleta $(\mathbf{Z}_h, \mathbf{D}_j, \mathbf{D}_{j'})$ corresponden a un análisis de coinercia.

El primer paso consiste en obtener la inter-estructura que proporciona los pesos óptimos γ_h de modo que la inercia de la tripleta $(\sum_h^H \gamma_h \mathbf{Z}_h, \mathbf{D}_J, \mathbf{D}_{J'})$ sea máxima con la restricción $\sum_1^H \gamma_h^2 = 1$.

Seguidamente se calcula el compromiso, que corresponde a una media ponderada de las tablas de productos cruzados utilizando pesos $\mathbf{Z} = \sum_1^H \gamma_h \mathbf{Z}_h$. El compromiso en STATICO, es el análisis del triplete $(\mathbf{Z}, \mathbf{D}_J, \mathbf{D}_{J'})$. Ofrece una visualización gráfica de las variables ambientales (filas de \mathbf{Z}) y de las especies (columnas de \mathbf{Z}).

Por último, la se determina la Infraestructura que se basa en las proyecciones habituales de filas y columnas.

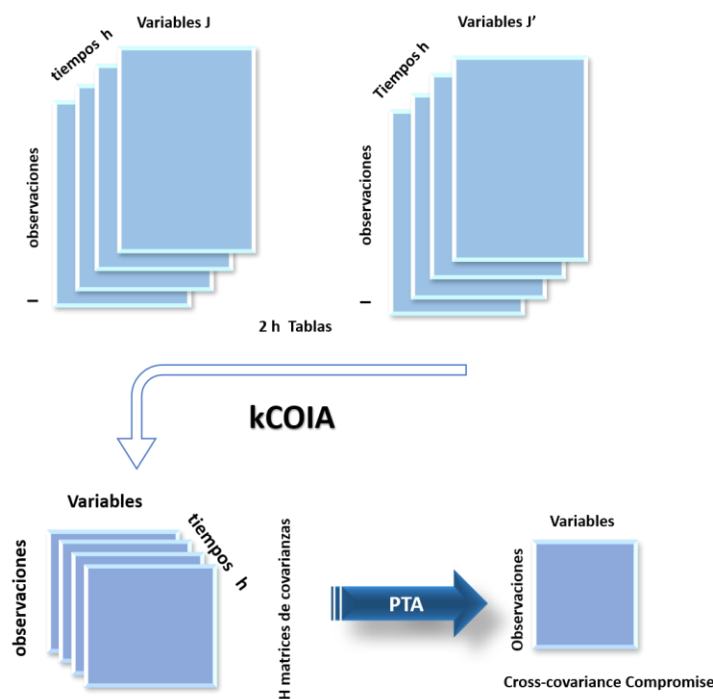


Figura 30. Esquema del método STATICO, adaptado de Thioulouse, 2011

2.4.5. DOUBLE-STATIS (DO-ACT)

Desarrollado por Vivien y Sabatier (2004), Double-STATIS-ACT o DO-ACT, es una extensión del STATIS para datos multibloque los cuales pueden presentar longitud variable. Existe un método que trabaja con dos multibloques, el STATICO (Simier et al., 1999), pero este se utiliza para el caso de que los dos

conjuntos de datos tengan la misma dimensión; es decir que los datos deben estar organizados en dos conjuntos de cubos que se puedan comparar. Mientras que con el DO-ACT se pueden estudiar distintos números de variables en cada matriz o tabla de datos y, además, tiene en cuenta la estructura de bloques de los datos en el proceso de los cálculos.

Se parte de dos multibloques, o dos conjuntos de matrices, $(\mathbf{X}_h)_h$ con $h=1, \dots, H$ e $(\mathbf{Y}_\ell)_\ell$ con $\ell=1, \dots, L$. H y L son las longitudes de los multibloques y supone que H y L no son simultáneamente iguales a uno. Cada \mathbf{X}_h e \mathbf{Y}_ℓ son tablas $I \times J_h$ y $I \times J_\ell$ respectivamente, donde todas las variables son medidas sobre los mismos individuos (u observaciones) y el supuesto es que deben estar centradas. A partir de las matrices que conforman los multibloques definimos los respectivos operadores \mathbf{W}_{X_h} y \mathbf{W}_{Y_ℓ} :

$$\mathbf{W}_{X_h} = \mathbf{X}_h \mathbf{X}_h^T \quad (47)$$

$$\mathbf{W}_{Y_\ell} = \mathbf{Y}_\ell \mathbf{Y}_\ell^T \quad (48)$$

Se determinan los compromisos representativos de cada multibloque con máxima correlación entre ellos. Es decir, se deben encontrar dos operadores compromiso, uno por bloque, de tal manera que cada compromiso es una combinación lineal de los operadores de partida y cuyos coeficientes maximicen la traza $(\mathbf{W}_{X_h} \mathbf{X} \mathbf{W}_{Y_\ell})$,

$$\mathbf{W}_X = \mathbf{v} \sum_{h=1}^H \gamma_h \mathbf{W}_{X_h} \quad (49)$$

y

$$\mathbf{W}_Y = \mathbf{v} \sum_{\ell=1}^L \delta_\ell \mathbf{W}_{Y_\ell} \quad (50)$$

con \mathbf{v} y \mathbf{v} dos coeficientes escalares.

Los dos operadores compromiso se conforman a partir de los vectores

$$\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_H)^T \text{ y } \boldsymbol{\delta} = (\delta_1, \dots, \delta_L)^T.$$

Los compromisos son simultáneamente sujetos a las restricciones

$$\|\gamma\|^2 = \gamma^T \gamma = \sum_{h=1}^H \gamma_h^2 = 1 \quad (51)$$

y

$$\|\delta\|^2 = \delta^T \delta = \sum_{h=1}^H \delta_h^2 = 1 \quad (52)$$

Es necesario utilizar una función Lagrangiana L , para hallar las soluciones. Sean σ y τ los multiplicadores asociados a las restricciones, entonces la función L será:

$$L(\gamma, \delta, \sigma, \tau) = \text{traza}(\mathbf{W}_X \mathbf{W}_Y) + \frac{1}{2} \sigma (1 - \|\gamma\|^2) + \frac{1}{2} \tau (1 - \|\delta\|^2) \quad (53)$$

Luego, las soluciones que maximizan la traza($\mathbf{W}_X \mathbf{W}_Y$), se expresan como,

$$\text{Max}_{\gamma, \delta} (\text{traza}(\mathbf{W}_X, \mathbf{W}_Y)) \quad (54)$$

y con las restricciones señaladas en (51) y (52) proporcionan los coeficientes buscados. Para más detalles consultar [Vivien y Sabatier, \(2004\)](#).

En los análisis de la técnica Double-STATIS (DO-ACT), se evalúan varios aspectos que listamos a continuación:

Las normas de Hilbert–Schmidt de los operadores de cada multibloque y los ángulos que se crean entre ellos para poder evaluar la proximidad entre operadores de cada bloque. El primer eje explica la mayor parte de la variabilidad, de la misma forma que en el STATIS.

Se contrasta la proximidad entre los operadores de uno y otro bloque.

Se examinan los RV del compromiso de cada mutibloque con los operados de ese bloque.

Se examinan los coeficientes RV entre los dos multibloques.

Se representan las observaciones compromiso sobre las dos primeras componentes del espacio euclídeo compromiso Double-STATIS.

Se calculan las correlaciones de las variables de cada matriz (para cada bloque) con el Double-STATIS compromiso.

Se analiza la asociación entre variables e individuos (observaciones) en el espacio Double-STATIS consenso.

2.4.6. (K+1)-STATIS O STATIS EXTERNO

El (K+1)-STATIS, explora la relación de **H** tablas con una tabla externa. Se cuenta con un conjunto de $\mathbf{H}_{I \times J_h}$ tablas, y una tabla externa (la tabla $k + 1$) con información sobre aspectos de los datos de las **H** matrices, esta tabla con información externa es representada por una matriz $\mathbf{K}_{I \times M}$.

El objetivo del método es analizar el patrón de similitud entre las **H** tablas en relación con la tabla ($K + 1$). La idea principal es calcular los pesos \mathbf{v} óptimos utilizando matrices de correlación o covarianza entre cada una de las tablas **H** y la tabla ($K + 1$), en lugar de las matrices de productos cruzados. Estos pesos óptimos se utilizan en lugar de los pesos \mathbf{v} de STATIS, para calcular una solución estándar para STATIS, basada en las **H** matrices. La matriz ($k + 1$) se proyecta como un elemento suplementario en la solución.

Se tienen **H** matrices \mathbf{X}_h , de dimensión $I \times J_h$; y una matriz externa $\mathbf{H}_{I \times M}$, todas ellas con información medida sobre en el mismo conjunto de **I** observaciones, y adecuadamente pre-procesadas.

El primer paso del análisis y dependiendo del tipo de preprocesamiento realizado, es calcular una matriz de correlación o covarianza entre cada una de las matrices \mathbf{X}_h y **K**. Estas matrices, denotadas \mathbf{S}_h y de dimensión M por J_h se calculan como

$$\mathbf{S}_h = \mathbf{K}^T \mathbf{X}_h \quad (55)$$

El primer paso de STATIS consiste en encontrar los pesos óptimos, para ello se utilizan las matrices de productos cruzados calculadas a partir de las matrices \mathbf{S}_h como

$$\mathbf{W}_h^* = \mathbf{S}_h \mathbf{S}_h^T = \mathbf{K}^T \mathbf{X}_h \mathbf{X}_h^T \mathbf{K} \quad (56)$$

Como el término $\mathbf{X}_h \mathbf{X}_h^T$ define el producto cruzado estándar, esta última ecuación puede reescribirse como:

$$\mathbf{W}_h^* = \mathbf{K}^T \mathbf{W}_h \mathbf{K} \quad (57)$$

Esto hace explícitas las relaciones entre (K+1)-STATIS y STATIS estándar.

A partir de las matrices de productos cruzados \mathbf{W}_h , entonces se calcula la matriz de producto interno denotada \mathbf{S}^* cuyo término genérico es,

$$\mathbf{s}_{h,h'}^* = \langle \mathbf{W}_h^*, \mathbf{W}_{h'}^* \rangle = \langle \mathbf{K}^T \mathbf{X}_h \mathbf{X}_h^T \mathbf{K}, \mathbf{K}^T \mathbf{X}_{h'} \mathbf{X}_{h'}^T \mathbf{K} \rangle \quad (58)$$

La descomposición propia de la matriz \mathbf{S}^* dada por,

$$\mathbf{S}^* = \mathbf{V}^* \mathbf{\Lambda}^* \mathbf{V}^{*T} \quad (59)$$

donde

$$\mathbf{V}^{*T} \mathbf{V}^* = \mathbf{I} \quad (60)$$

Del primer vector propio de \mathbf{S}^* , denotado \mathbf{v}_1^* , y se obtiene el conjunto de pesos óptimos para (K+1)-STATIS de la misma forma que para el STATIS,

$$\boldsymbol{\gamma}^* = \mathbf{a}_1^* \mathbf{x} (\mathbf{a}_1^{*T})^{-1} \quad (61)$$

Los elementos de $\boldsymbol{\gamma}^*$ se utilizan luego en los elementos de $\boldsymbol{\gamma}$ y el resto del procedimiento STATIS se aplica con estos pesos.

2.4.7. DISTATIS

La idea principal detrás de DISTATIS es definir y analizar la estructura común de un conjunto de matrices de distancia entre unidades, derivando de esta estructura un conjunto óptimo de pesos. Estos pesos se utilizan para calcular la mejor representación común de las unidades, denominada compromiso. El compromiso se obtiene como la suma ponderada de cada matriz de distancia multiplicada por el peso óptimo, es decir, el compromiso es una combinación lineal de las matrices de datos originales.

DISTATIS es una generalización del EMD clásico (Torgerson, 1958) en el enfoque STATIS (Escoufier, 1985) diseñado para analizar un conjunto de matrices de distancia. El EMD se obtiene a través de la DVS de la matriz de distancia doblemente centrada. El método permite analizar tanto la estructura relacional incrustada en cada capa individual como la estructura relacional global derivada como una combinación lineal (que se asemeja a una función de agregación de capas) de las capas con pesos controlados por datos y proporciona un amplio conjunto de resultados analíticos y gráficos.

El método trabaja de la siguiente forma: primero deriva una matriz de distancias de tres vías; luego aplica STATIS a la matriz de tres vías derivada de matrices

de distancia; interpreta los resultados y representaciones de DISTATIS en términos de datos relacionales.

Sean H estudios, en donde cada uno de ellos produce una matriz de distancias \mathbf{D}_h de $I \times I$, donde I es el número de objetos. Dada, la matriz de distancia de tres vías $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_H)$, el primer paso, de acuerdo con el procedimiento clásico de EMD (Escalamiento Multidimensional), cada matriz \mathbf{D}_h debe ser doblemente centrada para obtener la llamada matriz de productos cruzados $\tilde{\mathbf{R}}_h$. Dado el operador centrado $\Phi = \mathbf{I} - \mathbf{1}\mathbf{1}^T$, donde \mathbf{I} es una matriz de identidad n -dimensional, $\mathbf{1}$ es un vector unidad n -dimensional y n es un vector n -dimensional con elementos iguales a $\frac{1}{I}$, correspondiente a la masa de cada nodo, las matrices de productos cruzados se definen como,

$$\tilde{\mathbf{R}}_h = -\frac{1}{2} \Phi \mathbf{D}_h \Phi^T \quad (62)$$

En un segundo paso, las matrices $\tilde{\mathbf{R}}_h$ se combinan linealmente para obtener el compromiso. En DISTATIS, el compromiso es un promedio ponderado de las matrices de distancia utilizando un sistema doble de pesos. El primer conjunto de pesos tiene como objetivo normalizar cada tabla en términos de variabilidad, mientras que el segundo conjunto de pesos expresa la similitud entre las tablas. Se calcula el primer valor propio para cada $\tilde{\mathbf{R}}_h$ y normalizados con

$$\mathbf{R}_h = \beta_{1h}^{-1} \tilde{\mathbf{R}}_h \quad (63)$$

donde β_{1h} representa el primer valor propio de $\tilde{\mathbf{R}}_h$.

Para calcular el segundo conjunto de pesos, se evalúa la similitud entre tablas, para ello se utiliza el coseno entre las matrices \mathbf{R}_h . A medida que las matrices se normalizan, estos cosenos corresponden al coeficiente RV (Robert & Escoufier, 1976).

Sea la matriz $\mathbf{Q} = (q_{h,h'})$ una matriz de similitud que recoge todos los pares de coeficientes RV, es decir,

$$q_{h,h'} = \frac{\mathbf{r}_h^T \mathbf{r}_{h'}}{\|\mathbf{r}_h\| \|\mathbf{r}_{h'}\|} \quad (64)$$

siendo \mathbf{r}_h la vectorización de \mathbf{R}_h .

Dada la matriz \mathbf{Q} , calculamos sus valores y vectores propios, es decir, $\mathbf{Q} = \mathbf{G}^T \mathbf{\Xi} \mathbf{G}$. El primer vector propio \mathbf{g}_1 se usa para determinar el segundo conjunto de pesos,

$$\gamma_h = \frac{g_{k_1}}{\|g_1\|} \quad (65)$$

siendo g_{k_1} la coordenada de la capa h en el subespacio abarcado por el primer vector propio de \mathbf{Q} . Así, los γ reflejan las similitudes entre las matrices de productos cruzados normalizados. La matriz de compromiso $\mathbf{R} = \sum_{h=1}^H \gamma_h \mathbf{R}_h$ es la suma ponderada de las matrices normalizadas de productos cruzados, la matriz de compromiso sigue siendo una matriz de productos cruzados y, por lo tanto, puede descomponerse a través de la DVS en el marco de un EMD.

Finalmente, siguiendo el EMD clásico, realizamos la descomposición del valor propio de la matriz de compromiso $\mathbf{R} = \mathbf{\Lambda} \mathbf{\Gamma} \mathbf{\Lambda}^T$ para obtener las coordenadas factoriales para trazar los nodos en el espacio común

También es posible representar cada matriz de productos cruzados \mathbf{R}_h en el espacio de compromiso proyectando las matrices como puntos suplementarios. Las coordenadas se pueden calcular fácilmente como y y se llaman puntajes parciales. Representan la posición de cada nodo en cada capa, y todas las coordenadas tienen un espacio de referencia común dado por el compromiso.

En resumen, como indican [Abdi et al. \(2005\)](#), los pasos del DISTATIS son:

- ⊕ Transformar cada matriz de distancia (es decir, cada estudio) en una matriz de producto cruzado entre objetos.
- ⊕ Analizar la estructura de las matrices de productos cruzados.
- ⊕ Derivar un conjunto óptimo de pesos para calcular el compromiso.
- ⊕ Calcular el compromiso como una suma ponderada de las matrices individuales de productos cruzados.
- ⊕ Calcular la descomposición propia de la matriz de compromiso.
- ⊕ Trazar las proyecciones de las observaciones en el espacio de compromiso.
- ⊕ Trazar la trayectoria de las observaciones como puntos suplementarios en el espacio de compromiso.

2.4.8. CANOSTATIS O STATIS-CANONICO/Y STATIS CANONICO DUAL

[Vallejo-Arboleda et al, \(2007\)](#) desarrollan el CANOSTATIS, para situaciones en que los individuos tengan estructura de grupos. Es un el análisis de datos de tres vías, con la característica de que las filas tienen estructura de grupo con varias variables y diferentes momentos, escenarios o tiempos.

El CANOSTATIS se utiliza en casos en donde los grupos de individuos son los mismos y las variables pueda que sean ser diferentes. El objetivo es representar en un espacio de baja dimensión, el cambio de las distancias de Mahalanobis entre las medias de los grupos de cada estudio (tiempo o escenario).

Se parte de observaciones organizadas en H tablas de datos $\mathbf{X}_1, \dots, \mathbf{X}_h, \dots, \mathbf{X}_H$, donde cada \mathbf{X}_h es de tamaño $I \times J_h$ de I individuos divididos en G grupos por J_h variables continuas, para $h = 1, \dots, H$. Primeramente, se centran (o estandarizan) los datos por columnas y se construyen las matrices de medias de los grupos, que denotamos por $\bar{\mathbf{X}}_1, \dots, \bar{\mathbf{X}}_h, \dots, \bar{\mathbf{X}}_H$ y las matrices de covarianzas dentro de los grupos en cada una de las tablas y que se denotan como $\mathbf{S}_1, \dots, \mathbf{S}_h, \dots, \mathbf{S}_H$, que se asumen como definidas positivas. Se construye la matriz \mathbf{D} , una matriz diagonal con los tamaños de los grupos en la diagonal. Los individuos pueden estar estructurados, por ejemplo, de acuerdo a un tratamiento donde para cada tratamiento los individuos están replicados

Las matrices de medias de las variables de cada grupo, componen los estudios. La matriz de la métrica de las variables es la diagonal con los tamaños de los grupos y la matriz de la métrica de los grupos es la inversa de la matriz de covarianzas dentro de los grupos de individuos. Al seleccionar estas métricas y dado que las medias están centradas, los objetos representativos de los estudios, son matrices que contienen las distancias de Mahalanobis entre las medias de los grupos.

Para el CANOSTATIS los estudios son las ternas $(\bar{\mathbf{X}}_h, \mathbf{S}_h^{-1}, \mathbf{D})$, donde cada matriz $\bar{\mathbf{X}}_h$ es una matriz que contiene en sus filas los vectores de medias de los grupos, para las J variables, \mathbf{S}_h es la matriz de covarianzas dentro de los G grupos en la tabla h y es de tamaño $J \times J$; y \mathbf{D} una matriz diagonal que contiene los tamaños de los grupos,

$$\mathbf{S}_h = \frac{1}{I - G} (\mathbf{X}_h^T \mathbf{X}_h - \bar{\mathbf{X}}_h^T \mathbf{D} \bar{\mathbf{X}}_h) \quad (66)$$

El uso de esta matriz, como métrica de los grupos en cada ocasión, permite capturar en el análisis la variabilidad presente dentro de los grupos.

Como se utiliza \mathbf{S}_h^{-1} como matriz de métrica de los grupos, los objetos representativos de cada estudio son productos escalares generalizados. Igualmente, si los datos están centrados, estos productos escalares están

relacionados con las distancias de Mahalanobis entre las medias de los grupos en cada ocasión.

Las matrices \mathbf{W}_h , representan los productos escalares que generan las distancias de Mahalanobis entre las medias de los grupos; es decir, son los objetos representativos del estudio h en el CANOSTATIS, está dada por

$$\mathbf{W}_h = \bar{\mathbf{X}}_h \mathbf{S}_h^{-1} \mathbf{X}_h^T \quad (67)$$

Cada objeto \mathbf{W}_h es una matriz de productos escalares entre los vectores fila de $\bar{\mathbf{X}}_h$, pero con la métrica de Mahalanobis, denotada por \mathbf{S}_h^{-1} .

El producto escalar de Hilbert-Schmidt entre dos objetos \mathbf{W}_h y $\mathbf{W}_{h'}$ para este caso particular se expresa como:

$$\langle \mathbf{W}_h | \mathbf{W}_{h'} \rangle_{HS} = \text{traza} (\mathbf{D} \bar{\mathbf{X}}_h \mathbf{S}_h^{-1} \bar{\mathbf{X}}_h^T \mathbf{D} \bar{\mathbf{X}}_{h'} \mathbf{S}_{h'}^{-1} \bar{\mathbf{X}}_{h'}^T) \quad (68)$$

La norma del objeto \mathbf{W}_h está dad por:

$$\|\mathbf{W}_h\|_{HS}^2 = \langle \mathbf{W}_h | \mathbf{W}_h \rangle_{HS} \quad (69)$$

La distancia entre dos objetos \mathbf{W}_h y $\mathbf{W}_{h'}$ está dada por:

$$d_{HS} = \|\mathbf{W}_h - \mathbf{W}_{h'}\|_{HS} = \sqrt{\|\mathbf{W}_h\|_{HS}^2 - \|\mathbf{W}_{h'}\|_{HS}^2 + 2\langle \mathbf{W}_h | \mathbf{W}_{h'} \rangle_{HS}} \quad (70)$$

Sea \mathbf{S}^W la matriz de los productos escalares RV y de tamaño $H \times H$. La descomposición espectral de \mathbf{S}^W permite obtener la imagen Euclídea de cada uno de los objetos, a esta estructura se denomina inter-estructura, así,

$$\mathbf{S}^W = \mathbf{L} \mathbf{\Gamma} \mathbf{L}^T \quad (71)$$

con \mathbf{L} una matriz cuyas columnas son los vectores propios de \mathbf{S}^W y $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_h, \dots, \gamma_H)$ los valores propios ordenados.

Si $\mathbf{E}_1, \dots, \mathbf{E}_h, \dots, \mathbf{E}_H$ son los puntos asociados a los objetos $\mathbf{W}_1, \dots, \mathbf{W}_h, \dots, \mathbf{W}_H$, en la imagen Euclídea obtenida, las coordenadas de \mathbf{E}_h son las h filas de la matriz,

$$\mathbf{E} = \mathbf{L} \mathbf{\Gamma}^{\frac{1}{2}}.$$

La representación de los estudios en el plano principal permite visualizar la inter-estructura y se la proximidad entre dos puntos convenientemente representados en este plano, se interpretan como la existencia de una estructura de grupos comunes a las correspondientes tablas o estudios.

Es decir, se aplica la metodología del STATIS a los datos y matrices de medias pero con métricas específicas, con la finalidad de explorar la existencia o no de una estructura común, de las distancias de Mahalanobis entre las medias de los grupos. Comprobada la existencia de dicha estructura común se busca una

representación común de las medias y de las variables promedio de todos estudios en un único espacio. De no existir estructura común, no debe continuarse con el análisis, ya que la representación puede llevar a errores y principalmente si se trata de representar sobre esta misma estructura las trayectorias de las medias o de las variables de todas las ocasiones o estudios. La norma al cuadrado de cada objeto representativo se puede expresar como la suma de los valores propios al cuadrado de la matriz, $\mathbf{S}_h^{-1}\mathbf{K}_h$ la cual es la que se diagonaliza en el Análisis Canónico de Poblaciones, siendo \mathbf{K} :

$$\mathbf{K}_h = \frac{1}{(G-1)} \bar{\mathbf{X}}_h^T \mathbf{D} \bar{\mathbf{X}}_h \quad (72)$$

una matriz de covarianzas entre los grupos de las h tabla de tamaño y de tamaño $J_k \times J_h$.

El CANOSTATIS es una alternativa metodológica, que permite trabajar, de forma similar a como lo hace el Análisis Canónico de Poblaciones, pero representando la estructura en un espacio de referencia común a todas las ocasiones, configurado a través de un análisis STATIS específicamente diseñado para capturar la estructura estable en el tiempo. La proyección sobre este espacio de referencia común de las matrices originales nos permite visualizar, además, aquello que ha sido diferente a lo largo del tiempo. Por tanto, el CANOSTATIS permite obtener un sistema de referencia común para representar los Análisis Canónicos de Poblaciones realizados en cada una de las ocasiones. Tanto en el CANOSTATIS como en el STATIS DUAL CANÓNICO, se obtiene un espacio de representación común donde los ejes están en las orientaciones que hacen máximo el cociente (variabilidad entre / variabilidad dentro).

Con el STATIS CANÓNICO podemos representar, por un lado, la media compromiso, y por otro, la evolución de las medias de los tratamientos en el tiempo.

Con el STATIS CANÓNICO DUAL representamos, por un lado, las variables compromiso y sus trayectorias y por otro las variables compromiso y las medias de esas variables en las diferentes ocasiones. En ésta última representación es posible, incluso, proyectar cada una de las réplicas de los estudios originales.

Tenemos entonces dos aspectos importantes en el análisis de este tipo de datos: La estructura de los grupos en cada ocasión y la evolución de dicha estructura de grupos. También sería posible analizar si la estructura de covariación

existente entre las variables consideradas en el estudio se mantiene a lo largo de los tiempos/ocasiones de estudio, con el STATIS DUAL CANONICO. Obsérvese que si aplicamos a datos con esta estructura el STATIS o el STATIS DUAL clásicos los resultados serían diferentes ya que no se tendría en cuenta la variabilidad entre tratamientos.

En el análisis de la inter-estructura en el CANOSTATIS, se aplica el Análisis de Componentes Principales a la matriz \mathbf{S}^W y se obtiene la representación gráfica en el primer plano principal que proporciona la representación euclídea de la inter-estructura, donde, como es usual, se espera que el primer eje recoja la mayor parte de la información

La representación euclídea nos permite evaluar la absorción de inercia de los ejes. Si el eje 1 absorbe la mayor parte de la información se puede pensar que en el análisis de las diferentes ocasiones/tiempos, hay una parte común que está bien recogida por el primer eje de la inter-estructura, y por tanto tendrá sentido la interpretación de una estructura compromiso que nos permita representar geoméricamente las diferencias entre los tratamientos en un tiempo compromiso. Sin embargo, si el eje 2 recoge información de interés, eso significaría que las correlaciones con el eje 2 ponen de manifiesto que, a pesar de existir una parte de la información que es común a los diferentes tiempos, hay otra parte específica que marca diferencias.

[VALLEJO-ARBOLEDA et al, \(2007\)](#) demuestran que es posible no solo conocer la calidad de representación en el primer plano principal y de los diferentes tiempos, presentando el BILOT inducido por el CANOSTATIS, sino también la calidad de representación de los diferentes tratamientos (grupos compromiso) y de las variables. Al igual que en el STATIS clásico, es posible representar, en el BILOT STATIS CANÓNICO, las trayectorias de las variables. Los autores señalan que la representación BILOT de las técnicas, CANOSTATIS y CANOSTATIS DUAL, proporciona ventajas relevantes frente a la forma clásica de presentar los Métodos STATIS clásicos. La [Figura 31](#) muestra un esquema de los pasos del CANOSTATIS.

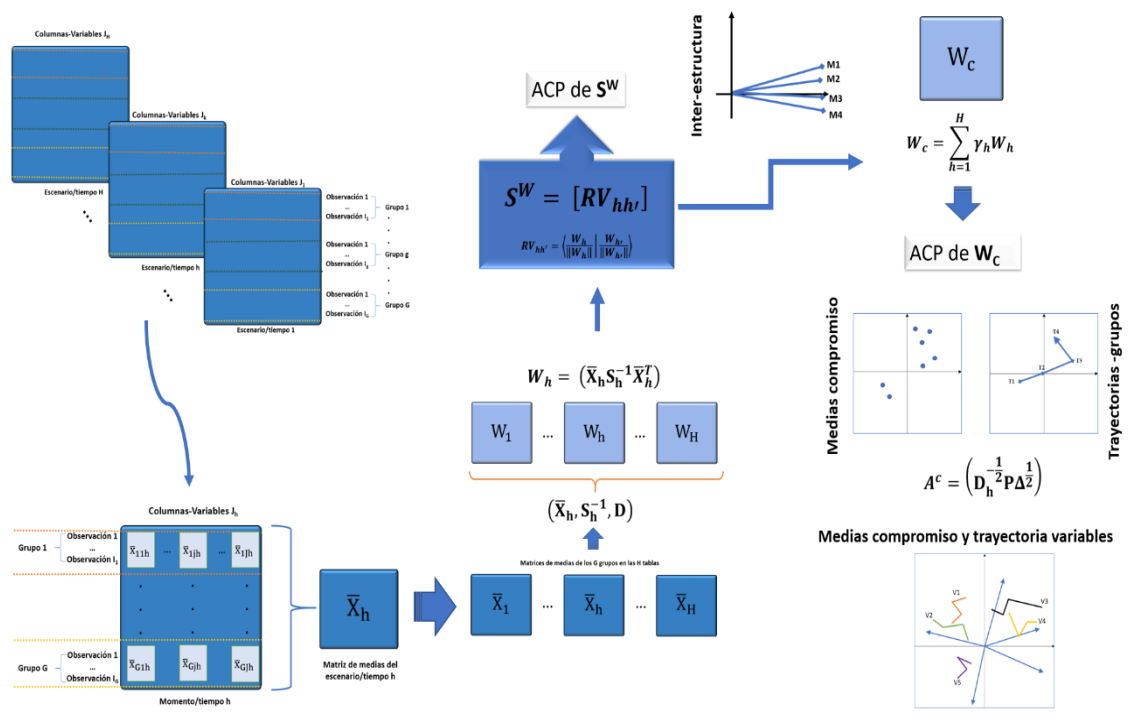


Figura 31. Esquema de la técnica CANOSTATIS

2.4.9. STATIS-4

En 2008, SABATIER y VIVEN, extienden el STATIS y proponen el STATIS- 4 que es presentado por sus autores como un nuevo método lineal para analizar tablas multibloque de 4 vías; es decir, más de dos conjuntos de tablas que tengan las filas comunes.

Como en el DO-ACT, a partir de las tablas se definen los operadores W (matrices de productos escalares entre los individuos de cada tabla), en un primer paso se busca una matriz compromiso para cada bloque y posteriormente una matriz compromiso de las matrices compromiso del paso previo. El estudio de la inter-estructura se resuelve algorítmicamente y no como un proceso de diagonalización de matrices. Como en el STATIS, todas las optimizaciones se hacen en el sentido de Hilbert-Schmidt y se usan productos escalares entre operadores a partir de la traza y los coeficientes de correlación RV.

Al ser una generalización del método STATIS para más de una tabla multibloque a través de la búsqueda de operadores compromiso $L+1$, tal que: los L operadores compromiso $\mathbf{W}_\ell^c \mathbf{D}$ son representativos del K_ℓ operador $\mathbf{W}_k^\ell \mathbf{D}$ de la ℓ -ésima tabla multibloque; y el compromiso global $\mathbf{W}^c \mathbf{D}$ es el mejor operador que representa los L compromisos anteriores.

De la misma forma que en el STATIS, todas las optimizaciones se realizan en el sentido del criterio de Hilbert-Schmidt, es decir, utilizar el producto escalar entre el operador utilizando la traza y el coeficiente RV como correlación.

Entonces, $\mathbf{W}_\ell^c \mathbf{D}$ contiene productos escalares entre n observaciones del compromiso del ℓ -ésimo multibloque y el cual es el más cercano a todos los operadores de $\mathbf{W}_k^\ell \mathbf{D}$. Del mismo modo, el compromiso global contiene productos escalares entre observaciones de compromiso que son los más cercanos a los de $\mathbf{W}_\ell^c \mathbf{D}$. Por lo tanto, los compromisos multibloques permiten resumir las similitudes entre las observaciones (líneas) en cada multiblock y el compromiso global permite resumir las similitudes entre las observaciones de cada multiblock.

Cada compromiso $\mathbf{W}_k^\ell \mathbf{D}$ ($\ell = 1, \dots, L$) y el compromiso global $\mathbf{W}^c \mathbf{D}$, son combinaciones lineales, con coeficientes desconocidos, de sus respectivos operadores:

$$\mathbf{W}_\ell^c \mathbf{D} = \sum_{k=1}^{K_\ell} \gamma_k^\ell \mathbf{W}_k^\ell \mathbf{D} \quad (73)$$

$$\mathbf{W}^c \mathbf{D} = \sum_{\ell=1}^L \mu_\ell \mathbf{W}_\ell^c \mathbf{D} \quad (74)$$

que maximiza la norma al cuadrado: $\text{traza}(\mathbf{W}^c \mathbf{D} \mathbf{W}^c \mathbf{D})$.

Esta optimización significa que se busca el compromiso global que tiene la norma máxima. La configuración de las observaciones asociadas tendrá la máxima "varianza".

El STATIS-4, utiliza un proceso iterativo para calcular un compromiso para cada conjunto de tablas y un compromiso global (es decir, el compromiso de los compromisos).

2.4.10. KERNEL-STATIS

Kernel-STATIS (K-S), es una propuesta de [Marcondes Filho et al, \(2011\)](#) la cual consiste en el desarrollo de gráficos de control (GC) basado en una modificación del STATIS, que también incorpora no linealidades presentes en los datos. Mediante el uso de kernels, se definen funciones no lineales de los datos para una mejor representación de la estructura a caracterizar por el método STATIS. Es una generalización del método STATIS que permite obtener una medida de correlación no lineal entre matrices \mathbf{X}_b . Esto se realiza usando la teoría del kernel ([Schölkopf, Smola, & Bach, 2002](#)), sumergiendo las matrices \mathbf{X}_b en un espacio extendido (denominado espacio característico) que contiene funciones no lineales $\Phi(x)_t^b$ de las observaciones originales $(x)_t^b$ de \mathbf{X}_b . Tal espacio puede por ejemplo, contener productos de segundo orden de los elementos de $(x)_t^b$, que en ese caso, se puede evaluar la correlación no lineal (cuadrática) entre matrices \mathbf{X}_b . Este objetivo se logra mediante el uso de una función de kernel adecuada a partir de observaciones $(x)_t^b$.

El método considera dos vectores de observación $(x)_t^b$ y $(x)_{t'}^b$, y una función $k(x_t^b, x_{t'}^b)$, llamado kernel, que genera un número real que representa la similitud entre los vectores, tal que $k(x_t^b, x_{t'}^b) = k(x_{t'}^b, x_t^b)$. Las medidas de similitud comúnmente utilizadas en el análisis de datos multivariados utilizan el producto interno canónico, definido a través de del kernel $k(x_t^b, x_{t'}^b) = x_{t'}^b \cdot x_t^b$. Esta medida permite trabajar con construcciones geométricas formuladas en términos de ángulos, longitudes y distancias.

Los autores señalan que las técnicas estadísticas lineales de análisis multivariado, como el PCA, el análisis discriminante y el método STATIS, son algoritmos que se pueden escribir de acuerdo con productos internos canónicos entre observaciones. Las versiones no lineales de estos algoritmos se pueden obtener a través de medidas de similitud más generales, representadas por productos internos modificados de los datos a través de kernels. Este resultado se conoce como truco del kernel y se detalla en ([Schölkopf et al., 2002](#)).

2.4.11. COVSTATIS

COVSTATIS se utiliza para analizar un conjunto de matrices de covarianza (o correlación). Cuando se trata de un conjunto de matrices de covarianza, el problema es similar al enfoque de productos cruzados para STATIS (tenemos matrices de covarianza en lugar de productos cruzados). Para el caso de matrices de covarianza, el problema de tener diferentes unidades para las diferentes matrices de covarianza debe eliminarse mediante algún esquema de normalización (por ejemplo, asegurando que todas las matrices de covarianza tengan una norma de 1, o tengan un primer valor propio igual a 1).

Sin embargo, existe un problema adicional, porque cada matriz de covarianza (o correlación) puede interpretarse como un espacio en el que los puntos están dispersos desde su origen, pero cada matriz tiene su propio origen. Las matrices de covarianza con orígenes diferentes no se pueden comparar de manera significativa. Para garantizar que todas las matrices de covarianza tengan un origen común, cada matriz debe estar doblemente centrada (esto es equivalente a restar la media para variables simples). Esto se obtiene calculando primero una matriz de centrado I por I, denotada por Φ , y definida como,

$$\Phi = [I - \mathbf{1m}^T] \quad (75)$$

Entonces, si R_h es una covarianza (normalizada) o matriz de correlación, la matriz de producto cruzado de doble centrado correspondiente (W_h) se obtendrá como,

$$W_h = \frac{1}{2} \Phi R_h \Phi \quad (76)$$

Estas matrices W_h se utilizan como en el enfoque estándar de productos cruzados para el procedimiento STATIS. Con COVSTATIS, las distancias en el esquema entre las observaciones son, aproximadamente, inversamente proporcional a su covarianza o correlación en las matrices de datos originales.

2.4.12. COSTATIS

La técnica COSTATIS (STATIS y CO-INERCIA) (Thioulouse, 2011), implica también organizar los datos en tres entradas las cuales corresponden a individuos, variables y escenarios. Es decir, la información se registra en tablas

(o matrices) que identifican a las observaciones (individuos) bajo estudio en filas, a las variables que se han medido sobre estos en columnas, y los tiempos o momentos de estudio en cada matriz por separado. La finalidad es analizar las diferencias y semejanzas entre los diferentes escenarios a través de las configuraciones de las observaciones y de las relaciones entre los diferentes grupos de variables.

Esta técnica, al igual que BGOIA (Franquet, Dolédec, & Chessel, 1995), STATICO (Simier et al., 1999) se utilizan para analizar en forma simultánea un par de cubos de datos (dos secuencias de tablas) y estudiar estabilidad de las relaciones entre pares de tablas en un período de tiempo o espacio. COSTATIS hace énfasis en las relaciones (co-estructuras) entre las estructuras estables extraídas de las dos series de tablas. La técnica se basa a su vez en dos técnicas conocidas ampliamente, las cuales son el Análisis de Co-inercia (COIA) (Doledec & Chessel, 1994) y el Análisis Triádico Parcial (PTA) (Thioulouse & Chessel, 1987). En la Figura 32 se muestran los pasos del análisis COSTATIS.

En el COSTATIS, el número de observaciones (individuos) y columnas (variables) de los dos cubos deben ser los mismos para todas las tablas. El primer paso de COSTATIS consiste en realizar dos PTA, uno sobre las H-tablas del primer cubo, y otro sobre las H-tablas del segundo cubo. El segundo paso es un COIA de los compromisos de estos dos PTA. El análisis de Co-inercia de los dos compromisos, descompone la Co-inercia total y maximiza la Co-inercia entre los dos grupos de variables. Cada compromiso representa la "estructura estable" de la serie correspondiente. COSTATIS pone de manifiesto las relaciones entre estas dos estructuras estables. Por lo tanto, es muy fácil de interpretar (como un análisis estándar de Co-inercia), pero conserva las propiedades de optimalidad de los compromisos de los dos análisis triádicos parciales. COSTATIS se desarrolló inicialmente para resolver problemas en el contexto ecológico, específicamente lo relacionado a variables ambientales y especies.

Sean X_1, \dots, X_H , H tablas de variables ambientales con las mismas I files (individuos, muestras, observaciones) y las mismas J variables. Sean Y_1, \dots, Y_H , H tablas de especies con las mismas I' filas y las mismas J' variables (que en este caso corresponden a las especies). Primeramente, se realizan dos PTA

(análisis triádicos parciales): uno para las tablas de variables ambientales y para las tablas de especies.

Sea $\mathbf{X} = \sum_1^H \gamma_h \mathbf{X}_h$ de $I \times J$ el compromiso del primer PTA, es decir el compromiso para las variables ambientales, con $\sum_1^H \gamma_h^2 = 1$. Sea $\mathbf{Y} = \sum_1^H \delta_h \mathbf{Y}_h$ de $I \times J'$, el compromiso del segundo PTA, en este caso de la tabla de especies, con $\sum_1^H \delta_h^2 = 1$. Sean $(\mathbf{X}, \mathbf{D}_J, \mathbf{D})$ y $(\mathbf{X}, \mathbf{D}_{J'}, \mathbf{D})$ las dos tripletas asociadas. En un segundo paso, el análisis de Coinertia combina estos dos PTA en uno solo para identificar qué estructuras son comunes a ambos compromisos. Por lo tanto, se define por la tripleta $(\mathbf{X}^T \mathbf{D} \mathbf{X}, \mathbf{D}_J, \mathbf{D}_{J'})$. El análisis de coinertia de estos dos compromisos descompone la coinertia total:

$$\mathbf{L} \mathbf{Y}^T \mathbf{D} \mathbf{X}, \mathbf{D}_J, \mathbf{D}_{J'} = \text{traza} (\mathbf{Y}^T \mathbf{D} \mathbf{X} \mathbf{D}_J \mathbf{X}^T \mathbf{D} \mathbf{Y} \mathbf{D}_{J'}) \quad (77)$$

y maximiza la coinertia entre especies y los puntajes de las variables ambientales. Se puede implementar un paso adicional, como en el método STATICO, es posible proyectar las filas y las columnas de todas las tablas de las dos secuencias como elementos suplementarios en el espacio multidimensional de este análisis de coinertia.

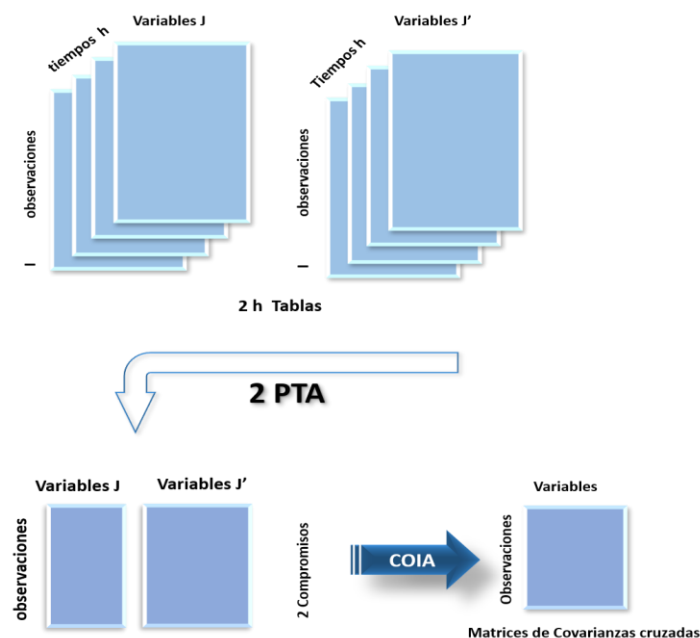


Figura 32. Esquema del procedimiento COSTATIS, adaptado de Thioulouse, 2011.

2.4.13. POWER-STATIS

Power-STATIS (Bénasséni & Bennani-Dosse, 2012) es una versión del STATIS en la cual el peso que se da a cada matriz difiere de la propuesta de L'Hermier en la cual el peso está relacionado con las componentes del primer vector propio de la matriz de correlaciones vectoriales.

Para analizar H tablas $X_1, \dots, X_h, \dots, X_H$, de dimensiones $I \times P_1, \dots, I \times P_H$, procedentes de I individuos, se calculan las matrices de productos cruzados (operadores) $\mathbf{W}_h = \mathbf{X}_h \mathbf{X}_h^H$, donde $h = 1, \dots, H$. Estas matrices \mathbf{W}_h contienen toda la información sobre las estructuras de similitud de los individuos a lo largo de los H tiempos o condiciones.

El cálculo se facilita al utilizar las matrices de productos cruzados en lugar de las matrices de datos, ya que evita la necesidad de rotar al buscar congruencia entre las matrices, sin embargo, su interpretación no es tan sencilla, (Robert & Escoufier, 1976).

Como sabemos, la matriz compromiso $\mathbf{W}_C = \sum_{h=1}^H \gamma_h \mathbf{W}_h$ es una combinación lineal de las \mathbf{W}_h , que se aproxima globalmente a estas matrices de productos cruzados en el sentido de maximizar el criterio,

$$\mathbf{Crit}_2 := \sum_{h=1}^H [\text{traza}(\mathbf{W}_C \mathbf{W}_h)]^2 \quad (78)$$

De acuerdo a Lavit et al. (1994) y a Vivien y Sabatier (2004), los γ_h que maximizan el \mathbf{Crit}_2 son las componentes del primer vector propio de la matriz de correlaciones vectoriales \mathbf{C} , entre los operadores \mathbf{W}_h .

Recordemos que una de las características del método STATIS basado en el criterio de la potencia $S=2$, es que factores para la combinación lineal, de matrices que no se parecen a la matriz compromiso, dan peso muy bajo (Schlich, 1996).

Las trazas ($\text{traza}(\mathbf{W}_C, \mathbf{W}_h)$), son siempre positivas, (Bénasséni & Bennani-Dosse, 2012) por tanto, no es necesario elevar al cuadrado esa traza por lo cual, se define un criterio general

$$\mathbf{Crit}_s := \sum_{h=1}^H [\text{traza}(\mathbf{W}_C \mathbf{W}_h)]^s \quad (79)$$

siendo $s \geq 1$, los autores estudian los alcances de variar la potencia a la que va elevadas esas trazas, con específicamente el caso $s=1$ por la sencillez del método y las propiedades asociadas a la matriz, compromiso.

Para el caso de $s=1$ el problema de maximización se sujeta a:

$$\mathbf{Crit}_1 := \sum_{h=1}^H [\text{traza}(\mathbf{W}_c \mathbf{W}_h)] \quad (80)$$

donde $\gamma_h \geq 0$ para $h = 1, \dots, H$ y $\gamma^T \gamma = 1$ con $\gamma^T = (\gamma_1, \dots, \gamma_H)$

Dadas las condiciones anteriores, tenemos

$$\sum_{h=1}^H [\text{traza}(\mathbf{W}_c, \mathbf{W}_h)] = \sum_{h=1}^H \sum_{\ell=1}^H \gamma_\ell [\text{traza}(\mathbf{W}_\ell, \mathbf{W}_h)] = \mathbf{e}_H^T \mathbf{C} \gamma \quad (81)$$

con \mathbf{e}_H es un vector de unos.

De acuerdo a la desigualdad de Cauchy-Schwarz se conoce que

$$\mathbf{e}_H^T \mathbf{C} \gamma \leq \|\mathbf{C} \mathbf{e}_H\| \|\gamma\| \quad (82)$$

Así, el máximo se obtiene cuando

$$\gamma = \frac{\mathbf{C} \mathbf{e}_h}{\|\mathbf{C} \mathbf{e}_h\|} \quad (83)$$

donde todos los elementos de γ son positivos, (Bénasséni & Bennani-Dosse, 2012).

El criterio \mathbf{Crit}_1 (potencia $s=1$) tiene como ventaja de que solamente se utilizan las sumas de filas de la matriz de correlaciones vectoriales \mathbf{C} y en los casos en que tienen muchas matrices en el estudio, implican un significativo ahorro computacional ya que no implica la descomposición de esa matriz en valores y vectores propios para poder calcular los factores de carga del primer vector propio que, en la propuesta de L'Hermier, era la base del cómputo de la matriz consenso. La interpretación de los pesos es la siguiente:

Una tabla X_h tendrá un gran peso en la matriz consenso si sus columnas están altamente correlacionadas con las columnas de las otras matrices.

\mathbf{Crit}_1 es simplemente una suma ponderada de los cuadrados de las covarianzas entre las columnas de las H tablas.

$$\mathbf{Crit}_1 = \sum_{\ell=1}^H \gamma_\ell \sum_{h=1}^H \sum_{i=1}^I \sum_{j=1}^J [\text{cov}(x_{hi}, x_{\ell j})]^2 \quad (84)$$

Los pesos incluidos en el \mathbf{Crit}_1 admiten interpretar el compromiso simplemente en términos de covariaciones entre columnas de las H tablas de datos.

El cálculo del compromiso requiere solo operaciones elementales poco elevadas en términos computacionales.

Cuando se trabaja con valores de $s > 2$, se habla de hipotéticas ventajas ya que a medida que aumenta s descende la importancia en el consenso de las matrices poco correlacionadas con el resto, pero los autores manifiestan que el trabajo es incipiente y debe ser estudiado con mayor profundidad.

2.4.14. ANISOSTATIS

En el método STATIS se le da el mismo peso, γ_h a cada variable de una tabla. Una manera más formal de expresar esta condición, es decir que STATIS es una técnica isotrópica. Una extensión interesante de STATIS es relajar la condición isotrópica y asignar un peso por variable (cualquiera que sea la tabla) en lugar de tener el mismo peso para todas las variables de una tabla.

Esta extensión de STATIS es anisotrópica y se llama ANISO-STATIS (Abdi et al., 2012). Entonces, para ANISOSTATIS, el objetivo es encontrar valores para el vector $\mathbf{c} = [\gamma_1 \mathbf{1}_1^T, \dots, \gamma_h \mathbf{1}_h^T, \dots, \gamma_H \mathbf{1}_H^T]$ que acercarán el mapa del compromiso al conjunto de tablas. Al generalizar STATIS, los dos criterios equivalentes

$$\mathcal{G} = \|\mathbf{W}_+\|^2 = \langle \mathbf{W}_+, \mathbf{W}_+ \rangle \quad (85)$$

$$\mathbf{W} = \sum_{h=1}^H \langle \mathbf{W}_h, \mathbf{W}_+ \rangle^2 \quad (86)$$

con $\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{1}$, maximizados por STATIS dan lugar a dos soluciones diferentes. Además, también es posible derivar un criterio de potencia 1 análogo al criterio utilizado por STATIS de potencia 1.

A continuación, presentamos estos tres criterios los cuales pueden adaptarse directamente para trabajar con PTA.

2.4.14.1. Criterio 1

El primer criterio maximizado por ANISOSTATIS, equivalente a la ecuación (85) se expresa como:

$$\mathcal{G} = \|\mathbf{W}_+\|^2 = \langle \mathbf{W}_+, \mathbf{W}_+ \rangle = \left\langle \sum_h \mathbf{W}_h \mathbf{A}_h \mathbf{W}_h^T, \sum_h \mathbf{W}_h \mathbf{A}_h \mathbf{W}_h^T \right\rangle \quad (87)$$

con $\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{1}$

La solución de este problema envuelve la matriz de covarianza de \mathbf{X} con todos sus elementos al cuadrado. Esta matriz de covarianza de $J \times J$ se denota $\widehat{\Psi}$ y se calcula como,

$$\widehat{\Psi} = \mathbf{X}^T \mathbf{M} \mathbf{X} \quad (88)$$

Cada entrada de $\widehat{\Psi}$ almacena la covarianza entre dos variables de \mathbf{X} . Esta matriz se usa para calcular una matriz denominada \mathbb{B} , que se llama matriz de covarianza al cuadrado de Hadamard y cuyos elementos son los cuadrados de los elementos de $\widehat{\Psi}$. Específicamente, \mathbb{B} se calcula como

$$\mathbb{B} = \widehat{\Psi} \circ \widehat{\Psi} = (\mathbf{X}^T \mathbf{M} \mathbf{X}) \circ (\mathbf{X}^T \mathbf{M} \mathbf{X}) \quad (89)$$

donde \circ denota el producto Hadamard. Esta matriz \mathbb{B} es semi-definida positiva, ya que $\widehat{\Psi}$ es semi-definida positiva y todos sus elementos, siendo números al cuadrado, son positivos.

Se obtiene la solución del primer criterio, directamente de la descomposición propia de \mathbb{B} que se expresa como,

$$\mathbb{B} = \mathbf{S} \boldsymbol{\Omega} \mathbf{S} \quad (90)$$

con $\mathbf{S}^T \mathbf{S} = \mathbf{I}$

Como lo indica el teorema de Perron-Frobenius, los elementos del primer vector propio de \mathbb{B} se pueden elegir para que sean todos positivos y, por lo tanto, el vector de peso óptimo se obtiene al reescalar este primer vector propio de modo que la suma de sus elementos sea igual a uno. Los demás pasos de ANISOSTATIS se pueden realizar de la misma manera que para la versión estándar de STATIS.

2.4.14.2. Criterio 2

El segundo criterio de ANISOSTATIS generaliza el criterio expresado en la ecuación (86) y se expresa de manera similar a,

$$\mathbf{W} = \sum_{h=1}^H \langle W_h, W_+ \rangle^2 \quad (91)$$

con $\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{1}$

La solución de este problema implica una nueva matriz \mathbf{G} de $J \times H$ en la cual un 1 en la fila j y la columna h indica que la j -estima variable pertenece a la tabla h y un valor de 0 que no pertenece a dicha tabla. La solución de este problema se obtiene calculando primero una matriz semi-definida, $(\mathbb{B}\mathbf{G}) \times (\mathbb{B}\mathbf{G})^T$, que luego se descompone como

$$(\mathbb{B}\mathbf{G}) \times (\mathbb{B}\mathbf{G})^T = \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^T \quad (92)$$

con $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$. Esta descomposición propia también se puede obtener de manera más eficiente a partir de la descomposición de valores singulares de la matriz $\mathbb{B}\mathbf{G}$.

Tal como señala el teorema de Perron-Frobenius, son elegido todos positivos y los elementos de peso óptimo del primer vector propio $(\mathbb{B}\mathbf{G}) \times (\mathbb{B}\mathbf{G})^T$, se obtiene al re-escalar este primer vector propio a una suma de 1. Todos los demás pasos se realizan de la misma forma que en la versión estándar del STATIS.

Al usar la descomposición de los valores singulares de $\mathbb{B}\mathbf{G}$, se pueden graficar los puntajes de los factores de manera similar a la gráfica de la matriz \mathbf{S} para el STATIS estándar, si bien en este caso, dado que se utilizan los valores al cuadrado de una matriz de covarianza, el primer valor singular, es en general, mucho más grande que el segundo por lo cual, es preferible trazar las dimensiones 2 frente a 3 para explorar la estructura de similitud de las tablas.

2.4.14.3. Criterio 3

De manera similar al power-1 STATIS, el segundo criterio puede modificarse para incluir la minimización de la suma de los productos internos de las matrices

de productos cruzados del compromiso, en lugar de los productos internos al cuadrado. Específicamente, el criterio para maximizar es,

$$\mathbf{W}'_1 = \sum_{h=1}^H \langle W_h, W_+ \rangle \quad (93)$$

con $\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{1}$

Al igual que el power-1 STATIS, la versión power-1 de ANISOSTATIS no requiere el cálculo de una descomposición propia. Específicamente, el vector a de pesos óptimos se obtiene como,

$$\boldsymbol{\gamma} = \mathbb{B} \mathbf{1} \times \|\mathbb{B} \mathbf{1}\|^{-1} \quad (94)$$

donde $\mathbf{1}$ es un vector J por 1 de unos. Este vector de pesos óptimos se usa como en el STATIS estándar para calcular puntajes de carga y factores.

Los autores señalan que los pesos óptimos y los puntajes de los factores son casi idénticos para las tres versiones de ANISOSTATIS.

2.4.15. STATIS-LDA

Es un nuevo enfoque y su algoritmo asociado para resolver la tarea de clasificación, cuando los grupos de individuos son los mismos en todas las tablas. Utiliza el análisis discriminante lineal habitual (LDA-discriminante lineal de Fisher) y el enfoque STATIS. El método propuesto, llamado STATIS-LDA, (Sabatier et al., 2013), utiliza la descomposición entre variaciones de cada bloque y permite cuantificar la importancia o peso de cada bloque con respecto a la clasificación.

El método propuesto consiste en utilizar conjuntamente el método STATIS y la tripleta AFD, para cada tabla de datos.

Se denomina STATIS-LDA, el método STATIS aplicado a las H triplete:

$[(\mathbf{G}_h, \mathbf{V}_h^{-1}, \mathbf{D}_1)]$. Los operadores serán $\mathbf{G}_h \mathbf{V}_h^{-1} \mathbf{G}_h^T \mathbf{D}_1$

La inter-estructura consiste en calcular, primero la matriz $\mathbf{C}^* = \{c_{h,h'}^*\}$ al igual que en STATIS estándar.

$$\begin{aligned} c_{h,h'}^* &= \text{traza} (\mathbf{G}_h \mathbf{V}_h^{-1} \mathbf{G}_h^T \mathbf{D}_I \mathbf{x} \mathbf{G}_{h'} \mathbf{V}_{h'}^{-1} \mathbf{G}_{h'}^T \mathbf{D}_I) \\ &= \text{traza} (\mathbf{G}_{h'}^T \mathbf{D}_I \mathbf{G}_h \mathbf{V}_h^{-1} \mathbf{G}_h^T \mathbf{D}_I \mathbf{G}_{h'} \mathbf{V}_{h'}^{-1}) \\ &= \text{traza} (\mathbf{V}_{inter_{h'h}} \mathbf{V}_h^{-1} \mathbf{V}_{inter_{hh'}} \mathbf{V}_{h'}^{-1}) \end{aligned} \quad (95)$$

donde $\mathbf{V}_{inter_{hh'}} = \mathbf{G}_h^T \mathbf{D}_I \mathbf{G}_{h'}$. Luego se diagonaliza la matriz \mathbf{C}^* para obtener la representación de la inter-estructura.

Seguidamente se obtiene el compromiso. Para ello primero se determina el vector de los coeficientes del compromiso, $\mathbf{v}^* = (v_1^*, \dots, v_h^*, \dots, v_H^*)^T$, como el primer vector propio de \mathbf{C}^* . Posteriormente, se calcula $\mathbf{W}_{\mathbf{C}^*} \mathbf{D}_I$ como una combinación lineal de los operadores iniciales:

$$\mathbf{W}_{\mathbf{C}^*} \mathbf{D}_I = \sum_{h=1}^H v_h^* \mathbf{G}_h \mathbf{V}_h^{-1} \mathbf{G}_h^T \mathbf{D}_I \quad (96)$$

Por último, en la intra-estructura, primero se diagonaliza $\mathbf{W}_{\mathbf{C}^*} \mathbf{D}_I$. Se obtienen los ejes de la infraestructura STATIS-LDA como vectores propios de $\mathbf{W}_{\mathbf{C}^*} \mathbf{D}_I$, con \mathbf{D}_I ortogonal, que se notará como $\{c_\alpha^{cdg}\}$; y los valores propios asociados $\{\mu_\alpha\}$, con $\alpha = 1, \dots, A$, donde A es el número de componentes principales retenidas.

STATIS-LDA tiene algunas propiedades importantes: se respeta la descomposición inter e intra; los resultados son independientes de las unidades de medida de las variables, los datos de grandes dimensiones (es decir, que tienen un gran número de observaciones) se pueden procesar; el análisis de los coeficientes v_h^* permite cuantificar la importancia de cada bloque; su estructura permite llevar a cabo procedimientos de validación, que son esenciales en todos los procedimientos de discriminación.

2.4.16. INTERSTATIS

Corrales y Rodríguez, (2014) presentan dos propuestas para aplicar el método STATIS cuando los datos de entrada son de tipo intervalo. Ambas propuestas explotan los estudios adicionales que proporcionan los objetos simbólicos, permitiendo el manejo de variación interna en los datos de entrada. Esto a su vez posibilita el uso de representaciones de datos que mejor reflejen la realidad de lo medido. Estos autores prueban, que el STATIS es un caso particular del propuesto INTERSTATIS con Aritmética de Intervalos (AI) de Moore (Moore, 1979) y ACP de Centros de Cazes (Cazes et al., 1997), validando su sentido estadístico.

2.4.16.1. Interstatis con AI y ACPC

INTERSTATIS con AI y ACP de Centros es la primera propuesta de estos autores del INTERSTATIS para intervalos, se utiliza el ACP de Centros y operaciones de la aritmética de intervalos únicamente. El procedimiento consiste en reemplazar las operaciones convencionales por sus homologas para intervalos, tomadas de la AI, y reemplazar también el ACP clásico por el ACP de Centros, cuyas entradas y salidas son de tipo intervalo. Durante el preprocesamiento de las matrices de entrada, es requisito centrarlas. Para tales efectos, se utilizó la definición de promedio para intervalos de Rodríguez [19], dada como:

$$\bar{Y} = \left(\frac{1}{I} \sum_{i=1}^I \bar{y}_i, \frac{1}{I} \sum_{i=1}^I \bar{y}'_i \right) \quad (97)$$

donde Y es una variable de tipo intervalo para $Y = \{(\bar{y}_1, \bar{y}'_1), \dots, (\bar{y}_i, \bar{y}'_i), \dots, (\bar{y}_I, \bar{y}'_I)\}$.

En INTERSTATIS se tiene los mismos individuos I , pero pueden tenerse diferentes variables que se denotan como J_h , en donde $J = \sum_{h=1}^H J_h$. De igual manera se cuenta con H tablas (o matrices) de datos X_h de tamaño $I \times J_h$, centradas con respecto a D , una matriz diagonal de pesos, invariable para las tablas.

La inter-estructura se calcula la matriz \mathbf{W}_h , dada como,

$$\mathbf{W}_h = \mathbf{X}_h \tilde{\mathbf{x}} \mathbf{X}_h^T \quad (98)$$

Luego se genera la matriz \mathbf{X} , como

$$\mathbf{X} = [\overline{\mathbf{W}}_1 | \dots | \overline{\mathbf{W}}_h | \dots | \overline{\mathbf{W}}_H] \quad (99)$$

Se aplica un análisis de componentes principales de centros para la tripleta $(\mathbf{X}, \mathbf{D}_{\frac{1}{\sigma^2}}, \frac{1}{I^2} \mathbf{L}_I^2)$. Se obtiene la matriz $\mathbf{T} = [\mathbf{CP}_1, \dots, \mathbf{CP}_h, \dots, \mathbf{CP}_H]$, correspondiente a las correlaciones principales, y el vector $u = (u_1, \dots, u_h, \dots, u_H)$, que contiene el primer vector propio y el primer valor propio λ_1 , los cuales se utilizarán para la construcción del compromiso. Por último, se gráfica el círculo de correlaciones a partir de la matriz \mathbf{T} . Cada punto representará una tabla de datos, puntos cercanos significan configuraciones de individuos similares, y vectores paralelos representan configuraciones homotéticas.

La intra-estructura, se efectúa por medio del compromiso $\sum_{h=1}^H \beta_h \mathbf{W}_h$ producto del ACP de Centros: $(\tilde{\mathbf{X}}, \mathbf{L}_\ell, \mathbf{D})$, donde $\beta = (\beta_1, \dots, \beta_H) = \frac{1}{\sqrt{\lambda_1}} u$, y $\tilde{\mathbf{X}} = [\beta_1 \hat{\mathbf{x}} \mathbf{X}_1 | \dots | \beta_H \hat{\mathbf{x}} \mathbf{X}_\ell]$. De esta manera se obtienen las componentes principales $\mathbf{M}_i = [C_1, \dots, C_\ell]$ y las correlaciones principales $\mathbf{E}_v = [\mathbf{CP}_1, \dots, \mathbf{CP}_1]$.

A partir de \mathbf{E}_v , se gráfica el círculo de correlaciones el cual permite estudiar la trayectoria de las variables. Utilizando \mathbf{M}_i , se gráfica el plano principal, en donde se representan los I individuos promedio.

Para estudiar la trayectoria de los individuos, se calcula la matriz \mathbf{T}_{IND} , definida por bloques como:

$$\mathbf{T}_{\text{IND}} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_H \end{bmatrix}$$

Las coordenadas de los individuos para las trayectorias, se calculan como,

$$\mathbf{E}_i = \mathbf{T}_{\text{IND}} \hat{\mathbf{x}} \mathbf{M}_i \quad (100)$$

A partir de \mathbf{E}_i , se gráfica el plano principal y se estudian las trayectorias de los individuos.

2.4.16.2. Interstatis con AA y ACP de Centros

La segunda propuesta es el INTERSTATIS con Aritmética Afín (AA) (Stolfi & De Figueiredo, 2003) y ACP de Centros (Cazes et al., 1997).

Consiste en H tablas de datos \mathbf{X}_h de I individuos y J_h variables, centradas respecto a $\mathbf{D} = \text{diag}(\frac{1}{i}, \dots, \frac{1}{i})$ una matriz de pesos, invariable para las H mediciones.

En la inter-estructura se compara la distribución espacial con las matrices $\mathbf{W}_h = \mathbf{A}_h \tilde{\times} \mathbf{A}_h^T$. Para lo cual es necesario utilizar la métrica,

$$\langle \mathbf{W}_h, \mathbf{W}_{h'} \rangle_{\Phi} = \text{traza}(\mathbf{W}_h \tilde{\times} \mathbf{W}_{h'}) = \langle \vec{\mathbf{W}}_h \tilde{\times} \vec{\mathbf{W}}_{h'} \rangle \quad (101)$$

donde $\vec{\mathbf{W}}_h$ es un vector formado por las filas de la matriz \mathbf{W}_h para $h=1,2,\dots,H$.

El primer paso consiste en transformar las matrices \mathbf{X}_h de intervalos a matrices afines \mathbf{A}_h . Luego se calcula las matrices $\mathbf{W}_h = \mathbf{A}_h \tilde{\times} \mathbf{A}_h^T$ y se genera la matriz $\mathbf{A} = [\vec{\mathbf{W}}_1 | \dots | \vec{\mathbf{W}}_H]$. Se obtiene la matriz \mathbf{X} , que corresponde a la transformación de la matriz \mathbf{A} a una matriz de intervalos.

Obtenida la matriz \mathbf{X} , se aplica un ACP de Centros a la tripleta $(\mathbf{X}, \mathbf{D}_{\frac{1}{\sigma^2}}, \frac{1}{i^2} \mathbf{I}_I^2)$, dando como resultado la matriz de correlaciones principales $\mathbf{T} = [\mathbf{CP}_1, \dots, \mathbf{CP}_h, \dots, \mathbf{CP}_H]$ la cual se utiliza para graficas el circulo de correlaciones, en donde cada punto representará una tabla de datos, puntos cercanos indican configuraciones de individuos similares, y vectores paralelos representan configuraciones homotéticas. En este proceso se obtiene también el primer vector propio, denotado como $u = (u_1, \dots, u_h, \dots, u_H)$, y el primer valor propio λ_1 , los cuales se utilizarán para la construcción del compromiso.

Es importante señalar, como indican los autores, que en esta fase no es necesario convertir los intervalos, ya que la AA define operaciones entre formas afines y reales.

En proceso de intra-estructura, se realiza el estudio del proceso evolutivo de los individuos y las variables a través del compromiso $\sum_{h=1}^H \beta_h \mathbf{W}_h$ producto del ACP de Centros: $(\tilde{\mathbf{A}}, \mathbf{L}_\ell, \mathbf{D})$, donde $\beta = (\beta_1, \dots, \beta_H) = \frac{1}{\sqrt{\lambda_1}} u$, y $\tilde{\mathbf{A}} = [\beta_1 \tilde{\mathbf{X}} \mathbf{A}_1 | \dots | \beta_H \tilde{\mathbf{X}} \mathbf{A}_\ell]$. Obtenida la matriz de bloques $\tilde{\mathbf{A}}$ se convierte a su forma de intervalos, $\tilde{\mathbf{X}}$ se aplica un ACP de Centros a la tripleta $(\tilde{\mathbf{X}}, \mathbf{L}_\ell, \mathbf{D})$, las componentes principales resultantes se asignan a la matriz $\mathbf{M}_i = (C_1, \dots, C_\ell)$ y las correlaciones principales a $\mathbf{E}_v = [CP_1, \dots, CP_I]$. A partir de \mathbf{E}_v se grafica el círculo de correlaciones, que permitirá estudiar la evolución de las variables. Con \mathbf{M}_i se gráfica el plano principal, en donde se representan los I individuos promedio.

Con respecto a la trayectoria de los individuos, se calcula la matriz \mathbf{E}_{IND} , definida por bloques como:

$$\mathbf{E}_{\text{IND}} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_H \end{bmatrix}$$

Las coordenadas de los individuos para las trayectorias, se calculan como,

$$\mathbf{E}_i = \mathbf{E}_{\text{IND}} \tilde{\mathbf{X}} \mathbf{M}_i \quad (102)$$

A partir de \mathbf{E}_i , se gráfica el plano principal y se estudian las trayectorias de los individuos.

2.4.17. SIR-STATIS

Villa-Vialaneix et al, (2016), proponen una extensión del STATIS-dual en el contexto de la regresión inversa partida o Sliced Inverse Regression, SIR (Li, 1991), para analizar conjuntos de datos, de varias tablas, con respecto a una variable numérica de interés.

La regresión inversa partida (SIR) supone la estimación, a partir de una muestra (\mathbf{x}'_i, y_i) , $1 \leq i \leq N$, una relación no paramétrica entre \mathbf{x} e y , donde la dimensión p del vector \mathbf{x} es grande. Lo cual no es posible a menos que N sea muy grande, que en general no es posible. Li (1991), propone el siguiente modelo no

paramétrico, para subsanar el problema, donde y depende de \mathbf{x} sólo a través de un reducido número, K , de combinaciones lineales

$$y = f(\beta_1'x, \dots, \beta_K'x, \varepsilon) \quad (103)$$

aquí y es la variable respuesta, \mathbf{x} es el vector p -dimensional de variables explicativas, ε el error, que es independiente de las \mathbf{x} , los β_i son vectores desconocidos en \mathbb{R}^p y f es una función arbitraria, $f: \mathbb{R}^{K+1} \rightarrow \mathbb{R}$. Las K combinaciones lineales $\beta_1'x, \dots, \beta_K'x$ capturan todo lo que es necesario conocer de las covariables \mathbf{x} para relacionarlas con y . Si K es pequeño y se logra estimar los vectores β_i en forma eficaz, se logra el objetivo de reducir el número de variables y hacer factible una estimación no paramétrica.

El método SIR-STATIS propuesto, está diseñado para analizar el caso en el que, un conjunto \mathbf{X} de datos que corresponde a un conjunto de p variables medidas T veces en los mismos n sujetos, está relacionado con una variable objetivo real \mathbf{y} , medida en los mismos n sujetos. El enfoque es un método exploratorio que tiene como objetivo comprender la evolución de la relación entre \mathbf{X} e \mathbf{y} a través del tiempo y consta de dos fases:

1. Un análisis entre estructuras, en la que se estudia la semejanza entre los diferentes momentos en el tiempo, calculando similitudes entre las estimaciones de la covarianza de la media de $\mathbf{X}_{..t}$ condicional a \mathbf{y} . De manera similar a SIR, la esperanza condicional se estima particionando el rango de \mathbf{y} . El resultado de este análisis es una matriz de covarianza de compromiso Γ_c , que captura una estructura de correlación de compromiso de $\mathbb{E}(\mathbf{X}_{..t} | \mathbf{y})$ sobre t ,
2. Un análisis de la intra-estructura, que es un ACP generalizado del compromiso. Este segundo paso da como resultado salidas gráficas que pueden usarse para explorar la estructura de covarianza entre variables y momentos en el tiempo condicionados a \mathbf{y} .

La [Figura 33](#) muestra un esquema de los pasos del SIR- STATIS.

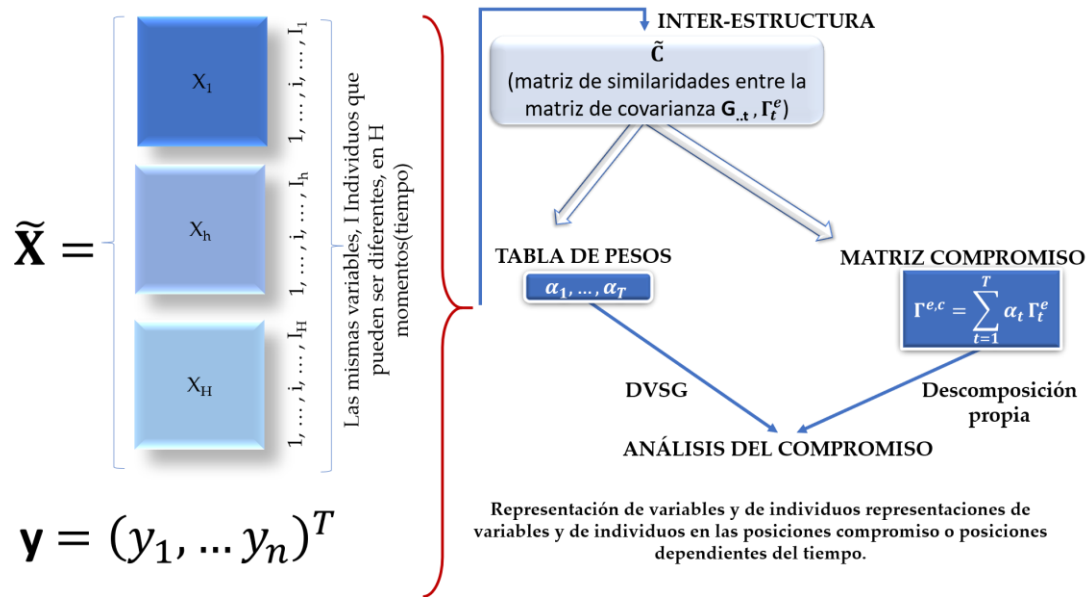


Figura 33. Diagrama del método SIR-STATIS (adaptado de Villa-Vialaneix, et al , 2016)

El proceso es el siguiente, para $n > p$, se obtienen las siguientes estimaciones,

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \tag{104}$$

y

$$\mathbf{\Gamma} = \frac{1}{n} \mathbf{X}^T (\mathbf{X} - \bar{\mathbf{X}}) \tag{105}$$

particionar el rango de Y en H diferentes partes: τ_1, \dots, τ_H y estimar

$$\mathbf{G} = \left(\frac{1}{n_h} \sum_{i: y_i \in \tau_h} \mathbf{z}_i \right) \tag{106}$$

con $n_h = |\{i : y_i \in \tau_h\}|$ y

$$\mathbf{\Gamma}^e = \mathbf{G}^T \mathbf{M} \mathbf{G} \tag{107}$$

con $\mathbf{M} = \text{Diag} \left(\frac{n_1}{n}, \dots, \frac{n_H}{n} \right)$

Seguidamente se obtiene la descomposición en valores propios generalizada de $\mathbf{\Gamma}^e$ con norma $\mathbf{\Gamma}$.

La idea básica es realizar un análisis STATIS dual a los centros de gravedad de las particiones ($G_{..t}$), en lugar de las variables originales. El análisis de compromiso es similar a encontrar un espacio de efectiva reducción de la dimensión (EDR) del compromiso.

En resumen y según sus autores, este método es capaz de enfocarse en una variable numérica de interés, en forma similar a SIR, de igual manera es capaz de explicar evoluciones longitudinales cuando el número de momentos en el tiempo es pequeño. Solo es válido para $n \geq p$, aunque indican que se está estudiando el enfoque de regularización para permitir el análisis de $n < p$ casos. Han desarrollado un paquete en R denominado SirSTATIS (Villa-Vialaneix, 2016).

2.4.18. HIDISTATIS

Kriegsman (2018a), señala que ningún método actual puede calcular compromisos de grupo óptimos. Por un lado el método **HMFA** (Le Dien & Pagès, 2003), calcula compromisos de grupo, y por otro, el **DiSTATIS** (Abdi et al., 2005), calcula un compromiso óptimo único. Para solucionar este problema, el autor combina los procedimientos de **HMFA** y **DiSTATIS** para desarrollar un nuevo método llamado **HiDiSTATIS**, un método discriminante que representa de manera óptima el diseño anidado de un factor en las tablas de arreglo de matrices de distancia ($I \times I \times C$ (D)).

La notación utilizada para denotar los estímulos y los participantes dependen del diseño experimental.

Los datos de clasificación sin procesar se almacenan inicialmente como una matriz de datos nominal, con los estímulos en las filas y los participantes en las columnas. Cada columna contiene etiquetas arbitrarias para denotar cómo un participante determinado clasificó los estímulos. Cada patrón de clasificación (columna) se recodifica como una matriz de distancia de clasificación. Es decir, la matriz (estímulos \times estímulos) en una matriz (estímulos \times estímulos \times estímulos) de matrices de distancia de clasificación. Una matriz de distancia de

clasificación es una matriz cuadrada simétrica de las relaciones por pares entre todos los estímulos. A un par de estímulos ordenados juntos se les asigna un valor de 0, mientras que a los estímulos ordenados en diferentes categorías se les asigna un valor de 1. Estos datos 0-1, llamados distancia de clasificación, definen una distancia euclidiana al cuadrado.

Sin en un diseño experimental, los estímulos se denotan como I , y los participantes se denotan como K , el arreglo ($I \times I \times K$) se denota como $S^{[K]}$. Sin embargo, con un diseño experimental $A(B)$, los estímulos se anidan en categorías B , y los participantes $C(D)$ se anidan en grupos D , para obtener el arreglo ($A(B) \times A(B) \times C(D)$) se denota como, $S_B^{[C(D)]}$. Esta notación apunta a la asimetría entre estímulos y los participantes. La naturaleza de los datos es que los participantes $C(D)$ clasifican los estímulos $A(B)$. Las diversas perspectivas incluyen: todos los participantes, $C(D)$; todos los participantes en un grupo dado, $C(d)$; todos los grupos, $\cdot(D)$; un grupo dado, $\cdot(d)$; o una sola perspectiva general, indicada $\cdot(\cdot)$

En este sentido, el método DiSTATIS jerárquico (**HiDiSTATIS**) soluciona el problema de los métodos actuales que no pueden mapear matrices de distancia en un espacio común que represente de manera óptima la diferencia entre los grupos.

El análisis jerárquico de factores múltiples (**HMFA**), extiende el procedimiento de MFA para acomodar tablas anidadas en grupos (de acuerdo con un diseño anidado de un factor en las tablas). Con reminiscencias de ANOVA, en HMFA, las tablas se normalizan y promedian dentro de los grupos para dar tablas representativas a nivel de grupo, llamadas compromisos de grupo. Luego, los compromisos grupales se normalizan y promedian para dar un gran compromiso. El gran compromiso se descompone entonces para representar la perspectiva general de los estímulos.

El método **HiDiSTATIS** consta de 2 pasos:

Paso 1: es similar a análisis jerárquico de factores múltiples (HMFA), el cual consiste en que las tablas individuales se integren dentro de los grupos para dar compromisos grupales. Además, adoptando el DiSTATIS, cada vez que se

calcula un compromiso, se utiliza un análisis de RV para definir pesos gamma para las tablas correspondientes. De esta manera, cada compromiso de grupo, representa de manera óptima, la perspectiva general de ese grupo, y el gran compromiso representa de manera óptima las perspectivas de los grupos.

En este paso, se calculan los compromisos de grupo y grandes compromisos. Esto es equivalente al primer paso de un HMFA, intercalado con el procedimiento para calcular pesos gamma cada vez que las tablas se integran en un compromiso.

Paso 2: se realiza la descomposición propia del gran compromiso para mapear el espacio factorial del gran compromiso. Los compromisos grupales sobreponderados luego se proyectan en el espacio del factor de gran compromiso para revelar cómo las perspectivas de los estímulos de los grupos difieren del gran compromiso. Del mismo modo, las tablas individuales sobreponderadas se pueden proyectar en el espacio del gran factor de compromiso para revelar cómo las percepciones de los estímulos varían dentro de los grupos. Como indicamos el método HiDiSTATIS analiza un arreglo ($I \times I \times C(D)$) de matrices de distancia. Cada matriz de distancia está doblemente centrada, para dar una matriz de similitud. Este paso asegura que la información dentro de cada matriz se centre en el origen. Luego, cada matriz $\tilde{\mathbf{S}}^{c(d)}$ se normaliza con MFA, de modo que cada matriz aporta una cantidad comparable de información.

$$\mathbf{S}^{c(d)} = \left(\tilde{\lambda}_1^{c(d)} \right)^{-1} \times \tilde{\mathbf{S}}^{c(d)} \quad (108)$$

Se realiza un análisis de RV en las tablas $C(d)$ dentro de cada grupo para proporcionar pesos gamma para esas tablas,

$$\mathbf{R}^{C(d)} = [\text{RV}\{\mathbf{S}^{c(d)}, \mathbf{S}^{c'(d)}\}] \quad (109)$$

Cada compromiso de grupo se calcula como el promedio ponderado de las tablas dentro de ese grupo, para proporcionar los compromisos de grupos D . Estos compromisos grupales se normalizan con MFA, se pasan a un análisis de RV para identificar los pesos gamma entre grupos, y su promedio ponderado proporciona el gran compromiso.

$$\text{eigen}\{\mathbf{R}^{C(d)}\} = \mathbf{V}^{C(d)} \mathbf{\Gamma}^{C(d)} (\mathbf{V}^{C(d)})^T \quad (110)$$

donde $(\mathbf{V}^{C(d)})^T \mathbf{V}^{C(d)} = \mathbf{I}$ y los correspondientes pesos gamma, se calculan como el primer valor propio escalado,

$$\boldsymbol{\gamma}^{C(d)} = \mathbf{V}^{C(d)} \times \left(\mathbf{1}^T \mathbf{v}_1^{C(d)} \right)^{-1} \quad (111)$$

en donde $\mathbf{v}_1^{C(d)}$ denota la primera columna de la matriz de vectores propios $\mathbf{V}^{C(d)}$, y $\mathbf{1}^T \mathbf{v}_1^{C(d)}$ es la suma de todas las columnas.

Luego se calcula el gran compromiso, para lo cual cada conjunto de ponderaciones gamma dentro del grupo, $\boldsymbol{\gamma}^{C(d)}$, se usa para calcular el compromiso de grupo correspondiente, como,

$$\tilde{\mathbf{S}}^{\cdot(d)} = \sum_{c(d)} \boldsymbol{\gamma}^{c(d)} \times \mathbf{S}^{c(d)} \quad (112)$$

que corresponde a la suma de ponderaciones gamma sobre las tablas normalizadas con MFA dentro de ese grupo. De esta manera, un compromiso grupal representa de manera óptima la perspectiva de su grupo.

Este procedimiento se repite a nivel de grupo. Los compromisos del grupo, $\tilde{\mathbf{S}}^{\cdot(D)}$, se normalizan como,

$$\mathbf{S}^{\cdot(d)} = \tilde{\gamma}_1^{\cdot(d)-1} \times \tilde{\mathbf{S}}^{\cdot(d)} \quad (113)$$

para que cada compromiso grupal proporcione una cantidad comparable de información.

A los compromisos grupales $\mathbf{S}^{\cdot(D)}$ luego se les aplica un análisis RV entre grupos para identificar las ponderaciones gamma a nivel de grupo, $\boldsymbol{\gamma}^{\cdot(d)}$. La matriz RV entre grupos, se calcula como,

$$\mathbf{R}^{\cdot(D)} = [RV\{\mathbf{S}^{\cdot(d)}, \mathbf{S}^{\cdot(d')}\}] \quad (114)$$

la descomposición propia,

$$eigen\{\mathbf{R}^{\cdot(D)}\} = \mathbf{V}^{\cdot(D)} \boldsymbol{\Gamma}^{\cdot(D)} (\mathbf{V}^{\cdot(D)})^T \quad (115)$$

donde $(\mathbf{V}^{\cdot(D)})^T \mathbf{V}^{\cdot(D)} = \mathbf{I}$, y las ponderaciones gamma a nivel de grupo se calculan como

$$\mathbf{Y}^{(D)} = \mathbf{v}_1^{(D)} \times (\mathbf{1}^T \mathbf{v}_1^{(D)})^{-1} \quad (116)$$

El gran compromiso se calcula como,

$$\mathbf{S}^{(\cdot)} = \sum_d^D \gamma^{(d)} \times \mathbf{S}^{(d)} \quad (117)$$

Se realiza la descomposición propia del gran compromiso

$$eigen \{ \mathbf{S}^{(\cdot)} \} = \mathbf{U}^{(\cdot)} \mathbf{\Lambda}^{(\cdot)} (\mathbf{U}^{(\cdot)})^T \quad (118)$$

donde $(\mathbf{U}^{(\cdot)})^T \mathbf{U}^{(\cdot)} = \mathbf{I}$, para obtener los puntajes factoriales del gran compromiso, $\mathbf{F}^{(\cdot)}$, la mejor representación individual de las perspectivas de los estímulos de los grupos.

$$\mathbf{F}^{(\cdot)} = \mathbf{U}^{(\cdot)} (\mathbf{\Lambda}^{(\cdot)})^{\frac{1}{2}} = \mathbf{S}^{(\cdot)} \mathbf{U}^{(\cdot)} (\mathbf{\Lambda}^{(\cdot)})^{-\frac{1}{2}} \quad (119)$$

$\mathbf{U}^{(\cdot)} (\mathbf{\Lambda}^{(\cdot)})^{-\frac{1}{2}}$ es la matriz de proyección para el gran compromiso, y puede ser utilizada para proyectar los compromisos del grupo en el espacio de factores del gran compromiso.

Los sobre-ponderados compromisos grupales se proyectan en el espacio factorial de gran compromiso para dar puntajes factoriales parciales grupales, $\mathbf{F}^{(D)}$, que se calculan como

$$\mathbf{F}^{(d)} = D \gamma^{(d)} (\lambda_1^{(d)})^{-1} = \mathbf{S}^{(\cdot)} \mathbf{U}^{(\cdot)} (\mathbf{\Lambda}^{(\cdot)})^{-\frac{1}{2}} \quad (120)$$

Las sobre-ponderaciones de participaciones individuales también se proyectan en el espacio factorial del gran compromiso para dar puntajes factoriales parciales individuales, $\mathbf{F}^{c(D)}$ y se expresan como,

$$\mathbf{F}^{c(d)} = \bar{\mathbf{S}}^{c(d)} \mathbf{U}^{(\cdot)} (\mathbf{\Lambda}^{(\cdot)})^{-\frac{1}{2}} \quad (121)$$

Estos puntajes factoriales se pueden inspeccionar de varias maneras, como todas las perspectivas de un estímulo dado o una perspectiva dada de todos los estímulos.

La magnitud de los efectos experimentales se puede expresar mediante varias sumas de cuadrados y valores de r^2 . La forma en que HiDiSTATIS segmenta los

efectos experimentales es única, porque las sumas de cuadrados calculadas en ANOVA y DiMDS representan las distancias al cuadrado desde los puntajes de los factores hasta el origen. Sin embargo, esto cuantifica la variabilidad de los estímulos. En HiDiSTATIS, las distancias al cuadrado al origen deben sustraerse para calcular la variabilidad entre y dentro de los grupos. Por esta razón, la desviación entre

$$SS^{C(D)} = \sum_{c(d)}^{C(D)} \text{traza} \{(\mathbf{F}^{c(d)})^T \mathbf{F}^{c(d)}\} \quad (122)$$

(suma de cuadrados de distancias de todos los puntos de datos), y

$$SS^{\cdot(D)} = C(d) \times \sum_d^D \text{traza} \{(\mathbf{F}^{\cdot(d)})^T \mathbf{F}^{\cdot(d)}\} \quad (123)$$

(suma de cuadrados de la perspectiva de grupos), da la variabilidad dentro del grupo,

$$SS^{\text{Within}} = SS^{C(D)} - SS^{\cdot(D)} \quad (124)$$

y la desviación entre $SS^{\cdot(D)}$ y $SS^{\cdot(\cdot)}$.

$$SS^{\cdot(\cdot)} = C(D) \times \text{traza} \{(\mathbf{F}^{\cdot(\cdot)})^T \mathbf{F}^{\cdot(\cdot)}\} \quad (125)$$

proporciona la variabilidad entre grupos,

$$SS^{\text{Between}} = SS^{\cdot(\cdot)} - SS^{\cdot(D)} \quad (126)$$

La variabilidad total de todos los participantes se particiona de la siguiente manera

$$SS^{C(D)} = SS^{\cdot(\cdot)} + SS^{\text{Between}} + SS^{\text{Within}} \quad (127)$$

La parte de variabilidad explicada por los grupos se calcula como,

$$r_{\text{Grupos}}^2 = \frac{SS^{\text{Between}}}{SS^{\text{Between}} + SS^{\text{Within}}} \quad (128)$$

Para probar cómo se generalizan los resultados de HiDiSTATIS, el autor presenta dos tipos de pruebas de inferencia no paramétricas: una prueba de permutación y una técnica de remuestreo llamada Bootstrap.

2.4.19. DIDISTATIS

Kriegsman (2018a), desarrolla un método llamado DiDiSTATIS o DISTATIS discriminante, el cual consiste en analizar una matriz de matrices de distancia con un diseño anidado de un factor en las filas y un diseño anidado de un factor en las tablas. Para hacerlo, integra el método de *escalamiento multidimensional discriminante (DiMDS)* y DiSTATIS jerárquico (**HiDiSTATIS**). DiDiSTATIS primero integra las tablas para dar el gran compromiso baricéntrico. El espacio factorial del gran compromiso baricéntrico representa de manera óptima cómo los grupos (de participantes) percibieron de manera diferente las categorías (de estímulos).

DiDiSTATIS toma utiliza el DiMDS para definir un subespacio discriminante baricéntrico que captura específicamente las partes de los patrones de clasificación relacionados con las diferencias percibidas entre las categorías de fila. Este subespacio discriminante baricéntrico es similar al análisis de contraste, definido por categorías de estímulo (tipo ANOVA). Como resultado, cuando los estímulos se proyectan en el subespacio discriminante baricéntrico, sus posiciones con respecto a los baricentros muestran cómo cada estímulo fue percibido, más como esto o aquella categoría. De esta manera, el espacio del factor de gran compromiso baricéntrico es el mejor espacio posible para evaluar cuán consistentemente se clasificaron los estímulos en las categorías hipotetizadas. DiDISTATIS también usa el HiDiSTATIS, para ponderar jerárquicamente las tablas con el fin de representar de manera óptima los grupos de participantes.

Kriegsman extiende el MDS/EMD (Torgerson, 1958), a escalamiento multidimensional discriminante (DiMDS), para realizar análisis discriminantes en las filas de una matriz de distancia. El DiMDS transforma cada matriz de distancias en matrices de similitud semi-definidas positivas, aplica el diseño

anidado hipotético de un factor en las filas de la matriz de similitud y luego hace la descomposición propia a esta matriz. El resultado (similar al análisis de contraste) es un espacio factorial que corresponde a la hipótesis, en el que la variabilidad de los estímulos sobre sus baricentros hipotéticos revela la calidad de la discriminación.

El enfoque discriminante baricéntrico se refiere al procedimiento de análisis discriminante baricéntrico (Abdi & Williams, 2010; Abdi, Williams, & Béra, 2018), que es similar al ANOVA multidimensional. Un baricentro es una media ponderada (*la media de las observaciones de una categoría dada, también llamada centro de gravedad, centro de masa, vector medio o centroide*), y el *análisis discriminante baricéntrico (BADA)* es un tipo de análisis discriminante que maximiza la variabilidad entre los centros de grupo. Es un enfoque diferente a otros métodos discriminantes, como el análisis discriminante lineal, que maximiza la relación entre la variabilidad entre grupos y dentro del grupo.

En **BADA**, las filas (estímulos) de una matriz rectangular, \mathbf{X} , se promedian dentro de las categorías, para dar una matriz rectangular con una fila para cada categoría, llamada matriz de baricentro. La matriz del baricentro se descompone para dar al subespacio discriminante baricéntrico, de \mathbf{X} , un espacio que maximiza las diferencias entre las categorías, (las filas de la matriz del baricentro). BADA segmenta la variabilidad total en \mathbf{X} en 3 fuentes de variabilidad:

el efecto entre categorías,
el efecto dentro de la categoría y
el residual.

El residuo corresponde a la porción de \mathbf{X} que no se proyecta (es decir, se ajusta) en el subespacio discriminante baricéntrico.

Como indica el autor, BADA da un espacio de menor dimensión que maximiza las diferencias entre las categorías de estímulo. Sin embargo, BADA se define solo para datos rectangulares y, por lo tanto, para ser utilizado para clasificar datos, BADA debe adaptarse para el análisis de datos semi-definidos positivos.

En BADA, los datos de estímulo rectangular se denotan con \mathbf{X} . La matriz de proyección barcéntrica, \mathbf{P}_B , se calcula como,

$$\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T \quad (129)$$

donde \mathbf{B} denota una matriz de diseño para las filas de \mathbf{X} . La matriz \mathbf{X} , se proyecta en el subespacio discriminante baricéntrico como,

$$\mathbf{P}_B\mathbf{X} \quad (130)$$

La matriz $\mathbf{P}_B\mathbf{X}$, se descompone para dar el subespacio discriminante baricéntrico. Entonces, \mathbf{X} , los estímulos de la fila, se pueden proyectar en el subespacio discriminante baricéntrico.

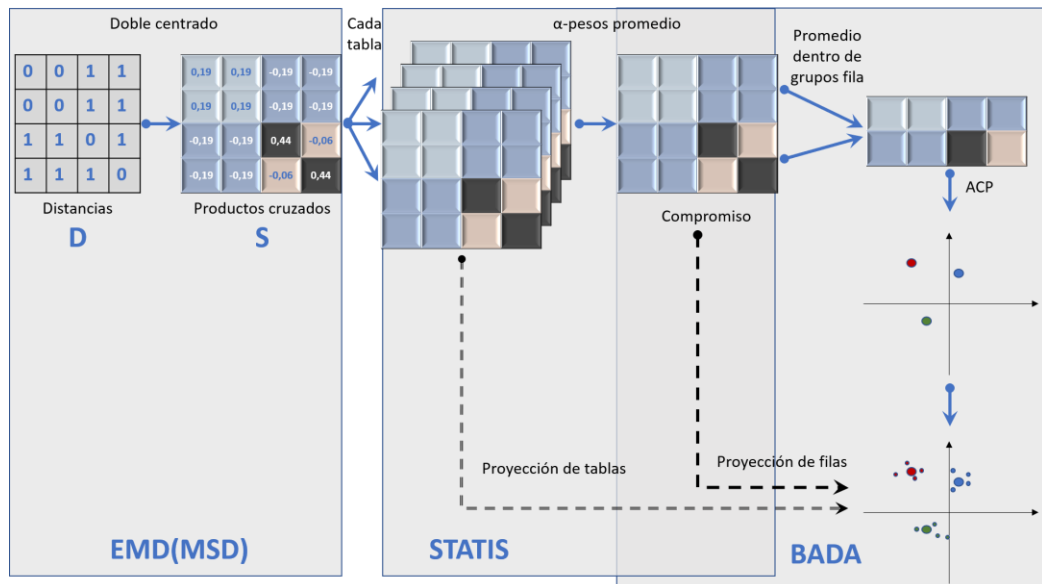


Figura 34. Proceso del método DiDISTATIS, adaptado de Kriegsman et al., 2015

El método DiDISTATIS inicia con un arreglo de (clasificación) matrices de distancia. Como en MDS, se realiza un doble centro de cada matriz de distancia, para obtener un arreglo de matrices de similitud. La Figura 34 muestra un esquema de los pasos del DiDISTATIS. En el Paso I de DiDISTATIS, las tablas se integran en una sola tabla representativa, llamada gran compromiso baricéntrico, $S_B^{(\cdot)}$. En el Paso II, se realiza una descomposición propia del gran compromiso baricéntrico. Los datos sobre-ponderados (que incluyen las perspectivas mayores, grupales e individuales sobre las categorías y sobre los

estímulos) se proyectan luego en el espacio factorial del gran compromiso baricéntrico. Las estadísticas de resumen se calculan para describir la muestra.

El procedimiento para integrar tablas en DiDiSTATI, intercala los procedimientos de HiDiSTATIS y DiMDS. Para comenzar, como en HiDiSTATIS, las tablas individuales $\tilde{\mathbf{S}}^{[c(d)]}$, son integradas dentro de sus grupos, para proporcionar grupos compromisos $\tilde{\mathbf{S}}^{[·(D)]}$. Entonces, como en DiMDS, cada uno de los grupos compromisos se proyectan en su propio subespacio baricéntrico para proporcionar un arreglo de grupos compromiso de baricéntrico, $\tilde{\mathbf{S}}_B^{[·(D)]}$. De nuevo, como en HiDiSTATIS, estos grupos compromisos de baricéntrico se integran una para obtener el gran compromiso baricéntrico, $\mathbf{S}_B^{(·)}$.

Este procedimiento permite que las perspectivas individuales de la estructura de similitud entre estímulos se acumulen a nivel grupal. Cada representación a nivel de grupo se reduce a su subespacio baricéntrico discriminante. Por lo tanto, cada compromiso de grupo baricéntrico $\tilde{\mathbf{S}}_B^{[·(d)]}$, indica cómo un grupo dado percibió las categorías de estímulo. El gran compromiso baricéntrico se calcula como el espacio común óptimo para ver cómo los grupos percibieron de manera diferente las categorías de estímulo. Los primeros tres pasos proceden precisamente como en HiDiSTATIS.

El procedimiento inicia con la normalización con MFA de cada tabla individual,

$$\mathbf{S}^{c(d)} = \left(\tilde{\lambda}_1^{c(d)} \right)^{-1} \times \tilde{\mathbf{S}}^{c(d)} \quad (131)$$

De tal forma que cada participante proporcione una cantidad comparable de información.

Se calculan pesos alfa para las tablas dentro de cada grupo para proporcionar,

$$\boldsymbol{\gamma}^{c(d)} = \mathbf{v}_1^{c(d)} \times \left(\mathbf{1}^T \mathbf{v}_1^{c(d)} \right)^{-1} \quad (132)$$

Cada compromiso de grupo se calcula como la suma sobre las tablas alfa ponderadas normalizadas con MFA dentro de ese grupo, como,

$$\tilde{\mathbf{S}}^{(d)} = \sum_{c^{(d)}}^{C^{(d)}} \gamma^{c^{(d)}} \mathbf{X} \mathbf{S}^{c^{(d)}} \quad (133)$$

El compromiso del grupo $\tilde{\mathbf{S}}^{(d)}$, representa el consenso de cómo los participantes en ese grupo clasifican los estímulos.

A partir de aquí, los procedimientos de HiDiSTATIS y DiDiSTATIS divergen, y se introduce un aspecto del DiMDS. Es decir, cada compromiso de grupo se proyecta en su subespacio discriminante baricéntrico. Esto proporciona los llamados compromisos de grupo baricéntrico, matrices que representan cómo cada grupo percibió las categorías (estímulo).

Cada grupo compromiso $\tilde{\mathbf{S}}^{(d)}$, se proyecta en su subespacio discriminante baricéntrico, como

$$\tilde{\mathbf{S}}_B^{(d)} = \mathbf{P}_B \tilde{\mathbf{S}}^{(d)} \mathbf{P}_B \quad (134)$$

donde \mathbf{P}_B es la proyección baricéntrica como se define en DiMDS.

Seguidamente, los procedimientos de DiMDS y HiDiSTATIS se entrelazan para integrar los compromisos del grupo baricéntrico para obtener el gran compromiso baricéntrico.

Cada grupo compromiso baricéntrico $\tilde{\mathbf{S}}_B^{(d)}$, es normalizado con MFA como,

$$\mathbf{S}_B^{(d)} = \left(\tilde{\lambda}_{B,1}^{(d)} \right)^{-1} \tilde{\mathbf{S}}_B^{(d)} \quad (135)$$

donde $\tilde{\lambda}_{B,1}^{(d)}$ el primer valor propio de $\tilde{\mathbf{S}}_B^{(d)}$. Esta normalización con MFA a nivel de grupo asegura que cada grupo contribuya con la misma cantidad de información.

El autor señala que, dependiendo de la perspectiva del experimentador, esta normalización a nivel de grupo $\left(\tilde{\lambda}_{B,1}^{(d)} \right)^{-1}$ puede no ser deseable, por lo que en su software da la opción de usar el coeficiente calculado para cada grupo, o reemplazarlo por la media de los coeficientes de normalización de grupo D.

A los compromisos del grupo baricéntrico normalizado con MFA se les aplica un análisis RV entre grupos para identificar los pesos gamma a nivel de grupo, $\gamma_B^{(D)}$. Es decir, la matriz RV entre grupos, $\mathbf{R}_B^{(D)}$, se calcula como,

$$\mathbf{R}_B^{(D)} = \left[\text{RV} \left\{ \mathbf{S}_B^{(d)}, \mathbf{S}_B^{(d')} \right\} \right] \quad (136)$$

y su descomposición propia está dada por,

$$\text{eigen} \left\{ \mathbf{R}_B^{(D)} \right\} = \mathbf{V}_B^{(D)} \mathbf{\Gamma}_B^{(D)} \left(\mathbf{V}_B^{(D)} \right)^T \quad (137)$$

donde $\left(\mathbf{V}_B^{(D)} \right)^T \mathbf{V}_B^{(D)} = \mathbf{I}$ y los correspondientes pesos alfa de nivel de grupo se calculan a partir del primer vector propio correspondiente, $\mathbf{v}_{B,1}^{(D)}$

$$\gamma_B^{(D)} = \mathbf{v}_{B,1}^{(D)} \mathbf{x} \left(\mathbf{1}^T \mathbf{v}_{B,1}^{(D)} \right)^{-1} \quad (138)$$

El gran compromiso baricéntrico, $\mathbf{S}_B^{(\cdot)}$, se calcula como,

$$\mathbf{S}_B^{(\cdot)} = \sum_d^D \gamma_B^{(d)} \mathbf{S}_B^{(d)} \quad (139)$$

En el Paso II, se realiza la descomposición propia del gran compromiso baricéntrico, y las otras tablas se proyectan en este espacio. Los puntajes factoriales gran baricéntricos proporcionan las posiciones de las categorías de estímulo. Los puntajes factoriales parciales del grupo baricéntrico dejar ver cómo los grupos percibieron de manera diferente las categorías de estímulo. Las puntuaciones factoriales parciales del grupo revelan cómo los grupos percibieron de manera diferente la variabilidad de los estímulos dentro de las categorías. La descomposición propia del gran compromiso baricéntrico se expresa como,

$$\text{eigen} \left\{ \mathbf{S}_B^{(\cdot)} \right\} = \mathbf{U}_B^{(\cdot)} \mathbf{\Lambda}_B^{(\cdot)} \left(\mathbf{U}_B^{(\cdot)} \right)^T \quad (140)$$

con $\left(\mathbf{U}_B^{(\cdot)} \right)^T \mathbf{U}_B^{(\cdot)} = \mathbf{I}$

Se realiza la descomposición propia del gran compromiso para obtener los puntajes factoriales del gran compromiso, $\mathbf{F}_B^{(\cdot)}$, las coordenadas de las categorías de estímulo desde una perspectiva general, se calculan como,

$$\mathbf{F}_B^{(\cdot)} = \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{\frac{1}{2}} = \mathbf{S}_B^{(\cdot)} \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (141)$$

$\mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}}$ es la matriz de proyección para el espacio factorial del gran compromiso baricéntrico.

Los puntajes factoriales parciales grupales, $\mathbf{F}_B^{(D)}$, se calculan como

$$\mathbf{F}_B^{(d)} = D \gamma_B^{(d)} \left(\tilde{\lambda}_{B,1}^{(d)} \right)^{-1} \tilde{\mathbf{S}}_B^{(d)} \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (142)$$

Las sobre-ponderaciones de los compromisos de grupo, también se proyectan.

Los puntajes factoriales parciales grupales, $\mathbf{F}_{Disc}^{[D]}$ y se expresan como,

$$\mathbf{F}_{Disc}^{[D]} = D \gamma_B^{(d)} \left(\tilde{\lambda}_{B,1}^{(d)} \right)^{-1} \tilde{\mathbf{S}}^{(d)} \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (143)$$

Estos puntajes factoriales se pueden inspeccionar de varias maneras, como todas las perspectivas de un estímulo dado o una perspectiva dada de todos los estímulos.

Luego las puntuaciones factoriales parciales individuales se calculan como,

$$\mathbf{F}_{Disc}^{[c(d)]} = \bar{\mathbf{S}}^{c(d)} \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (144)$$

donde $\bar{\mathbf{S}}^{c(d)}$ se define como una tabla de datos individuales sobreponderada.

El gran compromiso sobreponderado se calcula como,

$$\bar{\mathbf{S}}^{(\cdot)} = \sum_d^D \bar{\mathbf{S}}^{(d)} = \sum_d^D D \gamma_B^{(d)} \left(\tilde{\lambda}_{B,1}^{(d)} \right)^{-1} \tilde{\mathbf{S}}^{(d)} \quad (145)$$

La proyección del gran compromiso baricéntrico está dada por,

$$\mathbf{F}_{Disc}^{(\cdot)} = \bar{\mathbf{S}}^{(\cdot)} \mathbf{U}_B^{(\cdot)} \left(\mathbf{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (146)$$

Las tablas de datos baricéntricos sobreponderados a nivel de participantes se calculan como,

$$\bar{\mathbf{S}}^{c(d)} = \mathbf{P}_B \bar{\mathbf{S}}^{c(d)} \quad (147)$$

La proyección del gran compromiso baricéntrico se obtiene con,

$$\mathbf{F}_B^{c(d)} = \bar{\mathbf{S}}_B^{c(d)} \mathbf{U}_B^{(\cdot)} \left(\boldsymbol{\Lambda}_B^{(\cdot)} \right)^{-\frac{1}{2}} \quad (148)$$

La magnitud de los efectos experimentales se puede expresar mediante varias sumas de cuadrados, valores de r^2 y estadísticas que relacionan la calidad del modelo predictivo (matriz de confusión).

Las sumas de cuadrados para cada uno de los 6 tipos de puntajes factoriales anteriores se calculan como la suma de las distancias al cuadrado desde los puntajes factoriales hasta el origen, para más detalles ver (Kriegsman, 2018a). Todos los mapas DiDiSTATIS describen solo el subespacio discriminante baricéntrico, que captura solo una parte (la parte relacionada con las categorías hipotéticas) de los datos de clasificación total.

DiDiSTATIS utiliza las mismas estadísticas de resumen introducidas tanto en DiMDS como en HiDiSTATIS. Esto incluye estadísticas que relacionan la magnitud de los efectos experimentales (sumas de cuadrados y r^2) y estadísticas que relacionan la calidad del modelo predictivo (matriz de confusión).

Para evaluar la importancia, la estabilidad y la calidad de la predicción, respectivamente, el autor presenta pruebas de inferencia no paramétricas: permutación, bootstrap y validación cruzada.

2.4.20. CATATIS

CATATIS(Llobell, Cariou, et al., 2019), fue desarrollado para el análisis de datos CATA (Check all that apply-marcar todo los que corresponda) en el campo de la evaluación sensorial, consiste en considerar los conjuntos de datos 0-1 asociados con los sujetos que indican si los sujetos han verificado los atributos o no. Estos conjuntos de datos se promedian sobre los temas, produciendo una

tabla de frecuencias que tabula de forma cruzada los productos y los atributos. Esta tabla indica la proporción de sujetos que han verificado cada atributo para cada producto. Es una tarea simple que no requiere una formación específica de los sujetos (Ares y Jaeger, 2015; Varela y Ares, 2012). Dado un conjunto de productos, a cada sujeto se le presenta una lista de atributos y se le indica que verifique los atributos que se aplican a cada producto. Para una visión general de los métodos utilizados para analizar los datos obtenidos de este modo, nos remitimos al documento de Meyners, Castura y Carr (2013). Estos métodos abarcan estrategias para evaluar la importancia de las diferencias entre los productos por atributo y en todos los atributos, estrategias para evaluar las relaciones entre los atributos, representaciones gráficas para mostrar las relaciones entre productos y atributos, etc. Entre otros En los métodos, se recomienda el análisis de correspondencia (CA) para mostrar gráficamente los atributos y los productos que muestran cómo se relacionan entre sí y, por lo tanto, dar una caracterización de los productos (Ares, Barreiro, Deliza, Giménez y Gambaro, 2010). Más precisamente, CA se realiza en la tabla de contingencia o en la tabla de frecuencia que tabula de forma cruzada los productos y los atributos para todo el panel de temas.

2.4.21. STATICO-COA

Propuesto por Mérigot et al, (2019), STATICO-CoA, combina el análisis STATICO con análisis de correspondencia. Al igual que en el método, los datos analizados son una secuencia de pares de matrices, que en el caso de la propuesta son variables ambientales y las densidades de especies respectivamente. En esta propuesta, primero, cada matriz con las variables ambientales se analiza primero se aplicando un análisis de correspondencias-CoA (Benzécri, 1973) y un PCA a las matrices con las densidades de especies, luego se aplica un Análisis de Co-inercia (COIA) (Doledec & Chessel, 1994), para cada par de tablas y por último se aplica un XSTATIS o Análisis Triádico Parcial (PTA) (Thioulouse & Chessel, 1987).

El análisis STATICO “combina los objetivos de STATIS (encontrar la parte estable de la estructura de una serie de tablas) y los objetivos del análisis de co-inercia (encontrar la estructura común de dos tablas de datos)”. En comparación

con la versión clásica de STATICO (Thioulouse et al., 2004), STATICO-CoA se basa en la lógica del análisis de correspondencia (CoA).

2.4.22. CLUSTATIS

Este método de agrupamiento está estrechamente relacionado con el método STATIS, consiste en un análisis de agrupamiento jerárquico y un algoritmo de partición. Estas dos estrategias apuntan a optimizar el mismo criterio; ya sea localmente para la agrupación jerárquica o globalmente para el algoritmo de partición. Se pueden ejecutar de forma independiente o en combinación en un intento de lograr una solución aún mejor que la obtenida por la aplicación de solo una de las dos estrategias.

CLUSTATIS puede verse como una extensión del análisis de agrupamiento de variables alrededor de componentes latentes (Vigneau & Qannari, 2003) para el caso de bloques de variables. Sigue el mismo patrón de análisis y disfruta de las mismas propiedades. Funciona con datos multibloque, que es una configuración más grande que los datos de tres vías, ya que las variables pueden no ser las mismas para todos los datos disponibles.

Su objetivo es minimizar un criterio que refleja el hecho de que se buscan grupos homogéneos de conjuntos de datos. Más precisamente, se supone que los conjuntos de datos en cada grupo están altamente relacionados con una configuración latente que se determina mediante el método STATIS.

Sean $\mathbf{X}_1, \dots, \mathbf{X}_H$ los conjuntos de datos, que se supone que está centrado en columnas, medidas en los mismos I individuos, pero las variables (J_h) pueden ser diferentes de un conjunto de datos a otros.

Se calculan las matrices de productos escalares,

$$\mathbf{W}_h = \mathbf{X}_h \mathbf{X}_h^T \quad (149)$$

Como el el método STATIS el coeficiente R_v refleja la similaridad entre dos configuraciones

$$RV(\mathbf{W}_h, \mathbf{W}_{h'}) = \frac{\langle \mathbf{W}_h, \mathbf{W}_{h'} \rangle}{\|\mathbf{W}_h\| \|\mathbf{W}_{h'}\|} = \frac{\text{traza}(\mathbf{W}_h, \mathbf{W}_{h'})}{\sqrt{\text{traza}(\mathbf{W}_h, \mathbf{W}_h)} \sqrt{\text{traza}(\mathbf{W}_{h'}, \mathbf{W}_{h'})}} \quad (150)$$

Se asume que $\|\mathbf{W}_h\| = 1$, para $i=1, \dots, H$, lo que significa que estas matrices se han estandarizado dividiendo cada \mathbf{W}_h original por su norma, $\|\mathbf{W}_h\| = \sqrt{\text{traza}(\mathbf{W}_h, \mathbf{W}_h)}$.

CLUSTATIS busca grupos homogéneos de conjuntos de datos de tal manera que los conjuntos de datos en cada grupo están lo más cerca posible de un conjunto de datos promedio de grupo. Formalmente, CLUSTATIS tiene como objetivo minimizar el siguiente criterio:

$$D = \sum_{k=1}^K \sum_{\ell=1}^{G_k} \|\mathbf{w}_\ell - \gamma_\ell \mathbf{W}^{(k)}\|^2 \quad (151)$$

donde γ_ℓ son escalares a determinar y se supone que,

$$\sum_{\ell=1}^{G_k} \gamma_\ell^2 = 1 \quad (152)$$

siendo K el número de grupos del conjunto de datos, G_k es el k -ésimo grupo y, para k ($k=1, \dots, K$), $\mathbf{W}^{(k)}$ es el compromiso del grupo G_k . Obviamente, cuando solo hay un grupo de conjuntos de datos ($K=1$), se obtiene el mismo criterio que subyace al método STATIS.

Este criterio señala el hecho de que se buscan grupos de conjuntos de datos de tal manera que, en cada grupo G_k , las matrices \mathbf{W}_h se agrupan alrededor de una matriz promedio, $\mathbf{W}^{(k)}$.

Los pesos γ_ℓ en cada grupo están dados por el método STATIS. Por lo tanto, aquellos conjuntos de datos en un grupo dado que tienen menos acuerdo con los otros conjuntos de datos en este grupo tendrán pesos pequeños. Por el contrario, aquellos conjuntos de datos que estén más en línea con el punto de vista general del grupo en consideración recibirán pesos relativamente más mayores.

Este procedimiento implica dos estrategias de agrupamiento complementarias. La primera estrategia consiste en un análisis de conglomerados jerárquico. La segunda estrategia consiste en un algoritmo de partición similar al algoritmo K-means. Ambas estrategias apuntan a optimizar el mismo criterio local o globalmente y, en la práctica, se complementan entre sí. Más precisamente, el análisis jerárquico de conglomerados puede ayudar a seleccionar el número apropiado, K , de conglomerados y proporciona una partición inicial de los conjuntos de datos que puede mejorarse mediante el algoritmo de partición.

El algoritmo jerárquico sigue un ascenso (o fusión) por sí mismo. Se inicia con la situación en la que cada conjunto de datos forma un grupo por sí mismo. Obviamente, en este caso, $D = 0$. En cada paso, fusionamos dos conjuntos de datos o, más generalmente, a medida que avanza el algoritmo, dos grupos de conjuntos de datos hasta que todos los conjuntos de datos se fusionen en un solo clúster. En cada paso, podemos mostrar que el criterio D aumenta. Más precisamente, podemos mostrar que cuando los dos grupos A y B se fusionan, el criterio D aumenta en $delta = \delta_1^{(A)} + \delta_1^{(B)} - \delta_1^{(AUB)}$, donde $\delta_1^{(A)}$, $\delta_1^{(B)}$ y $\delta_1^{(AUB)}$ son respectivamente el mayor valor propio de la matriz de coeficientes RV entre los pares de configuraciones de los clusters A, B y AUB .

El fundamento de la estrategia de agregación es fusionar esos dos grupos A y B (por ejemplo) que resultan en el aumento más pequeño del criterio D . Se debe examinar el aumento del criterio D a medida que avanza el agrupamiento jerárquico porque refleja la pérdida de heterogeneidad cuando se fusionan los grupos A y B . Un salto de esta cantidad indica que se está tratando de fusionar dos grupos que son heterogéneos. En la práctica, estas cantidades se reflejan en el árbol jerárquico (o dendrograma) como la altura de las ramas que conectan dos nodos incrustados. Alternativamente, estas cantidades podrían representarse como un gráfico de barras que muestra su evolución a medida que disminuye el número de grupos.

El problema de agrupamiento basado en el criterio D dado anteriormente puede resolverse también mediante un algoritmo de partición similar al algoritmo K-means (Everitt, Landau, Leese, & Stahl, 2011; Lloyd, 1982). A medida que trabaja el algoritmo, se permite que los conjuntos de datos entren y salgan de

los grupos logrando en cada paso una disminución del criterio D. Este algoritmo asume que el número de grupos, K , se proporciona de antemano y consiste en tres pasos:

Paso 1: Correspondiente a la partición inicial de los conjuntos de datos. Se proporciona los K grupos de conjuntos de datos. Estos grupos podrían elegirse mediante una asignación aleatoria de los conjuntos de datos. Se puede realizar una mejor inicialización a partir de los resultados del agrupamiento jerárquico.

Paso 2: Determinación de las matrices de productos escalares de compromiso de conglomerados. En el conglomerado G_k , la matriz de productos escalares de compromiso, $\mathbf{W}^{(k)}$ y los pesos asociados γ_i se determinan mediante el método STATIS.

Paso 3: Cambio de clústeres. Los nuevos grupos de conjuntos de datos se forman moviendo cada conjunto de datos, \mathbf{X}_h , al clúster G_k para el cual $R_V(\mathbf{W}^{(h)}\mathbf{W}^{(k)})$ es el más grande.

Los pasos 2 y 3 se repiten hasta que no haya cambios en los integrantes del clúster. Esto significa que el criterio D para de disminuir. En la práctica, tanto los algoritmos jerárquicos como los de particionamiento debe realizarse para alcanzar una mejor solución. En primer lugar, la estrategia jerárquica se puede utilizar para sugerir un número apropiado de grupos examinando la evolución del criterio de agregación en el curso del proceso de agregación. En segundo lugar, los conjuntos de datos se someten al algoritmo de partición utilizando como solución inicial, la partición obtenida seccionando el árbol jerárquico en el nivel indicado, es decir, con el número seleccionado de grupos. Al permitir el cambio de integrantes del clúster, la solución obtenida por el agrupamiento jerárquico puede mejorarse ya que es probable que el criterio D se minimice más.

Dado que el algoritmo jerárquico puede llevar mucho tiempo en situaciones en las que el número de conjuntos de datos disponibles es grande, solo se puede usar el algoritmo de partición. Sin embargo, se recomienda realizar una partición aleatoria de inicio múltiple, ejecutando este algoritmo utilizando varias particiones (aleatorias) como puntos de partida. Finalmente, se guarda la solución final que corresponde al valor más pequeño del criterio D.

Asociado con cada grupo de sujetos G_k , la técnica produce una matriz de promedios ponderados $\mathbf{W}^{(k)}$, que es el compromiso de los productos escalares obtenidos por el STATIS aplicado a los conjuntos de datos en G_k . Una descomposición propia de la matriz $\mathbf{W}^{(k)}$ hace posible escribir $\mathbf{W}^{(k)} = \mathbf{C}^{(k)}(\mathbf{C}^{(k)})^T$, en donde la matriz $\mathbf{C}^{(k)}$ representa la configuración promedio del grupo en el conglomerado G_k .

Hay varios índices asociados con la solución final que son de mucha importancia:

En primer lugar, se considera, para cada grupo G_k , ($k = 1, \dots, K$), el índice $I_k = \frac{\lambda_1^k}{H_k}$, donde λ_1^k es el valor propio más grande de la matriz de coeficientes RV entre los conjuntos de datos en el grupo G_k y H_k es el número de conjuntos de datos en este grupo. Este índice oscila en el intervalo $(\frac{1}{H_k}, 1)$ y refleja la homogeneidad en G_k . Se puede interpretar como el porcentaje de variación en los conjuntos de datos del grupo G_k que se explica por la configuración promedio del grupo asociada con $\mathbf{W}^{(k)}$.

Dentro del grupo G_k , se puede calcular para cada conjunto de datos \mathbf{X}_h , el coeficiente RV entre \mathbf{W}_h y $(\mathbf{W}^{(k)})$. Este índice refleja cómo cada conjunto de datos está cerca de su configuración de grupo asociada. Alternativamente, se podría considerar el coeficiente γ_h ($h = 1, \dots, H$) que proporciona la misma idea. Un índice general para evaluar la calidad de la partición de las tablas de datos obtenidas por el enfoque de agrupamiento está dado por la ponderación promedio de los índices I_k , y está dado por,

$$I = \frac{\sum_{k=1}^K m_k I_k}{m} = \frac{\sum_{k=1}^K \lambda_1^{(k)}}{m} \quad (153)$$

Este índice puede ser interpretado como el porcentaje de variación en los conjuntos de datos originales explicados por las configuraciones promedio de grupo en los diversos grupos.

Finalmente, para evaluar cómo los diferentes grupos de conjuntos de datos están cerca uno del otro, se calculan los coeficientes RV entre sus configuraciones promedio de grupo asociadas.

Una descomposición interesante similar a ANOVA es la siguiente:

$$\sum_{h=1}^H \|\mathbf{w}_h\|^2 = \sum_{k=1}^K \|\mathbf{w}^{(k)}\|^2 + \sum_{k=1}^K \sum_{i \in G_k} \|\mathbf{w}_i - \gamma_i \mathbf{w}^{(k)}\|^2 \quad (154)$$

Esta ecuación establece que la variación total en los conjuntos de datos medidos por $\sum_{h=1}^H \|\mathbf{w}_h\|^2$ es igual a la suma de la variación explicada por el modelo de agrupación en los K grupos, $\sum_{k=1}^K \|\mathbf{w}^{(k)}\|^2$, y la variación no explicada $\sum_{k=1}^K \sum_{i \in G_k} \|\mathbf{w}_i - \gamma_i \mathbf{w}^{(k)}\|^2$ que se busca minimizar. Se puede mostrar que el porcentaje de variación explicada dada por $I = \frac{\sum_{k=1}^K \|\mathbf{w}^{(k)}\|^2}{\sum_{h=1}^H \|\mathbf{w}_i\|^2}$ es igual a $\frac{\sum_{k=1}^K \lambda_1^{(k)}}{H}$ donde $\lambda_1^{(k)}$ es el valor propio más grande de la matriz de coeficientes RV por pares entre los conjuntos de datos en el k -ésimo conglomerado. Este índice refleja la homogeneidad general dentro de los conglomerados. CLUSTATIS también se puede definir como un método que busca maximizar la cantidad $\sum_{k=1}^K \sum_{i \in G_k} RV^2(\mathbf{w}_i, \mathbf{w}^{(k)})$, que refleja la homogeneidad dentro del conglomerado.

CAPÍTULO III: SOFTWARE PARA STATIS

3.1. INTRODUCCIÓN

En los primeros desarrollos computacionales para los métodos STATIS (STATIS y STATIS dual), se trabajó en la plataforma CISIA (*Centre International de Statistique et d'Informatique Appliquées*)-(CISIA, 1989) , a través de códigos ejecutables en PC con sistema operativo MS-DOS. Se incluía también el código Fortran 77 que se ejecutaba en varios "mainframe". A finales de 1993 se implementó un software flexible más general en SPAD (*Système Pour l'Analyse des Données*), distribuido también por CISIA. Este software permitía escribir y ejecutar el procedimiento ACT en su propio entorno y era bastante simple ya que solo se requería rutinas habituales sobre cálculos de matriz y los propios sistemas. En la versión 3.2 implementada en el año 1997, se incluye el módulo SPADTM para el análisis de tablas múltiples mediante los métodos STATIS y Análisis Factorial Múltiple (AFM). La versión actual del programa es la 8.2 y presenta mejoras en las salidas gráficas. El programa permite elegir entre varias opciones: se puede realizar STATIS y STATIS dual, analizar tablas normalizadas o no, o decidir el número de ejes retenidos para el análisis.

Paralelamente se implementa el *ade-4* (Thioulouse, Chessel, Dolédec, & Olivier, 1997), un programa desarrollado en el Laboratorio de Biometría y Biología Evolutiva (UMR 5558) de la Universidad de Lyon I - Claude Bernard. Fue desarrollado para aplicar funciones de análisis para datos ecológicos y ambientales en el marco de los métodos de exploración euclidiana, en francés "*d'Analyse de Données destinée d'abord à la manipulation des données Écologiques et Environnementales avec des procédures Exploratoires d'essence Euclidienne*",(Funciones de análisis de datos para analizar datos ecológicos y ambientales en el marco del método exploratorio euclidiano), de ahí el nombre *ade-4*, (ver [Figura 35](#)).

El programa *ade4* propone una gran variedad de métodos explicativos para analizar conjuntos de datos multivariados, aunque el paquete está dirigido a la investigación ecológica, ha sido utilizado en otros campos. Este programa se caracteriza por: (a) la implementación de funciones estadísticas y gráficas; (b) la provisión de datos digitales; (c) la redacción de una documentación técnica y temática y; (d) la inclusión de referencias bibliográficas.



Figura 35. Programa ade-4, Thioulouse et al. (1997).

Ade-4 incluye métodos de análisis de una tabla como el análisis de componentes principales; métodos de acoplamiento de dos tablas como análisis de correspondencia canónica, análisis de redundancia y análisis de co-inercia; y además métodos para el análisis de tablas K, es decir, tablas de tres vías, particularmente los métodos de la escuela francesa como lo son: STATIS, X-STATIS, STATICO, entre otros métodos de análisis y gráficos. Desde 2002, ade4 es un paquete de software R, disponible en sistemas Windows, Mac OS, Linux y Unix.

En la actualidad existe una gran variedad de programas informáticos diseñados para implementar métodos de análisis de tablas múltiples, sin embargo, en nuestro trabajo nos enfocaremos en los paquetes MultBiplot desarrollado en la Universidad de Salamanca y en otros paquetes y/o librerías existentes en el lenguaje R, que aplican el método STATIS.

3.2. MULTBILOT

El programa MultBiplot (Vicente-Villardón, 2010), desarrollado en la Universidad de Salamanca, plasma la experiencia del Grupo de Estadística Aplicada del Departamento de Estadística de la Universidad de Salamanca en el trabajo con Biplots. El programa MULTBILOT es un programa para realizar Análisis Biplot (Biplot clásico, HJ-Biplot, Biplots logísticos, entre otros). Además de contener algunas de las técnicas multivariadas clásicas, también incluye métodos STATIS, como son: STATIS-ACT, Dual STATIS-ACT y X-STATIS (análisis parcial triádico-PTA).

El análisis STATIS en este programa genera los resultados numéricos y gráficos: valores propios de la interestructura, la interestructura, valores propios de la infraestructura, círculos de correlación con las variables, círculos de correlación de las componentes y las trayectorias para filas y para columnas.

El programa es muy amigable, ya que trabaja con menús y permite importar datos desde Excel.

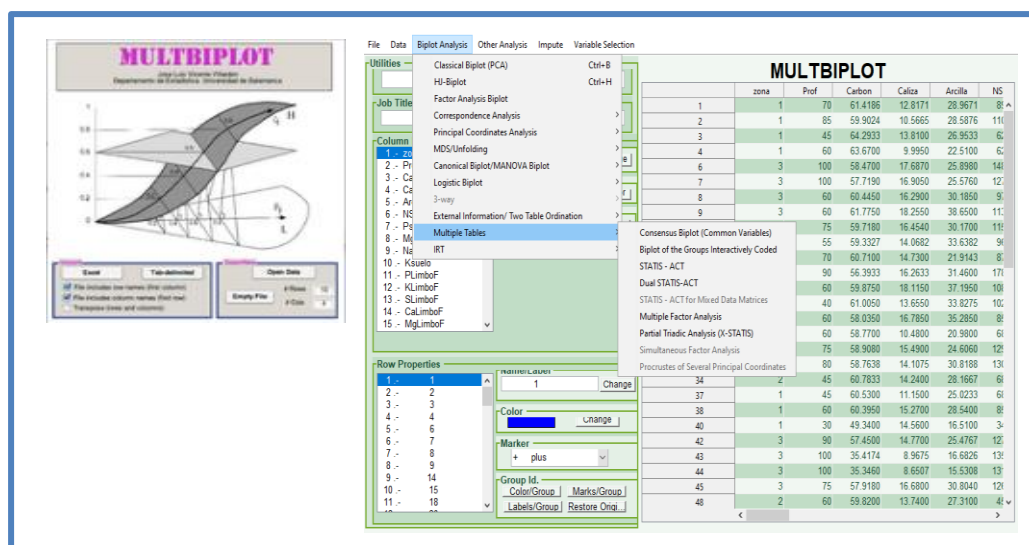


Figura 36 Métodos STATIS en el programa MULTBILOT. Vicente-Villardón, J.L, (2014)

Recientemente se creó nuevo paquete para R, **MultBiplotR: Multivariate Analysis using Biplots. R package version 0.1.0**, (ver Figura 36), que contiene todas las técnicas de la versión de escritorio de MultBiplot, agregado una interfaz gráfica para los métodos más básicos, entre estos el STATIS-dual, (Vicente-Villardón, 2015).

Los resultados generados por el programa MULTBILOT son los siguientes:

Resultados numéricos:

- Inter-estructura: matriz de productos escalares entre estudios u ocasiones; matriz de coeficiente de correlación RV entre estudios u ocasiones; configuración euclídea para estudios u ocasiones (valores propios, inercia e inercia acumulada); coordenadas para la configuración euclídea de las ocasiones u ocasiones; cosenos cuadrados (contribuciones) para la configuración euclídea de los estudios u ocasiones;
- Configuración Compromiso: tabla con los pesos de la configuración compromiso; representación de la configuración compromiso (valores propios, inercia e inercia acumulada); coordenadas principales para las filas del compromiso; porcentaje de contribución de las filas; calidades de representación acumuladas;
- Bondad de ajuste de los estudios u ocasiones en el biplot (% acumulativo); bondad de ajuste de los estudios u ocasiones en el biplot separado (% acumulativo); bondad de ajuste de las variables; correlaciones entre variables observadas y factores de consenso; correlaciones entre los componentes separados y del consenso.

3.3. LIBRERIAS Y PAQUETES EN R PARA STATIS

Unos de los lenguajes de programación que ha aportado grandes avances a la aplicaciones de métodos estadísticos ha sido el lenguaje R , creado por Ihaka y Gentleman (Ihaka & Gentleman, 1996), un sistema para análisis estadísticos y gráficos. Al ser R un lenguaje de programación de distribución gratuita, es una de las mejores opciones para el análisis de datos ya que es muy completo. Puede ser utilizado en sistemas Alpha Unix, Linux, Macintosh y Windows, ya que el código fuente está escrito básicamente en C, aunque algunas rutinas están escritas en Fortran.

Su versatilidad permite trabajar con pocos datos, al igual que con una gran cantidad de información, muy necesario en la era del Big Data. La flexibilidad

del programa, de permitir programar procesos propios, adecuados a lo que se necesita, es otra de las grandes ventajas de este.

En los últimos años se ha desarrollado una gran variedad de paquetes en R, y se ha implementado una interfase gráfica denominada GUI (Interfaz Gráfica del Usuario), que permite manejarse a través de menús, lo que facilita el uso a aquellos usuarios con poco conocimiento de programación.

Con el objetivo de conocer si existían paquetes que ofrecieran funcionalidades similares o relacionadas con el tema, se llevó a cabo una fase exploratoria. Para realizar dicha búsqueda consultaremos en el CRAN ([R Development Core Team, 2019](#)), GitHub («[GitHub](#)», 2019), R-Forge («[R-Forge](#)», 2019) y en Bioconductor ([Bioconductor, 2003](#)).

La búsqueda generó los siguientes resultados: [Clustblock \(Lobell, Vigneau, Cariou, & Qannari, 2019\)](#), [dad \(Boumaza et al., 2019\)](#), [ade4 \(Daniel Chessel, Dufour, & Thioulouse, 2004; Dray & Dufour, 2007; Dray, Dufour, & Chessel, 2007; Thioulouse et al., 2004\)](#), [DiDiSTATIS \(Kriegsman, 2018a\)](#), [SirStatis \(Villa-Vialaneix, 2016\)](#), [multigroup \(Eslami, Qannari, Bougeard, & Sanchez, 2015\)](#), [MultbiplotR \(Vicente-Villardón, 2015\)](#), [DistatisR \(D. Beaton, Fatt, & Abdi, 2013\)](#) y [MExPosition \(Chin Fatt, Beaton, & Abdi, 2015\)](#), que abordaban técnicas STATIS. En la [Tabla 2](#) se ha recogido información de cada uno de ellos con su respectivo nombre y una breve descripción.

A continuación, presentaremos un resumen de los paquetes y/o funciones disponibles en R, que aplican la técnica STATIS-ACT. Con la finalidad de mostrar la forma en que trabajan cada uno de estos, presentaremos el proceso de aplicación de estos y los ilustraremos con un ejemplo general para observar las diferentes salidas que ofrecen.

Tabla 3. Paquetes en el lenguaje R que aplican las técnicas de la familia STATIS

Paquete (Fecha de última actualización)	Título	Descripción
MExPosition (Chin Fatt et al., 2015) CRAN 2013-06-10 (Ver 2.0.3)	Multi-table ExPosition	MExPosition es una Exposición de múltiples tablas e incluye la familia del método STATIS, como STATIS simple, DISTATIS, STATIS dual y ANISOSTATIS.
DistatisR (D. Beaton et al., 2013) CRAN 2013-07-10 (Ver 1.0.1)	DiSTATIS Three Way Metric Multidimension al Scaling	Implementa DiSTATIS y CovSTATIS.
MultbiplotR (Vicente-Villardón, 2015) GitHub 2015-01-13 (0.3.3.1)	MULTivariate Analysis Using BIPLoTs	Varias técnicas multivariadas desde una perspectiva biplot. Es la traducción a R del paquete preventivo MultBiplot desarrollado en Matlab.
Multigroup (Eslami et al., 2015) CRAN 2015-03-09 (Ver 0.4.4)	Methods for multigroup data analysis	Este paquete incluye varios métodos para estudiar datos de múltiples grupos, donde se mide el mismo conjunto de variables en diferentes grupos de individuos.
SirStatis (Villa-Vialaneix, 2016) GitHub 2016-01-07 (Ver 0.1)	STATIS, dual STATIS and SIR-STATIS for cubic data	Realiza análisis similares a STATIS de cubos de datos. El paquete es muy adecuado para explorar conjuntos de datos de curso de tiempo, eventualmente en relación con una variable numérica objetivo (enfoque SIR).
KTensorGraphs (Rodriguez-Rosa, Miguel) CRAN 2019-10-01 (ver 0.2)	COSTATIS, STATICO,	Proporciona una función llamada COTUCKER3 () (método Co-Inertia Analysis + Tucker3), incluye otras funciones como: PCA, BGA, COIA, PTA, TUCKER3, BGOIA, STATICO, COSTATIS.
DiDiSTATIS (Kriegsman, 2018b) GitHub 2018-01-18 (Ver 0.1.0)	Multi-way discriminant analysis for distance data	Este paquete ejecuta los métodos estadísticos multivariados, a saber, DiMDS, HiDiSTATIS y DiDiSTATIS.

Paquete (Fecha de última actualización)	Título	Descripción
ade4 (Bougeard & Dray, 2018; Daniel Chessel et al., 2004; Dray & Dufour, 2007; Dray et al., 2007; Dray, Dufour, & Thioulouse, 2004)	Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences	Herramientas para el análisis de datos multivariados, como análisis de componentes principales, análisis de correspondencia, análisis de coinertia, análisis de redundancia, análisis k tabla con STATIS, entre otros.
CRAN		
2018-08-30 (Ver 1.7-13) 2002-12-10 (Ver 1.0.1)		
dad (Boumaza et al., 2019)	Three-Way Data Analysis Through Densities	El paquete proporciona métodos funcionales (análisis de componentes principales, escalamiento multidimensional, análisis de conglomerados, análisis discriminante, ...) para tales densidades de probabilidad.
CRAN		
2019-07-02 (Ver 3.3.0)		
ClustBlock (Llobell, Vigneau, Cariou, et al., 2019)	Clustering of Datasets	Algoritmos jerárquicos y de partición de bloques de variables. El algoritmo de partición incluye una opción llamada clúster de ruido para separar bloques atípicos de variables.
CRAN		
2019-07-08 (Ver 2.0.0) 2019-03-06 (Ver 1.0.0)		

3.3.1. MExPosition

Este paquete realiza análisis exploratorio de tablas múltiples con técnicas de la familia STATIS utilizando la descomposición de valores singulares (SVD). Incluye STATIS, STATIS dual, X-STATIS(PTA), DISTATIS, ANISOSTATIS, k+1 STATIS, CANOSTATIS y COVSTATIS.

La función **mpSTATIS**, se combina todos los pasos del proceso de análisis: preprocesamiento, procesamiento, optimización y proyecciones complementarias que se calculan utilizando el método de análisis STATIS-ACT para múltiples tablas con filas comunes:

mpSTATIS:

- `mpSTATIS(data, column.design, make.column.design.nominal = TRUE, row.design = NULL, make.row.design.nominal = FALSE, statis.prepro.option = 'Plain_STATIS', DESIGN = NULL, make.design.nominal = TRUE, graphs = TRUE)`

cuyos argumentos son los siguientes,

Data:

- Matriz de datos.

column.design:

- Matriz con la identificación de las diferentes tablas.

make.column.design.nominal:

- Booleano. Si es TRUE (por defecto), la tabla es un vector que indica grupos (y tendrá un código ficticio). Si es FALSE, la tabla es una matriz codificada ficticia.

row.design:

- Matriz que identifica los diferentes grupos..

make.row.design.nominal:

- Booleano. Si es TRUE (por defecto), la tabla es un vector que indica grupos (y tendrá un código ficticio). Si es FALSE, la tabla es una matriz codificada ficticia

statis.prepro.option:

- Opción de cadena para los presentar STATIS. Las siguientes opciones están disponibles: 'Plain_STATIS', 'MFA', 'Sum_PCA', 'Plain_Multitable', 'Plain_ANISOSTATIS' y 'Customization'.

graphs:

- Booleano. Si es TRUE (por defecto) se muestran los gráficos.

DESIGN:

- una matriz de diseño para indicar si las filas pertenecen a grupos.

make.design.nominal:

- Booleano. Si es TRUE (por defecto), DESIGN es un vector que indica grupos (y tendrá un código ficticio). Si es FALSE, el diseño es una matriz codificada ficticia.

3.3.2. MultBiplotR

Este paquete contiene muchas técnicas de biplot y es una traducción a R del paquete MULBILOT programado en MATLAB. Incluye la técnica STATIS, junto con muchas ayudas gráficas, utilizando la función **StatisBiplot**. El procedimiento realiza la metodología STATIS-ACT. Cuando las variables son las mismas para todas las ocasiones, también se pueden trazar trayectorias para las variables.

StatisBiplot:

- `StatisBiplot(X, InitTransform = "Standardize columns", dimens = 2, SameVar = FALSE)`

cuyos argumentos son los siguientes,

X:

- Una lista que contiene varias tablas con filas comunes.

InitTransform:

- Transformación inicial de las matrices de datos.

dimens:

- Dimensión de la solución final.

SameVar:

- ¿Las variables son las mismas para todas las ocasiones?.

La estructura de la matriz requerida por la función Statis, es una matriz de dos vías. Para transformar el conjunto de matrices de datos, en una matriz de dos

vías, se utiliza la función *Three2TwoWay*, en donde para separar variables y ocasiones o estudios se utiliza el guion bajo (_).

Three2TwoWay:

- `Three2TwoWay(X, whatlines = 2)`

cuyos argumentos son los siguientes,

X:

- Una lista que contiene las múltiples tablas

whatlines:

- Concatenar por filas (1) o por columnas (2)

La función genera un objeto de clase `statis`.

Las gráficas del análisis STATIS se generan con la función `plot.statis`:

plot.statis:

- `plot(x, A1 = 1, A2 = 2, ...)`

cuyos argumentos son los siguientes,

x:

- objeto de clase `statis`

A1:

- primera dimensión del gráfico

A2:

- segunda dimensión del gráfico

y los parámetros adicionales que sean necesarios.

3.3.3. Multigroup

Este paquete incluye varios métodos para estudiar datos de múltiples grupos, donde el mismo conjunto de variables se mide en diferentes grupos de individuos. La técnica STATIS en este programa es el STATIS-dual.

Su funcionamiento se da a través de la función **DSTATIS**:

DSTATIS:

- `DSTATIS(Data, Group, ncomp = NULL, Scale = FALSE, graph = FALSE)`

cuyos argumentos son los siguientes,

Data:

- Una matriz numérica o dataframe.

Group:

- Un vector de factores asociados con la estructura de grupos

ncomp:

- Número de componentes, si es NULL el número de componentes es igual a 2.

Scale:

- Variables de escala, por defecto es FALSE. Por defecto, los datos se centran dentro de los grupos.

graph:

- se debe trazar la carga y el componente

Los resultados generados son: datos originales (**Data**), los datos centrados concatenados (**Con.Data**), los datos centrados en el grupo (**split.Data**), grupos como vectores de factores (**Group**), la matriz de coeficientes RV (**RV**), vector de pesos (**weights**), matriz de varianza-covarianza del compromiso (**compromise.matrix**), matriz de cargas (**loadings.common**).

3.3.4. ade4

El paquete **ade4** para el entorno estadístico R propone una gran cantidad de métodos multivariados. Fue inicialmente desarrollado para ser utilizado en la identificación y comprensión las estructuras de las comunidades ecológicas. Su implementación sigue la tradición de la escuela francesa de "Analyze des Données" y se basa en el uso del diagrama de dualidad. Es uno de los más completos con relación a las técnicas de la familia STATIS. Tiene disponibles las técnicas STATIS, STATIS dual, Análisis Triádico Parcial, COSTATIS, STATICO, A continuación, se presentan las clases y funciones principales.

El análisis STATIS se realiza a través de la función **statis**, esta función también se puede utilizar para realizar el STATIS dual, configurando la matriz correspondiente.

statis:

- `statis(X, scannf = TRUE, nf = 3, tol = 1e-07)`

Con los argumentos,

X:

- un objeto de clase *ktab*.

scannf:

- Lógico. Indica si se debe preguntar el número de ejes a retener para el compromiso.

nf:

- Si `scannf` es FALSE, será un número entero que indica el número de ejes a retener para el compromiso.

tol:

- Un umbral de tolerancia para probar si la matriz de distancia es Euclídea: un valor propio se considera positivo si es mayor que $-tol * \alpha_1$ donde α_1 es el valor propio más grande.

Las gráficas correspondientes se obtienen con la función **plot**

plot:

- `plot(x, xax = 1, yax = 2, option = 1:4, ...)`

$con a_1$

x:

- un objeto de clase *statis*

xax, yax:

- es el número del eje que se representa en el eje x y y respectivamente.

option:

- un número entero entre 1 y 4; de lo contrario, se muestran los 4 componentes del gráfico.

La función **statis** proporciona un objeto de clase *statis*, la cual contiene: la matriz de coeficiente $RV(RV)$, el vector de valores propios($RV.eig$), un vector de pesos de cada matriz ($RV.tabw$), matrices con coordenadas de filas($C.II$) y de columnas($C.Co$) respectivamente, la matriz con los vectores principales para cada tabla($C.T4$), la matriz con los puntajes(scores)($RV.coo$), un vector con los nombres de las diferentes matrices($tab.names$), número de ejes retenidos($C.nf$) y el rango en el análisis($C.rank$).

Para utilizar la función **statis**, es necesario crear el objeto de clase *ktab*, y previa a esto se debe realizar un análisis de componentes principales normalizado, sobre cada una de las matrices, utilizando la función **withinpca**. El objeto *ktab* se obtiene con la función **ktab.within**.

3.3.5. SirSTATIS

Es paquete realizar análisis similares al STATIS. El paquete es muy adecuado para explorar conjuntos de datos de cursos de tiempo, eventualmente en relación con una variable numérica objetivo (**enfoque SIR: sliced inverse regression**). Dentro de las técnicas STATIS.

Su funcionamiento se da a través de la función **statis**:

statis:

- `statis <- function(data, method = c("sir", "classic", "dual"), y = NULL, H=NULL, scale=TRUE)`

cuyos argumentos son los siguientes,

data:

- La lista de tablas para estudiar.

method:

- Método a utilizar ("sir", "classic" o "dual"). Por defecto es "sir".

y:

- Una variable numérica de interés (obligatorio para "sir").

H:

- Valor entero. Número de secciones para cortar en "y" (obligatorio para "sir").

scale:

- Lógico. ¿Se debe calcular la interestructura con el coseno?. (TRUE) predeterminado, o (FALSE) producto punto.

Los resultados generados son: objetos representativos según el método elegido (**objects**), matriz utilizada para la descomposición propia (**corr**), valores propios de la descomposición del coeficiente RV o matriz de coeficientes de correlación (**values**), vectores propios de la descomposición de los coeficientes RV o matriz de coeficientes de correlación(**vectors**), matriz consenso entre estructuras(**consensus**), pesos calculados para cada tabla(**tweight**).

3.3.6. dad

En el paquete '**dad**', los datos consisten en un conjunto de variables medidas en varios grupos de individuos. A cada grupo se asocia una función de densidad de

probabilidad estimada. El paquete proporciona herramientas para crear o administrar dichos datos y métodos funcionales (análisis de componentes principales, escalamiento multidimensional, análisis de conglomerados, análisis discriminante, etc.) para tales densidades de probabilidad. Dentro de las funciones tiene `dstatis.inter` que realiza la primera etapa (entre estructuras) del método dual STATIS para describir una carpeta de datos, que consta de H grupos de individuos en los que se observan p variables. Devuelve un objeto de clase `dstatis`. La función `dstatis.inter`, calculan las matrices de covarianza (si `data.scale` es FALSE) o las matrices de correlación por grupo (si es TRUE). Luego se calcula la matriz \mathbf{W} de los productos escalares entre estas matrices de covarianza.

Para aplicar el método STATIS, usa la función `DSTATIS` del paquete `multigrupo`. El objeto de clase `dstatis` proporciona los valores propios y el porcentaje de inercia; las contribuciones de los primeros componentes principales de los `nb.factors`; las calidades en los primeros `nb.factores` factores principales; puntajes de los primeros `nb.factors` puntajes de la descomposición espectral de \mathbf{W} ; vector de las normas L_2 de las densidades; lista de medias, matrices de covarianza; matrices de correlación; coeficientes de asimetría; coeficientes de curtosis.

`dstatis.inter`:

- `dstatis.inter(xf, normed = TRUE, centered = FALSE, data.scaled = FALSE, nb.factors = 3, nb.values = 10, sub.title = "", plot.eigen = TRUE, plot.score = FALSE, nscore = 1:3, group.name = "group", filename = NULL)`

cuyos argumentos son los siguientes,

`xf`:

- objeto de la clase `folder`. Sus elementos son dataframes con p columnas numéricas. Si hay columnas no numéricas, genera un error. El elemento t^{th} ($t = 1, \dots, T$) coincide con el grupo t^{th} .

`normed`:

- logico. Si es TRUE (predeterminado), los productos escalares están normalizados.

`centered`:

- logico. Si es TRUE (el valor predeterminado es FALSE), los productos escalares están centrados.

data.scaled:

- lógico. Si es **TRUE**, los datos de cada grupo están centrados y escalados. El análisis se realiza luego con las matrices de correlación. Si es **FALSE** (predeterminado), el análisis se realiza con las matrices de covarianza..

nb.factors:

- numérico. Número de puntuaciones principales devueltas (por defecto **nb.factors** = 3).

nb.values:

- numérico. Número de valores propios devueltos (predeterminados **nb.values**= 10)

sub.title:

- cadena. Si se proporciona, el subtítulo para los gráficos.

plot.eigen:

- lógico. Si es **TRUE** (predeterminado), se traza la gráfica de barras de los valores propios.

plot.score:

- lógico. Si es **TRUE**, se trazan los gráficos de las puntuaciones principales. Se abre un nuevo dispositivo gráfico para cada par de puntajes principales definidos por el argumento **nscore**.

nscore:

- vector numérico. Si **plot.score** = **TRUE**, se trazan los números de las puntuaciones principales. Por defecto **nscore** = 1:3. Sus componentes no pueden ser mayores que **nb.factors**.

group.name:

- cadena. Nombre de la variable de agrupación. Valor predeterminado: **groupname** = "group".

filename:

- cadena. Nombre del archivo en el que se guardan los resultados. Por defecto (**filename** = **NULL**) los resultados no se guardan.

3.3.7. CLUSTBLOCK

Es paquete realiza algoritmos jerárquicos y de partición de bloques de variables. El algoritmo de partición incluye una opción llamada cluster de ruido para separar bloques atípicos de variables. El método CLUSTATIS (para bloques cuantitativos) es el núcleo de este paquete. Realiza agrupación jerárquica de bloques cuantitativos seguida de un algoritmo de partición (consolidación). Cada grupo de bloques está asociado con un compromiso calculado por el método STATIS. Además, se puede configurar un grupo de ruido.

El análisis STATIS en diferentes bloques de variables cuantitativas se realiza con la función **statis**, también genera salidas complementarias.

statis:

- `statis(Data,Blocks,NameBlocks=NULL,Graph_obj=TRUE,Graph_weights=TRUE, scale=FALSE)`

cuyos argumentos son los siguientes,

Data:

- matriz de datos o dataframe. Corresponde a todos los bloques de variables fusionadas horizontalmente.

Blocks:

- vector numérico. El número de variables de cada bloque. La suma debe ser igual al número de columnas de datos.

NameBlocks

- vector de cadena. Nombre de cada bloque. La longitud debe ser igual a la longitud del vector de bloques. Si es NULL, los nombres son B_1, \dots, B_m . Predeterminado: NULL.

Graph_obj:

- lógico. Mostrar la representación gráfica de los objetos?
Predeterminado: TRUE

Graph_weights:

- lógico. ¿trazar la gráfica de barras de los pesos?
Predeterminado: TRUE

scale:

- lógico. ¿Se deben escalar los datos de variables?
Predeterminado: FALSE

Las salidas generadas por esta función son: la matriz con el coeficiente RV entre bloques de variables(**RV**), una matriz compromiso de los bloques (similar a un promedio ponderado) (**compromise**), los pesos asociados con los bloques para construir el compromiso (**weights**), el primer valor propio de la matriz RV (**lambda**), el error para el criterio STATIS (**overall error**), el error de la configuración (criterio STATIS)(**error_by_conf**), el coeficiente RV de cada bloque con el compromiso(**rv_with_compromise**), homogeneidad de los bloques (en porcentaje) (**homogeneity**), las coordenadas de cada objeto (**coord**), Los valores propios de la descomposición de SVD (**eigenvalues**), el porcentaje de varianza total explicado por cada eje (**inertia**), el error para objeto (criterio STATIS)(**error_by_obj**), los factores escalares de cada bloque (**scalefactors**), la proyección de cada objeto de cada configuración en los ejes (presentado por configuración)(**proj_config**) y la proyección de cada objeto de cada configuración en los ejes(presentado por objeto)(**proj_objects**).

En la [tabla 4](#) se muestra una comparativa de los resultados generados por los paquetes en R que realizan Statis o Statis-dual.

Tabla 4. Comparativa de los resultados que arrojan las librerías que realizan Stasis* y/o Stasis-dua**

Paquete	MExposition (*) (**)	MultBiplotR (*)	Multigroup (*)	ade4 (*)	SirSTATIS (*)(**)	Dad (**)	Clustblock (*)
Matrices pre-procesadas (X_k)	✓	✓	✓				✓
Productos escalares (S_k)	✓	✓					
Matriz de coeficientes RV	✓	✓	✓	✓	✓	✓	✓
Matriz de vectores propios	✓	✓		✓	✓	✓	✓
Matriz de valores propios	✓	✓		✓	✓	✓	✓
Vector de Pesos	✓	✓	✓	✓	✓		✓
Matriz compromiso	✓	✓	✓	✓	✓		✓
Matriz de "scores" del compromiso	✓	✓	✓	✓	✓		✓
Gráfico de la interestructura	✓	✓		✓		✓	
Gráfico del compromiso	✓	✓	✓	✓			
Gráfico de la Intraestructura	✓	✓		✓			

CAPÍTULO IV: SPARSE STATIS-DUAL

4.1. INTRODUCCIÓN

El siglo XXI ha generado una revolución hacia el uso de datos, entendiendo la relevancia de los datos para la resolución de problemas, la sociedad se ha percatado de que los datos pueden hablar por sí mismos, siempre que se disponga de una gran cantidad de información, y se realicen análisis de patrones de comportamiento de los mismos. A tal efecto, se encontró la utilidad de gran cantidad de datos recopilados de varias fuentes y en diferentes formatos, lo que conocemos como “Big Data”. El Big Data, no es más que un gran volumen de datos, tanto estructurados como no estructurados. Normalmente, estos datos siempre estaban disponibles; no obstante, no se entendía mucho su importancia, ya que no era posible recopilar y analizar una enorme cantidad de datos. Con el desarrollo de herramientas basadas en la tecnología de la información y comunicación (TIC's) se ha podido recopilar, procesar y analizar los datos.

En muchos casos la complejidad de estos datos, como por ejemplo genomas, hace necesario utilizar técnicas con la capacidad de simplificar la información original y proporcionar significado a los resultados obtenidos. Recientemente algunos métodos estadísticos se han adaptado para procesar Big Data como modelos de regresión lineal, métodos de agrupamiento, entre otros.

En este sentido los métodos de análisis multivariante, son esenciales para el análisis de estos datos, ya que engloba los métodos y técnicas estadísticas que permiten estudiar y tratar, en su conjunto, gran cantidad de variables que han sido medidas en gran cantidad de individuos. Estos se basan en la descomposición matricial que exploran grandes cantidades de información, valiéndose de su alta dimensionalidad.

La técnica más conocida es el Análisis de Componentes Principales-ACP (Hotelling, 1936; Pearson, 1901), que se aplica a través de la Descomposición en Valores Singulares (DVS) de Eckart and Young (1936).

Dentro de las técnicas de análisis multivariante, se dispone de los métodos STATIS que, desde el punto de vista algebraico, son una generalización del

análisis de componentes principales, y por consiguiente se basan en la DVS de una matriz de datos.

Vamos a presentar una variante del modelo STATIS-dual, la propuesta, corresponde a la regularización Elastic-net, a la que denominaremos Sparse Stasis-dual. Se presentan los resultados de la aplicación de la propuesta sobre datos simulados y reales.

4.2. SPARSE PCA

Con la finalidad de mejorar la interpretación de los resultados del análisis de componentes principales se han desarrollado muchas propuestas, como por ejemplo técnicas de rotación y restricciones sobre las cargas factoriales del ACP. [Hausman \(1982\)](#), propuso restringir el valor que se le puede asignar a las cargas de las componentes principales a un conjunto de números enteros $\{-1,0,1\}$ con la finalidad de obtener componentes simplificados. Luego, [Vines \(2000\)](#) defiende la idea de [Hausman \(1982\)](#) y propone utilizar números enteros arbitrarios. [McCabe \(1984\)](#) idea otro método consistente en seleccionar un subconjunto de variables, que denomina variables principales, partiendo de un criterio de optimización, prescindiendo del ACP. [Cadima & Jolliffe \(1995\)](#) proponen una solución al problema presentando el método “umbral simple”, que se fundamenta en convertir las cargas factoriales, con valores absolutos menores que cierto umbral, en cargas nulas. Para simplificar la estructura de las CPs y facilitar su interpretación, habitualmente, se han utilizado las técnicas de rotación ([Jolliffe,1995](#)), sin embargo, la reducción de la dimensionalidad de los datos no siempre es suficiente para facilitar la interpretación de las CPs. Una manera, *ad hoc*, se fundamenta en utilizar las técnicas de regularización que, aunque requieren un parámetro de restricción para inducir vectores de proyección con cargas modificadas (nulas o cercanas a cero) y controlar el peso de las cargas, proporcionan una importante mejora para la interpretación de los resultados. Seguidamente, [Tibshirani \(1996\)](#) formula el método *LASSO (Least Absolute Shrinkage and Selection Operator)*, en el que combina un modelo de regresión con un procedimiento de contracción de algunos parámetros hacia cero

imponiendo una penalización sobre los coeficientes de regresión. Posteriormente, [Jolliffe & Uddin \(2000\)](#) introducen una solución que modifica el enfoque tradicional de CPs en dos etapas (ACP + rotación). Desarrollan la técnica de componentes simplificados ScoT (Simplified component Technique), en la cual las CPs originales seguidas de la rotación VARIMAX se combinan en un solo paso para inducir cargas con poca densidad, manteniendo la disminución de la proporción de varianza explicada. [Jolliffe, Trendafilov, & Uddin \(2003\)](#), proponen el algoritmo SCoTLASS (Simplified Component Technique subject to LASSO), ya que las cargas obtenidas mediante ScoT proporcionaban valores pequeños distintos de cero, pero no nulos. Esta nueva propuesta se basa en imponer una restricción de forma tal que algunas cargas se hacen completamente nulas, pero sacrificando la varianza. Es así que [Zou et al. \(2006\)](#) presentan un algoritmo de regresión penalizado usando la técnica “*Elastic Net*” (llamado *Sparse PCA*) ([Zou & Hastie, 2005](#)) que resolvieron eficazmente utilizando la regresión de ángulo mínimo ([Efron, Hastie, Johnstone, & Tibshirani, 2004](#)). Otras propuestas fueron las de [Moghaddam et al. \(2006\)](#), que sujeto a restricciones de cardinalidad (número de cargas cero, por componente), construyen un algoritmo para componentes *Sparse*; y la de [d’Aspremont et al. \(2007\)](#) que explican la restricción de cardinalidad en base a programación semidefinida.

[Shen & Huang \(2008\)](#) enlazan el ACP con la DVS de los datos y crean el CPs *Sparse* mediante penalizaciones de regularización (sPCA-rSVD). Seguidamente, [Witten, Tibshirani, & Hastie \(2009\)](#) fusionan el enfoque de aproximación matricial de bajo rango de [Shen & Huang \(2008\)](#) con el criterio de máxima varianza de [Jolliffe et al. \(2003\)](#) y el método *sparse PCA* de ([Zou et al., 2006](#)), para dar una solución general al problema de *sparse PCA*. En este mismo sentido, [Farcomeni \(2009\)](#) propone maximizar la varianza explicada, penalizada por la cardinalidad de las CPs *Sparse*. Al mismo tiempo, [Vichi & Saporta \(2009\)](#) exponen una variación al PCA, que permite reconocer componentes de máxima varianza y garantiza que cada variable contribuya solo a uno de los ejes factoriales. Posteriormente, [Qi, Luo, & Zhao \(2013\)](#) plantean una forma relativamente simple para afrontar el problema, mediante una extensión del ACP clásico. Construyen componentes *sparse* sustituyendo la norma L_2 , en

problemas de valores propios tradicionales, con una nueva norma que es la combinación de las normas L_1 y L_2 .

Desde un enfoque bayesiano, en el Sparse PCA, la forma estándar del ACP es reemplazada por su versión probabilística (Trendafilov, 2014) manteniendo la misma idea que los métodos precedentes.

Una profunda revisión de los principales enfoques para la interpretación del PCA, es realizada por Trendafilov (2014). De igual manera, Zhang, Xu, Yang, Li, & Zhang (2015) proporcionan un panorama general de los algoritmos *Sparse* desde la perspectiva de la teoría de la optimización matemática.

4.3. SPARSE STATIS-DUAL

Dentro de la familia de técnicas STATIS, no se han encontrado propuestas para penalizar o contraer las cargas de las componentes principales, que permitan la interpretación de la información que aportan los datos de alta dimensionalidad. Sin embargo, dentro de los métodos de análisis de tablas de tres vías se han propuesto restricciones basadas en las teorías Ridge (Hoerl & Kennard, 1970) y LASSO (Tibshirani, 1996).

Así tenemos el modelo CP-Ridge (Giordani & Rocci, 2013a), resuelve el problema de la degeneración al admitir un nivel máximo general de colinealidad entre los componentes. La peculiaridad de CP-Ridge consiste en relajar las restricciones de ortogonalidad admitiendo un nivel máximo general de colinealidad entre los componentes. Esto se logra considerando la descomposición \mathbf{QR} de una de las matrices componentes, sin pérdida de generalidad $\mathbf{A} = \mathbf{QR}$, y limitando la magnitud de los elementos fuera de la diagonal de \mathbf{R} una vez que las columnas de \mathbf{A} se reescalan para obtener elementos en la diagonal de \mathbf{R} iguales a 1. Para este propósito, se agrega un término de regularización en \mathbf{R} a la función de pérdida para minimizarla. Tal término de regularización proviene del análisis de regresión Ridge. Esto hace que CP-Ridge sea notablemente diferente de las correcciones utilizadas para la degeneración. De acuerdo al autor, CP-Ridge no debe considerarse un modelo

alternativo con respecto al CP estándar. Es un método capaz de producir una solución CP que no sufre degeneración. Por lo tanto, CP-Ridge representa un procedimiento de estimación para el modelo CP.

Otra de las propuestas de este mismo autor es Candecomp / Parafac con restricción LASSO (CP LASSO) (Giordani & Rocci, 2013b), una variante del modelo de CP que relajan las restricciones de ortogonalidad en una matriz de componentes, teniendo en cuenta que pueden ser demasiado restrictivas para la matriz de datos en observación. Esto se hace descomponiendo la matriz de componentes en la forma \mathbf{QR} , digamos $\mathbf{A} = \mathbf{QR}$ donde \mathbf{Q} es en forma de columna ortonormal y \mathbf{R} triangular superior con $\text{diag}(\mathbf{R}) = 1$, y restringe uno o más elementos de \mathbf{R} igual a cero. Se aplica para evitar la llamada “degeneración” (solución con componentes divergentes y no interpretables), estimulando la solución CP para que sea posiblemente ortogonal. CP-Lasso reduce una de las matrices de componentes hacia la ortogonalidad en columnas. Con esto, permite relajar las restricciones de ortogonalidad evitando la degeneración y, debido a la geometría de Lasso, a menudo produce estimaciones cero para los elementos de \mathbf{R} , es decir, componentes ortogonales (parciales).

Igualmente, el autor señala que el método CP-Ridge debe verse como una herramienta alternativa para resolver la degeneración de CP con respecto a CP-Lasso, donde la restricción de Lasso se impone a \mathbf{R} . Ridge y Lasso son procedimientos de reducción, pero solo el último tiene la propiedad de selección de variables, es decir, la tendencia a producir coeficientes cero. Lo que implica ortogonalidad (parcial) entre pares de componentes. La ortogonalidad (parcial) de las soluciones y la eficiencia de los algoritmos representan los temas clave para poder elegir entre CP-Lasso y CP-Ridge. El autor señala que, en el caso de deficiencias del CP, CP-Ridge y CP-Lasso son alternativas valiosas. Si se cree que algunos de los componentes subyacentes son ortogonales (parciales) hasta cierto punto y, por lo tanto, está interesado en obtener pares de componentes ortogonales (parciales), se debe considerar CP-Lasso, de lo contrario, CP-Ridge debería ser la opción a utilizar.

Dentro de los métodos de regularización, nos enfocaremos en Elastic Net (Zou et al., 2006). Como señalamos anteriormente, los métodos de restricción han

sido ampliamente cubiertos en la literatura cuando se aplican a modelos clásicos de regresión lineal y componentes principales, y en algunas técnicas de análisis de datos de tres vías. Sin embargo, no han sido estudiados en el contexto de las técnicas STATIS.

En esta línea de trabajo, este documento propone nuevas alternativas STATIS que consisten en adaptar restricciones para contraer y/o producir cargas nulas en las componentes, con base en la teoría de regularización de Elastic Net. El procedimiento para la implementación en el STATIS-dual, se presenta en detalle, teniendo en cuenta las particularidades o características propias del modelo.

En cada caso se demuestra el funcionamiento de los algoritmos mediante la creación en lenguaje R del paquete “*Sparse Stasis-dual*”, diseñado para dar soporte a la nueva metodología planteada. La implementación del paquete se realiza con los datos Environmental Development Index (EPI).

4.3.1. Stasis-dual restringido via elastic net

Zou & Hastie (2005) formulan el método *Elastic Net*, que integra los métodos de regularización *Ridge* y *LASSO*, dentro del análisis de regresión. Se trata entonces de un método de regularización que penaliza el tamaño de los coeficientes de regresión en base a las normas L_1 y L_2 .

Uno de los componentes más importantes del método *Elastic Net*, son los coeficientes estimados $\hat{\beta}^{elasticnet}$, los cuales son los valores que minimizan:

$$\hat{\beta}^{elasticnet} = \arg \min_{\beta} \|y_i - X\beta\|^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (155)$$

donde $\lambda_1 > 0$ y $\lambda_2 > 0$ son parámetros de complejidad.

Así tenemos que el término $\lambda_1 \sum_{j=1}^p |\beta_j|$ apunta a soluciones *sparse*. Mientras que el término $\lambda_2 \sum_{j=1}^p \beta_j^2$ apunta a que predictores altamente correlacionados consigan coeficientes estimados semejantes.

De manera similar se considera,

$$\hat{\beta}^{elasticnet} = \arg \min \|y_i - X\beta\|^2 \quad (156)$$

condicionado a $\sum_{j=1}^p \beta_j^2 \leq \lambda$ y $\sum_{j=1}^p |\beta_j| \leq \lambda$

Sobre la base de esta condición, Zou & Hastie (2005) plantean construir los nuevos β por medio de la siguiente fórmula:

$$V^{soft} = \text{sign}(V) \frac{(|V - \lambda_1|)_+}{1 + \lambda_2} \quad (157)$$

en dónde se integran las normas L_1 y L_2 aplicando el operador soft-thresholding.

La regularización Elastic Net se puede poner en práctica en el STATIS-dual, ajustando LASSO y Ridge para derivar cargas modificadas. Para ello se utiliza el modelo:

$$X = Q\Lambda^{\frac{1}{2}} + E \quad (158)$$

A partir del método de regularización *Elastic Net* se derivan cargas modificadas para el STATIS-dual, de la forma:

$$V_{elasticnet} = \arg \min \|X - Q\Lambda^{1/2}\|^2 + \lambda_2 \sum_{j=1}^p V_j^2 + \lambda_1 \sum_{j=1}^p |V_j| \quad (159)$$

en dónde λ_1 es el parámetro de penalización *LASSO* para propiciar esparsidad y λ_2 es el parámetro de regularización para reducir las cargas

Ahora, tomando en cuenta los primeros k ejes factoriales, se definen las matrices:

$$A_{p \times k} = [\alpha_1, \alpha_2, \dots, \alpha_k] \text{ y } B_{p \times k} = [\beta_1, \beta_2, \dots, \beta_k].$$

Para algún $\lambda_2 > 0$, sea:

$$(\hat{A}, \hat{B}) = \arg \min \sum_{i=1}^n \|x_i - AB^T x_i\|^2 + \lambda_2 \sum_{j=1}^k \|\beta_j\|^2 + \lambda_{1,j} \sum_{j=1}^k \|\beta_j\|_1 \quad (160)$$

condicionado a $A^T A = I_{k \times k}$. Entonces $\hat{\beta}_j \propto V_j$ para $j = 1, 2, \dots, k$

Considerando que

$$\begin{aligned}
\sum_{i=1}^n \|x_i - AB^T x_i\|^2 &= \|X - XBA^T\|^2 & (161) \\
&= \|XA\|^2 + \|XA - XB\|^2 \\
&= \|XA\|^2 + \sum_{j=1}^k \|X\alpha_j - X\beta_j\|^2
\end{aligned}$$

La solución se obtiene alternando la optimización sobre A y B usando el algoritmo LARS-EN (Zou & Hastie, 2005).

Para A fija se tiene:

$$\begin{aligned}
\hat{\beta}_j &= \operatorname{argmin} \|X\alpha_j - X\beta_j\|^2 + \lambda_2 \|\beta_j\|^2 + \lambda_{1,j} \|\beta_j\|_1 & (162) \\
&= (\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) + \lambda_2 \|\beta_j\|^2 + \lambda_{1,j} \|\beta_j\|_1
\end{aligned}$$

en dónde cada $\hat{\beta}_j$ es un estimador Elastic Net.

Para B fija: no se toma en cuenta la parte de penalización y se minimiza:

$$\operatorname{argmin} \sum_{i=1}^n \|x_i - AB^T x_i\|^2 = \|X - XBA^T\|^2 \quad (163)$$

condicionado a $A^T A = I_{K \times K}$

Lo que lleva a un problema de Procrustes, y la solución es suministrada por la DVS, $(X^T X)B = UDV^T$ y se determina $\hat{A} = UV^T$.

En resumen, Elastic Net contempla lo siguiente:

- Usa las normas L_1 y L_2 .
- Selección de variables
- Penaliza las cargas.
- Contrae algunas cargas hacia cero, y otras las anula.

El método de proyección variable (Erichson, Zheng, & Manohar (2018) se plantea como otra solución al problema de optimización.

No existen métodos evidentes y determinados para ajustar los parámetros λ_2 y λ_1 . Se propone ensayar varias combinaciones y optar por la que provea equilibrio entre la varianza explicada y la esparsidad, dando preponderancia a la varianza.

Los pasos para la implementación del método de regularización Elastic Net en el STATIS-dual se detallan en el siguiente algoritmo.

Algoritmo 1: Algoritmo SPARSE STATIS-dual usando la regularización Elastic Net

1. Considere una matriz de datos $n \times p$.
2. Se fija un valor de tolerancia (1×10^{-5}).
3. Se transforman los datos (centrar o estandarizar).
4. Se obtienen las matrices de productos cruzados S_h .
5. Se obtiene la matriz de cosenos entre estudios C
6. Se realiza un PCA sobre C
7. Se obtiene la matriz compromiso S
8. Se realiza la descomposición en DVS de la matriz compromiso.
9. Se toma A como las cargas de las primeras k componentes $V[, 1:k]$.
10. Se calcula β_j mediante:

$$\beta_j = (\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) + \lambda_2 \|\beta_j\|^2 + \lambda_{1,j} \|\beta_j\|_1$$

11. Se actualiza A mediante la DVS de $X^T X \beta$:

$$X^T X \beta = U D V^T \rightarrow A = U V^T$$

12. Se actualiza la diferencia entre A y B

$$dif_{AB} = \frac{1}{p} \sum_{i=1}^p \frac{1}{|\beta_i|^2 |\alpha_i|^2} \sum_{j=1}^m \beta_{ij} - \alpha_{ij}$$

13. Se repiten los pasos 4,5 y 6 hasta que $dif_{AB} < tolerancia$

14. Se normalizan las columnas $\hat{V}_j^{EN} = \frac{\beta_j}{\|\beta_j\|}, j = 1, \dots, k$

15. Se obtienen las cargas restringidas para la proyección de las variables sobre el compromiso.

16. Se grafica el *Elastic Net* STATIS-dual obtenido mediante los pasos anteriores.
-

En la [Figura 37](#) se presentan los pasos que describen la aplicación de los métodos de regularización sobre el STATIS-dual, que llevan a la obtención de una matriz de cargas modificada.

Se ha diseñado una aplicación en el lenguaje R, para obtener la representación *Elastic Net* STATIS-dual. Esta aplicación forma parte del paquete denominado "SparseSTATISdual, el cual se detalla en la sección 4.4.

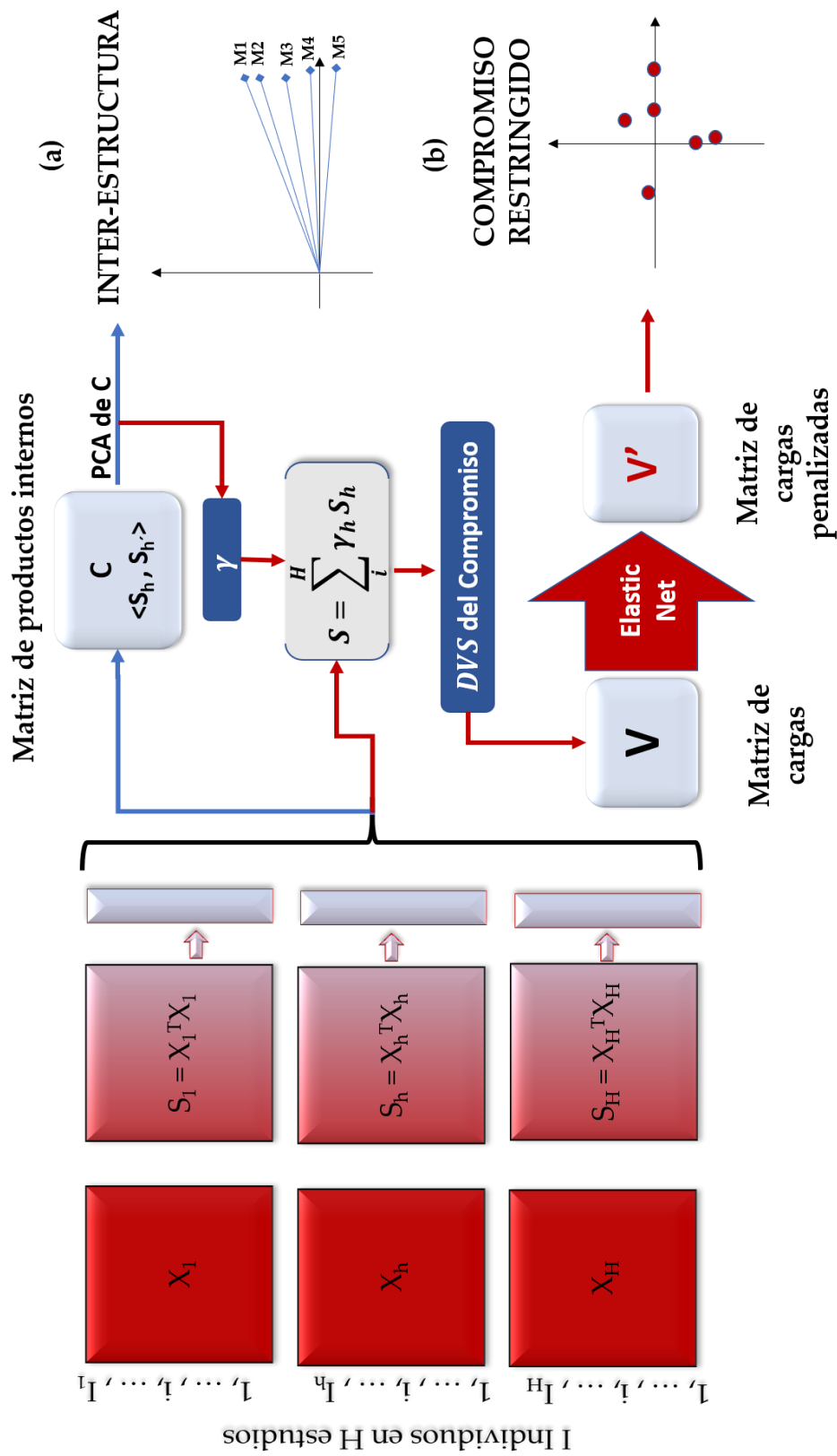


Figura 37. Diagrama de los pasos del Sparse STATIS-dual

4.4. ALGORITMO EN R: SPARSE STATIS-DUAL

En el interés de proporcionar una herramienta que implemente los algoritmos descritos en las secciones anteriores, se ha creado un paquete en el lenguaje R.

Este paquete, denominado `SparseStatisdual`, permite utilizar el algoritmo que desea con sus propios datos y generar un gráfico con la representación de interés; además, permite obtener resultados numéricos.

Este algoritmo se utiliza para realizar la selección automática de variables y contracción simultáneamente. Con este método, el modelo obtenido es más sencillo y más interpretable.

Su funcionamiento se da a través de las funciones **`Elastic_net_Statis`**, **`Plot_Interstructure`**, **`plot_Compromise`**, y **`Plot_Intrastructure`**, que hemos desarrollado para la implementación del paquete **`SparseStatisdual`** y se encuentra disponible en el repositorio GitHub con el link: github.com/CCRM07/SparseSTATISdual.

`Elastic_net_Statis`:

- `Elastic_net_Statis(Y, Transform.Data = 'center', Alpha = 0.001, Lambda = 0.01)`

cuyos argumentos son los siguientes,

Y:

- La lista de tablas para estudiar.

Transform.Data:

- Método de pre-procesamiento a utilizar ('center' o 'scale'). Por defecto es 'center'

Alpha:

- Parámetro de ajuste de la penalización LASSO. Valores más altos conducir a componentes más dispersos

Lambda:

- Parámetro de ajuste de la contracción Ridge

Los resultados generados son: objetos:

- `statis_$X_Transformed` → (X_tr) → matriz pre-procesada
- `statis_$Scalar_Products` → (S) → matrices de productos escalares
- `statis_$Cosine` → (Cosine) → matriz de cosenos
- `statis_$Interstructure` → (Interstructure) → interestructura
- `statis_$Weight` → (alpha) → pesos para cada tabla
- `statis_$Compromise` → (SC) → matriz compromiso
- `statis_$Quality` → (Q) → calidad de representación
- `statis_$Scores` → (F_) → factores de cargas restringidos
- `statis_$Projection_Matrix` → (G) → matriz de proyecciones

Plot_Interstructure:

- `Plot_Interstructure(X, color = "red") + ggtitle(label = "Interestructura")`

cuyos argumentos son los siguientes,

X:

- `X$Interstructure`ear, matriz de interestructuras.

El resultado generado es un gráfico que proyecta la interestructura.

plot_Compromise:

- `plot_Compromise(X) + ggtitle(label = "Compromiso")`

cuyos argumentos son los siguientes,

X:

- `X$scores`, factores de carga restringidos.

El resultado generado es un gráfico que proyecta las variables sobre el compromiso restringido.

El paquete desarrollado en R, también incluye las funciones `Statis` y `Statis-dual`

Statis:

- `Statis(Y, Transform.Data = 'center')`

cuyos argumentos son los siguientes,

Y:

- La lista de tablas para estudiar.

Transform.Data:

- Método de pre-procesamiento a utilizar ('center' o 'scale'). Por defecto es 'center'

Los resultados generados son: objetos

- `statis$X_Transformed` → (`X_tr`) → matriz pre-procesada
- `statis$Scalar_Products` → (`S`) → matrices de productos escalares
- `statis$Cosine` → (`Cosine`) → matriz de cosenos

- `statis$Interstructure` →(Interstructure)→interestructura
- `statis$Weight` →(alpha) → pesos para cada tabla
- `statis$Compromise` →(SC) → matriz compromiso
- `statis$Quality` →(Q) → calidad de representación
- `statis$Scores` → (F_) → factores de cargas
- `statis$Projection_Matrix`→ (G) → matriz de proyecciones
- `statis$Coordinates`→Coordinatesrepresentativos)→coordenadas de la infraestructura.

Statis-dual:

- `Statis_d(Y, Transform.Data = 'center')`

cuyos argumentos son los siguientes,

Y:

- La lista de tablas para estudiar.

Transform.Data:

- Método de pre-procesamiento a utilizar ('center' o 'scale'). Por defecto es 'center'

Los resultados generados son: objetos

- `statisd$X_Transformed` →(X_tr) → matriz pre-procesada
- `statisd$Scalar_Products`→(S) → matrices de productos escalares
- `statisd$Cosine` →(Cosine) → matriz de cosenos
- `statisd$Interstructure` →(Interstructure)→interestructura
- `statisd$Weight` →(alpha) → pesos para cada tabla
- `statisd$Compromise` →(SC) → matriz compromiso
- `statisd$Quality` →(Q) → calidad de representación
- `statisd$Scores` → (F_) → factores de cargas
- `statisd$Projection_Matrix`→ (G) → matriz de proyecciones

-
- `statisd$Coordinates→Coordinatesrepresentativos)→coordenadas` de la infraestructura

Igualmente, estas dos funciones permiten proyectar la infraestructura, el compromiso y la infraestructura, utilizando las funciones **Plot_Interstructure**, **plot_Compromise**, y **Plot_Infrastructure**.

4.5. ANÁLISIS DE DATOS REALES: APLICACIÓN DE SPARSE STATIS-DUAL

4.5.1. Base de datos

A continuación, aplicaremos la técnica propuesta, Sparse STATIS-dual, para ello utilizaremos la base de datos compuesta por el Índice de Desempeño Ambiental (Environmental Performance Index-EPI), (Wendling et al., 2020).

El EPI proporciona un indicador a escala nacional de cómo los países están cerca de los objetivos establecidos de la política ambiental. Se compone de dos objetivos ambientales generales, estos son: (i) reducir las tensiones ambientales para la salud humana, y (ii) promover la vitalidad del ecosistema y la gestión de recursos naturales.

Para nuestro análisis, hemos agregado dos variables adicionales. El EPI proporciona una base práctica para el análisis de políticas y un entorno para valorar el desempeño. Este índice es obtenido por las Universidad de Yale y la Universidad de Columbia en colaboración con el Foro Económico Mundial.

El índice es un marco métrico para el análisis de políticas ambientales globales, clasificando a 180 países según 32 indicadores de desempeño de 11 categorías sobre temas que cubren desde la salud ambiental hasta la vitalidad del ecosistema. La metodología completa, los datos y los resultados están disponibles en línea en epi.yale.edu.

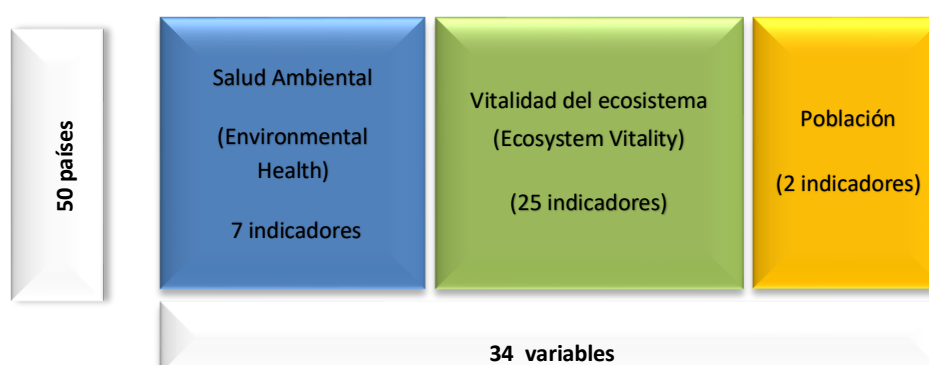


Figura 38. Estructura de la base de datos

En total la base de datos que utilizaremos cuenta con 34 variables y 50 países (ver Figura 38), la descripción de las mismas es proporcionada en la tabla 4.

Tabla 5. Descripción de Variables

Área	Factor	Var	Descripción
Salud Ambiental	Calidad del aire	PMD	Exposición a PM2.5 (partículas en suspensión)
		HAD	Combustibles sólidos domésticos
		OZD	Exposición a ozono
	Saneamiento y Agua potable	USD	Saneamiento inseguro
		UWD	Agua potable insegura
	Metales pesados	PBD	Exposición al plomo
	Gestión de residuos	MSW	Residuos sólidos controlados
Vitalidad del ecosistema	Biodiversidad y Hábitat	TBN	Protección del bioma terrestre (ponderaciones nacionales)
		TBG	Protección del bioma terrestre (ponderaciones globales)
		MPA	Áreas marinas protegidas
		PAR	Índice de representatividad de áreas protegidas
		SHI	Índice de hábitat de especies
		SPI	Índice de protección de especies
		BHV	Índice de hábitat de biodiversidad
	Servicios de ecosistema	TCL	Pérdida de cobertura arbórea
		GRL	Pérdida de pastizales
		WTL	Pérdida de humedales
	Pesca	FSS	Estado de la población de peces
		RMS	Índice trófico marino
		FGT	Pescado capturado por arrastre
	Cambio climático	CDA	Tasa de crecimiento de emisiones ajustada para dióxido de carbono
		CHA	Tasa de crecimiento de emisiones ajustada para metano
		FGA	Tasa de crecimiento de emisiones ajustada para gases fluorados

Área	Factor	Var	Descripción	
		NDA	Tasa de crecimiento de emisiones ajustada para óxido nitroso	
		BCA	Tasa de crecimiento de emisiones ajustada para carbono negro	
		LCB	Tasa de crecimiento de las emisiones de dióxido de carbono de la cubierta terrestre	
		GIB	Tasa de crecimiento de la intensidad de los gases de efecto invernadero	
		GHP	Emisiones de gases de efecto invernadero per cápita	
	Emisiones contaminantes	SDA	Tasa de crecimiento de emisiones ajustada para dióxido de azufre	
		NXA	Tasa de crecimiento de emisiones ajustada para óxidos nitrosos	
	Agricultura	SNM	Índice de gestión sostenible del nitrógeno	
	Fuentes de agua	WWT	Tratamiento de aguas residuales	
	Población	Economía	IEF	Índice de libertad económica
		Urbanización	URP	Porcentaje anual de población a mitad de año que reside en zonas urbanas

4.5.2. ANÁLISIS

Primeramente, presentamos los resultados de STATIS-dual, para luego hacer los análisis de Sparse STATIS-dual.

El primer paso del Stasis Dual, es realizar el análisis de la inter-estructura con el fin de determinar si las estructuras de covariancias son similares en los diferentes años. Recordemos que se tienen, cinco matrices de datos correspondientes a los años 2000, 2005, 2010, 2015 y 2020.

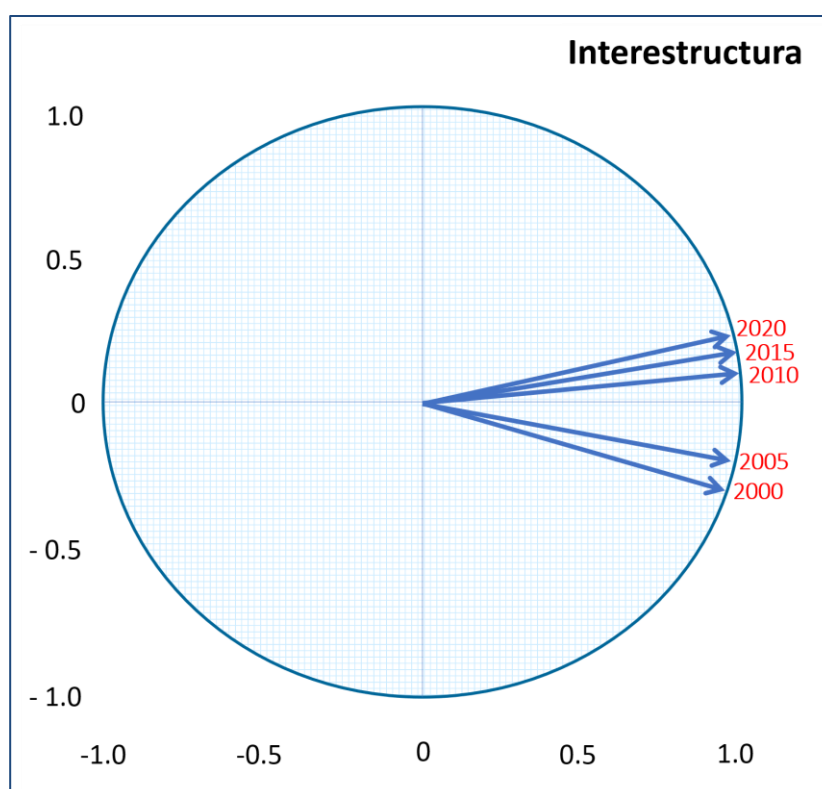


Figura 39. Interestructura del análisis STATIS-dual.

Como se puede observar en las Figura 39 y 41 todos los años correlacionan con el primer eje, no obstante, los años 2000 y 2005 tienen una estructura de covariancias diferentes respecto de los años 2010, 2015 y 2020, lo cual puede indicar que los indicadores ambientales analizados han presentado cambios importantes. Todos los años tienen una representación euclídea importante, dado que, a mayor longitud de la flecha, mejor es la representación.

La segunda parte, es la construcción del compromiso, la [Figura 40](#) nos muestra las variables sobre el compromiso, esta proyección se obtiene con la descomposición en valores y vectores singulares, que proporciona una imagen euclídea, en el que es posible proyectar las variables.

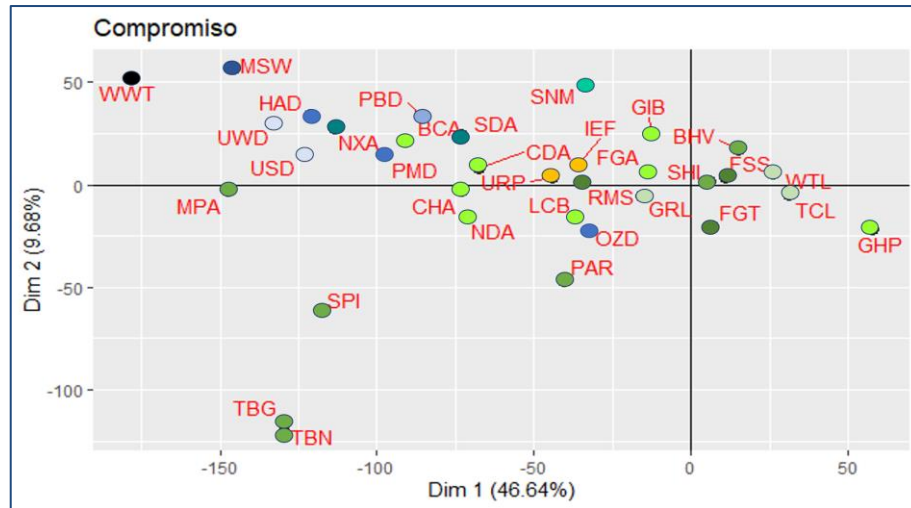


Figura 40. Proyección del espacio Compromiso, análisis STATIS-dual.

Seguidamente presentamos los resultados de la aplicación del Sparse STATIS-dual y haremos la respectiva comparación con los resultados del análisis STATIS-dual.

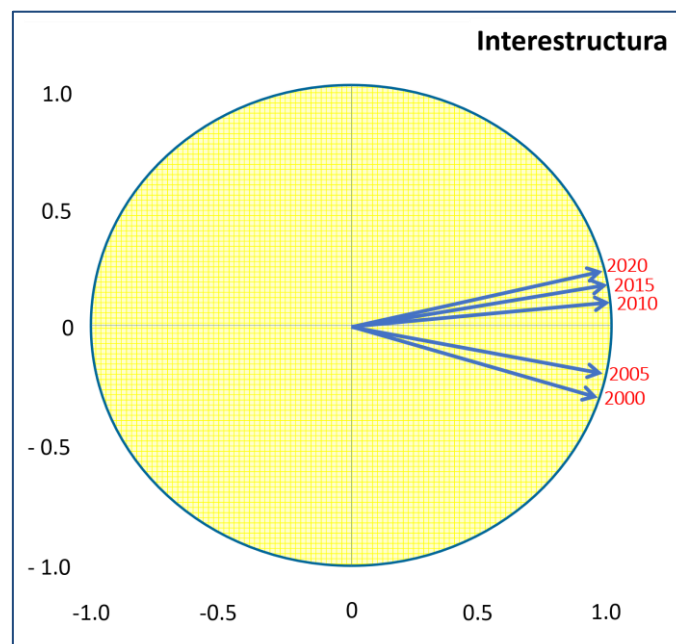


Figura 41. Interestructura generada por el análisis Sparse STATIS-dual

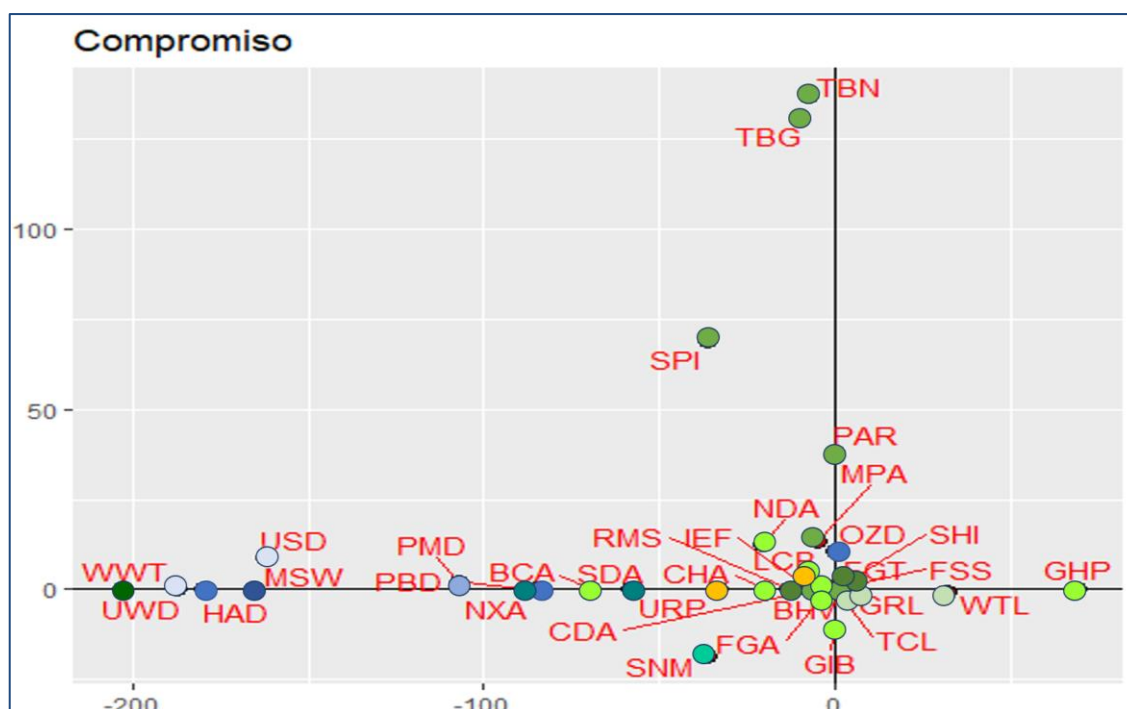


Figura 42. Proyección de las variables sobre el compromiso penalizado, Sparse STATIS-dual.

Los resultados obtenidos con la penalización Elastic Net, muestran que algunas cargas se anulan totalmente, tanto en la primera componente, como en la segunda, quedando representados sobre los ejes factoriales. Observamos en la [Figura 42](#), como algunas variables se han movido hacia la primera y segunda componente, quiere decir que algunas variables cuyas cargas en la primera componente o en la segunda componente tenían valores muy bajos, han sido forzadas a cero. Aun así, se puede observar que ciertos indicadores se mantienen cerca de la primera o de la segunda componente, sin anularse. Así, aunque son pocas variables, es factible proporcionar una mejor interpretación de los resultados obtenidos mediante esta técnica y precisar que variables contribuyen más a la primera componente y cuáles a la segunda componente. En este caso podemos observar que, la primera componente recoge los indicadores del factor salud ambiental; mientras que la segunda componente recoge los indicadores del factor vitalidad del ecosistema.

Como nos muestra la [Figura 43](#), los países de la región europea muestran mejores indicadores ambientales, sobre todo en aquellos del factor salud ambiental.

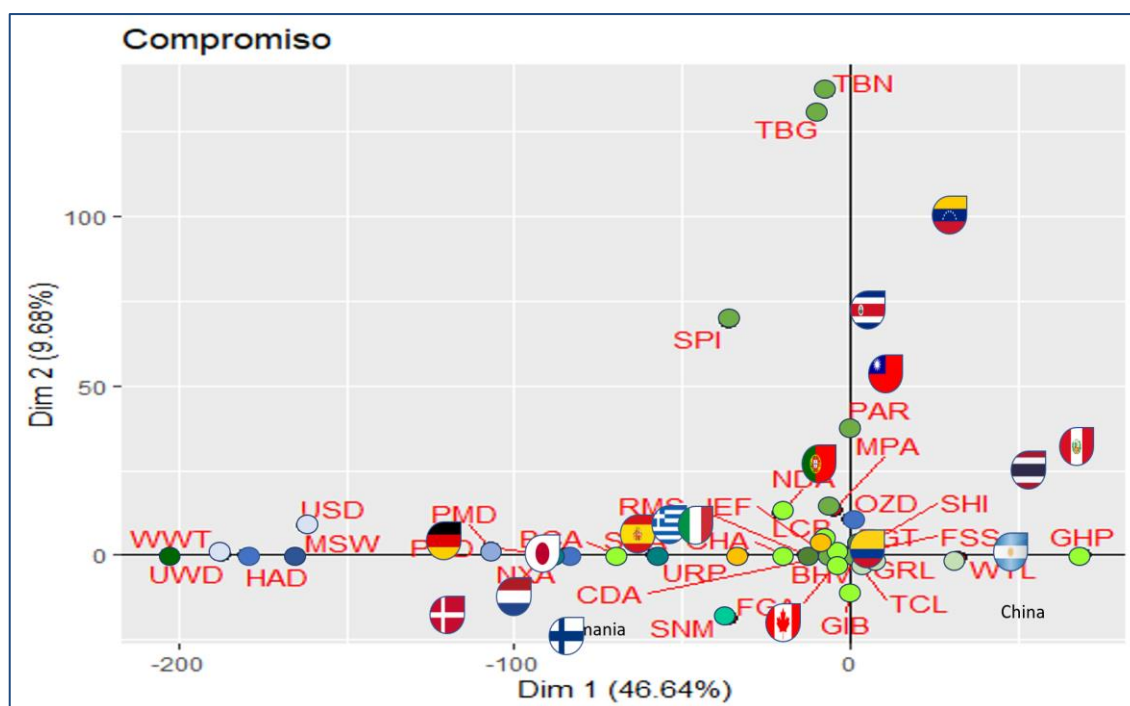


Figura 43. Proyección de las variables y variables sobre el compromiso penalizado, Sparse STATIS-dual.

Podemos observar en la [tabla 6](#) que la variabilidad explicada por cada eje en el Elastic Net varía con relación a las del STATIS-dual, ya que, el Elastic Net como método de regularización se centran en quedarse con las variables más importantes para cada componente, haciendo nulas las cargas menos importantes. Es así que, en las primeras componentes principales se concentra la proporción de varianza explicada, con mayor o menor porcentaje, esto de acuerdo al valor de penalización que se haya utilizado.

Tabla 6. Cargas de las tres primeras componentes principales obtenidas mediante el STATIS-dual y el método de regularización Elastic Net

Var	STATIS-dual			Elastic Net STATIS-dual		
	CP1	CP2	CP3	CP1	CP2	CP3
BCA	-0.0373	0.0422	0.1680	-0.0285	0.0000	-0.0026
BHV	0.0061	0.0337	-0.0080	0.0000	-0.0009	0.0000
CDA	-0.0278	0.0161	-0.0193	-0.0019	0.0000	0.0076
CHA	-0.0299	-0.0057	-0.0228	-0.0082	0.0000	0.0230
FGA	-0.0057	0.0139	0.0676	-0.0008	0.0000	0.0000
FGT	0.0029	-0.0399	0.0200	0.0000	0.0020	0.0000
FSS	0.0045	0.0066	-0.0455	0.0000	0.0000	0.0000
GHP	0.0238	-0.0432	-0.1049	0.0280	0.0000	0.0000
GIB	-0.0051	0.0496	-0.0518	0.0000	-0.0218	0.0000
GRL	-0.0063	-0.0112	-0.0753	0.0000	0.0000	0.0001
HAD	-0.0498	0.0667	0.1576	-0.0738	0.0000	0.0000
LCB	-0.0150	-0.0316	0.0570	-0.0032	0.0104	0.0000
MPA	-0.0605	-0.0058	-0.3834	-0.0021	0.0265	0.4007
MSW	-0.0596	0.1121	-0.1279	-0.0678	0.0000	0.1009
NDA	-0.0294	-0.0306	0.0244	-0.0085	0.0246	0.0000
NXA	-0.0466	0.0539	0.0617	-0.0350	0.0000	0.0000
OZD	-0.0133	-0.0439	-0.0515	0.0000	0.0215	0.0000
PAR	-0.0166	-0.0921	-0.0321	0.0000	0.0748	0.0000
PBD	-0.0351	0.0646	-0.0177	-0.0363	0.0000	0.0120
PMD	-0.0399	0.0300	0.0368	-0.0440	0.0045	0.0000
RMS	-0.0143	0.0029	0.0538	-0.0056	0.0000	0.0000
SDA	-0.0300	0.0476	0.0182	-0.0240	0.0000	0.0000
SHI	0.0024	0.0029	0.0111	0.0000	0.0000	0.0000
SNM	-0.0140	0.0950	0.0361	-0.0149	-0.0364	0.0000
SPI	-0.0483	-0.1223	0.0528	-0.0150	0.1364	0.0000
TBG	-0.0532	-0.2277	0.0305	-0.0041	0.2591	0.0000
TBN	-0.0530	-0.2386	0.0125	-0.0032	0.2706	0.0000
TCL	0.0129	-0.0098	0.0402	0.0000	0.0000	0.0000
USD	-0.0505	0.0305	0.0970	-0.0666	0.0175	0.0000
UWD	-0.0545	0.0600	0.1084	-0.0768	0.0000	0.0000
WTL	0.0108	0.0108	-0.0926	0.0129	-0.0014	0.0231
WWT	-0.0731	0.1022	-0.0952	-0.0832	0.0000	0.1019
IEF	-0.0148	0.0195	-0.0231	-0.0031	0.0000	0.0000
URP	-0.0182	0.0077	0.0483	-0.0136	0.0000	0.0000
Sparsity	0	0	0	9	19	25
% var expl.	66.07	9.19	6.69	60.40	12.10	7.40
% total	81.95			79.90		

CONCLUSIONES

1. La revisión bibliográfica ha mostrado que los métodos STATIS se han desarrollado ampliamente, sobre todo a partir del año 2000. Desde la propuesta original de L'Hermier des Plantes en 1976 con STATIS y STATIS-dual(L'Hermier des Plantes, 1976), y la de Jaffrenou en 1978 con X-STATIS o PTA, no se habían presentado nuevos métodos de esta familia. A partir del año 2004 se presentan 17 nuevas propuestas, de las cuales las más recientes son: Sir-STATIS, HiDiSTATIS , DiDiSTATIS, CATATIS y CLUSTATIS.
2. De la exhaustiva revisión bibliográfica realizada, hemos encontrado que los métodos sparse, se han aplicado a técnicas de análisis de datos de tres vías como lo son el Candecomp/Parafac y más recientemente a los métodos Tucker.
3. No se ha encontrado en la literatura referencias que implementen técnicas de restricción en los métodos STATIS. Es así, que nuestro principal aporte es una nueva propuesta de STATIS (Sparse STATIS-dual), basado en las restricciones Elastic Net. Al combinar las normas l_1 y l_2 , en la penalización, tiende a dar menos cargas establecidas en cero que un ajuste en un valor de la norma l_1 , y una mayor contracción de los otros coeficientes.
4. Consideramos que el uso de las técnicas de regularización sparse en el STATIS, permite obtener soluciones eficientes a problemas de alta dimensionalidad de los datos.
5. Para dar soporte al nuevo método del Sparse STATIS-dual planteado, se ha implementado una librería en el lenguaje de programación R, facilitando su aplicación a cualquier conjunto de datos mediante funciones específicas. El software desarrollado en R, llamado SparseSTATISdual, permite aplicar la propuesta teórica de esta tesis doctoral.
6. Se han desarrollado dos librerías más para los métodos STATIS (clásico) y STATIS-dual.
7. Se han comparado los resultados obtenidos mediante el STATIS-dual y el Sparse STATIS-dual, aplicados a datos sobre indicadores ambientales a nivel mundial, compilados en el Environmental Performance Index de la

Universidad de Yale. De esta manera, se ha demostrado que al pasar del STATIS-dual a los resultados del Sparse STATIS-dual se ha reducido el número de variables, lo que hace más obtener una mejor interpretación de las variables.

8. Esta contribución abre las puertas para el desarrollo de nuevas aplicaciones de las Sparse en otras técnicas de la familia STATIS.

ARTICULO PUBLICADO



Article

Exploring Relationships between Environmental Performance, E-Government and Corruption: A Multivariate Perspective

Carmen C. Rodríguez-Martínez ^{1,2,*}, Isabel María García-Sánchez ³,
Purificación Vicente-Galindo ^{1,4} and Purificación Galindo-Villardón ^{1,4}

¹ Departamento de Estadística, Universidad de Salamanca, 37007 Salamanca, Spain; purivg@usal.es (P.V.-G.); pgalindo@usal.es (P.G.-V.)

² Departamento de Estadística, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, 0824-Estafeta Universitaria, Panama

³ Instituto Multidisciplinar de Empresa-IME, Campus Miguel de Unamuno, Universidad de Salamanca, 37007 Salamanca, Spain; lajefa@usal.es

⁴ Instituto de Investigación Biomédica (IBSAL), 08028 Salamanca, Spain

* Correspondence: cc.rodruigm@usal.es; Tel.: +34-698-704-345

Received: 29 September 2019; Accepted: 13 November 2019; Published: 18 November 2019



Abstract: In order to improve environmental performance, the participation of effective and transparent citizens and governments that help counteract corruption in environmental matters is crucial. In this sense, this work focuses on exploring relationships between e-participation, e-government, the corruption index and environmental performance indicators. To this end, a sample comprising 116 countries from varying geographic regions is used in conjunction with indicators of environmental performance, e-participation, e-government and the corruption index. Through the use of the HJ-biplot and STATIS multivariate statistical techniques, it will be possible to observe the role that these variables play in countries' behavioural patterns with respect to environmental performance. The results show a correlation between the indicator 'perception of corruption' and environmental performance; therefore, the lower the level of corruption, the higher the environmental performance index. We conclude that countries that exhibit more e-participation, lower levels of corruption and better level income are more likely to follow policies and programmes aimed at achieving better environmental performance.

Keywords: e-participation; sustainability; STATIS; HJ-biplot; environmental performance; multivariate analysis; corruption

1. Introduction

This research focuses on examining how environmental performance is related to e-government, e-participation and the corruption index. We seek to understand relationships between the different variables and how they explain the similarities and differences between countries, within the context of their level income. We base our study on previous works such as Wang et al. [1] that show the effect of corruption on poor environmental performance; Burakov et al. [2] highlights that corruption and inequality increase environmental degradation and; Anisenco et al. [3] indicate that the development of e-government is positively related to sustainable development and better governance. Thus, we have selected the indicators e-participation, e-government and corruption to explore these relationships as a whole, using multivariate techniques, all seen from the point of view of the income levels of the countries.

We make a comparison at the country level since the level of development of citizen participation, and environmental sustainability is determined by the different policies that governments develop at

the local, regional and national levels. Although the regional contribution differs depending on the level of federalist structure of each country, the role of the state administration and, especially, of local governments is unquestionable in all the countries analysed. The relevance that local policies have over national policies in the inclusion of the opinion of citizens in sustainable development should be noted because a high percentage of society lives in cities, with urban areas being the demarcations used for (i) the carrying out all types of citizen consultations; and (ii) the establishment of policies that allow the design of sustainable cities

To achieve development with environmental sustainability, a set of regulatory, administrative and operational actions is necessary, which must be promoted by the State and local governments; this is known as environmental management [4,5]. In this context, environmental information is fundamental within environmental management processes, since it configures the base that facilitates the quantification and measurement of environmental quality. Several indicators have been implemented in order to understand and manage all sustainability issues. Within these, the Environmental Performance Index (EPI) [6] provides a nationwide indicator of how close countries are to their established environmental policy objectives.

The rapid development of information and communication technologies (ICT) has generated the growth of e-participation, whereby governments use digital tools to boost citizen participation, developing e-government. According to Justice, Melitski and Smith [7], e-governments improve transactions between the government and other agents, changing both the provision of public services and the broader field of interactions between citizens and the government [8]. E-governments allow citizens to take part in democratic institutions and political processes [7,9], producing an image of accessible, transparent, responsible, effective and participatory governance [10]. However, the lack of transparency of online information by institutions exerts a negative influence regarding sustainability [11].

E-participation contributes to the increase and progress of communications between the citizens themselves and the ruling politicians [12] and therefore, in democratic processes [13]. In this sense, e-participation is of vital importance for the development of policies aimed at improving sustainability [14].

Regarding the above, the United Nations [15] notes that governance for sustainable development goes hand-in-hand with national strategic planning to ensure efficiency and transparency. According to Voß et al. [16], environmental and sustainability issues are characterised by their complexity and, as indicated by Renn and Schweizer [17], special attention should be paid to the structure of the participation process in order to achieve the stated objectives.

On the other hand, corruption has become a global problem and compromises governance, transparency and accountability [18], becoming a very worrying problem in developing countries [19]. This is particularly with regard to environmental issues since these are extremely vulnerable to corruption [20–22], affecting environmental quality through different routes [23]. In this sense, we ask ourselves, could e-participation and e-governments play an important role in anti-corruption policy and therefore contribute to the improvement of environmental performance?

Environmental performance as a dimension of sustainable development is of great importance to today's society, which justifies the need to know the interests of citizens and their specific demands in relation to the environment. In this sense, we explore how the environmental development index relates to other variables linked to transparent governments in 116 countries from varying geographical areas. Using a variety of sources, we examine the relationship of the environmental development index with three indicators: e-government, e-participation and corruption index. Also, countries were analyzed according to their income level.

Our objective is to explore how the environmental development index is related to these variables, and if any relationship is clearly stronger; in the same way, we sought to understand the structures underlying the data, specifically to relations between the different indices and how these indices manage to explain the similarities and differences between countries. We also contribute to studies on environmental performance and citizen perception through recently developed multivariate techniques such as HJ-biplot and STATIS.

In terms of previous arguments and empirical evidence, we defend the proposal that countries with higher levels of e-participation, better e-government indicators and lower levels of corruption generally have better environmental performance indicators and that these indicators behave according to the economic level of the countries. Considering the findings of these previous studies, we hypothesised that e-participation, e-government, corruption index and income level are related to the environmental performance index. These variables, therefore, were addressed in this study from the perspective that a better understanding of the possible relationships of these variables on environmental performance could be used as the basis for developing specific studies of these relationships, leading to improved levels of environmental performance.

The document is divided into five sections: Section 1 contains a brief introduction to the topic; Section 2 presents the sample and the methodology used, Section 3 contains the main results and Sections 4 and 5 present the discussions and conclusions of this research.

2. Materials and Methods

2.1. Population and Sample

To achieve the objective of our work, we considered the 236 countries reported in the EPI report for the year 2018 [6]. The sample analysed was from 116 countries from varying geographical regions and level income. Countries that did not have information available for all the variables under analysis were excluded.

The database used in this study was compiled from three repositories, the first, the EPI [24], which presents the EPI index and performance indicators in ten thematic levels that involve environmental health (HLT-E) and ecosystem vitality (ECO-E); the second, the United Nations E-Government Development Database (EGDI) [25] provides the indexes of e-government (E-GOV) and e-participation (EPI-P); the third corresponds to the international transparency repository, which provides the corruption perception index (CPI) [26].

In Table 1, we present a description of the variables used. The environmental health dimension (HLT-E) with three indicators, the ecosystem vitality dimension (ECO-E) with seven indicators and the dimension we term perception (PER), which includes the three indices: e-government, e-participation and corruption index.

Table 1. Description of variables used.

Index	Dimension	Indicator	Indicator Code	Data Source
Environmental Performance Index (EPI-E)	Environmental Health (HLT-E) (Matrix 1)	Air Quality	AIR	EPI
		Water & Sanitation	H2O	EPI
		Heavy Metals	HMT	EPI
	Ecosystem Vitality (ECO-E) (Matrix 2)	Biodiversity & Habitat	BDH	EPI
		Forests	FOR	EPI
		Fisheries	FSH	EPI
		Water Resources	WRS	EPI
		Agriculture	AGR	EPI
		Air Pollution	APE	EPI
		Climate & Energy	CCE	EPI
Perception (PER) (Matrix 3)	Perception (PER) (Matrix 3)	e-government	E-GOV	EGDI
		e-participation	EPI-P	EGDI
		Corruption Perception index	CPI	CPI

Appendix A (Table A1) includes the nominal type variable level of income (level income) classification, according to the GNI 2018 [27]. This classification is valid for the 2020 fiscal year and will allow us to differentiate countries according to their level of income, in order to obtain a more accurate general evaluation of environmental performance according to the economic capacity of each country.

2.2. Methodology I: HJ-biplot Method

The HJ-biplot method is an exploratory statistical technique for the representation of multidimensional data [28]; it is a variant of the biplot graphic presentation proposed by Gabriel [29]. The HJ-biplot technique allows rows and columns of the data matrix to be presented as points in a small vector space. The quality of representation for both individuals and variables is superior to that achieved with similar techniques [28], which demonstrates its robustness. The method is, in some ways, similar to correspondence analysis but is not restricted to frequency data [30].

The technique starts from a matrix $X_{n \times p}$, composed of n individuals measured in p variables. The columns of X are usually standardised to make them comparable. The singular value decomposition (SVD) of X is defined as $X = U\Lambda V^T$, where U is a matrix whose column vectors are orthonormal and correspond to the eigenvectors of XX^T ; V is a matrix whose column vectors are also orthonormal and correspond to the eigenvectors of $X^T X$; and Λ is a diagonal matrix that contains the singular values arranged in descending order [31]. $X^T X$ captures the information of the covariation between the variables and XX^T captures the similarity of individuals (countries) in relation to the variables under study.

The results facilitate the simultaneous representation of the data and allow us to obtain optimal representation quality for both the rows and the columns of the data matrix. In the biplot representation, the b_p column markers (variables) are shown as arrows and the a_n^T row markers (countries) as points.

HJ-biplot provides clusters like the classic cluster analysis, but with the additional advantage that, by doing a cluster analysis on the coordinates of the HJ-biplot, we can also know why countries are grouped, which is not possible with classic cluster analysis.

Interpretation rules of the HJ-biplot are a combination of the rules used in other classical biplots, correspondence analysis, multidimensional scaling techniques and factor analysis. The length of the vector approximates the variability of each variable, and the angle approximates the correlation between variables; the closer markers (countries) are more similar; this allows for the identification of clusters of individuals (countries) with similar profiles. As an aid to identify the relevant clusters, any hierarchical or non-hierarchical cluster technique can be used, this allows for the identification of clusters of individuals (countries) with similar profiles; the cosines of the angles among the column vectors approximate the correlations among variables, right angles are associated to non-correlated variables, obtuse angles near to the straight angle are associated to variables with high negative correlation, and small acute angles are associated to variables with high positive correlation.

The main purpose of the HJ-biplot analysis is to extract and describe the relationships between the environmental index, e-government, e-participation and corruption index, and classify countries according to them. This information is contained in a matrix $X_{116 \times 4}$.

The countries are represented as points and variables as arrows (vectors). In this way, we can visualise countries with similar behaviours, that is, the distance between points is interpreted as similarity so that the nearest countries have similar profiles; see Figure 1.

The relationships between environmental performance indicators, electronic government and corruption perception measures are described by the angles between the vectors; acute angles are associated with a positive correlation. To classify the countries by their indicators of environmental performance, e-government, e-participation and corruption index, that is, by the orthogonal projections of the points (countries) on the vectors (variables), we can order the different countries in relation to each variable.

This method has been applied by Carrasco et al. [32] in hydrology; Cubilla-Montilla et al. [33], Amor-Esteban et al. [34] and García-Sánchez et al. [35] in the context of corporate social responsibility; Castellanos-Martín et al. [36] in breast cancer research; Gallego-Alvarez et al. [37] and Tejedor-Flores [38] to analyse sustainability; Nieto-Librero et al. [39] in geochemical studies; Hernández et al. [40] to study bioactive compounds in agriculture; Mendes et al. [30] in environmental science; García-Talegón [41], in geology, and; Orfao et al [42] in immunology.

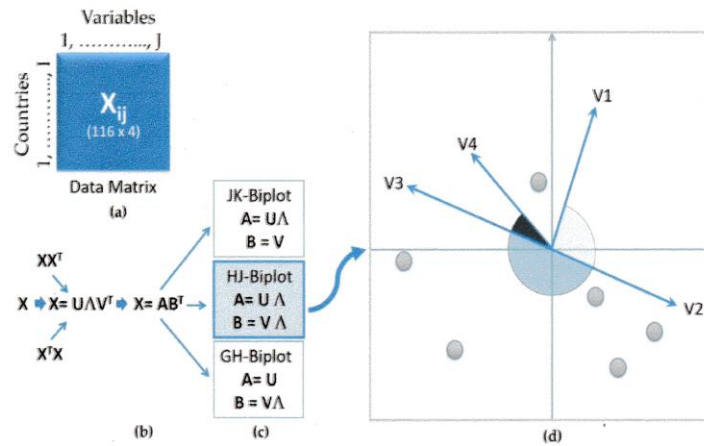


Figure 1. (a) Structure of the data matrix; (b) phases of the biplot process; (c) types of biplot; (d) representation of individuals and variables.

2.3. Methodology II: STATIS Method

The STATIS method, “structuration des tableaux à trois indices de la statistique” [43], is essentially a generalisation of the principal component analysis (PCA), whose objective is to analyse several sets of variables within a single set of observations in what is called multiple set data. It is a method that allows the exploration of the structure of several data tables. It consists of reducing the dimensionality of the data using a similarity measure based on Euclidean distances between point configurations. The variables may be different in each of the tables.

The STATIS method starts with K tables, with X_1, \dots, X_K , k data matrices with J_1, \dots, J_K quantitative variables, respectively, observed in the same group of n individuals.

Next, the scalar product between matrices W_k and $W_{k'}$ is calculated. To this purpose, a correlation coefficient between matrices is defined; in this case, the vector correlation coefficient RV [44].

The data are arranged in three tables; the first table corresponds to the information of n individuals (countries) measured on p variables (environmental health/HLT-E); the second table has information on the same n individuals measured on q variables (ecosystem vitality/ECO-E), and the third table corresponds to information on the same n individuals measured on r variables (perception/PER). The STATIS method is based on the development of the following steps (see Figure 2):

- Inter-structure analysis: its objective is to make a global comparison of data matrices. A matrix of similarities between individuals (countries) is generated for each matrix, which describes the relationship between matrices.
- Compromise matrix: summarises the matrices into one that is the weighted average of all of them. This matrix synthesises the information of each one of the 116 countries in the variables for the three matrices, ‘filtering the noise’ and representing the statistically significant information.
- Intra-structure analysis: a representation of the countries and variables, which allow exploring the differences or similarities that exist between the data matrices, to define the average position of the countries in the different matrices.

The first works in the framework of this methodology were those carried out by L’Hermier des Plantes and Thiébaud [45] and Lavit [46]. Subsequently, related techniques have been developed such as X-STATIS or Triadic Partial Analysis (PTA) [47], DISTATIS [48] and COVSTATIS [49], all of which take into account data from departure. On the other hand, STATICO (STATIS and COINERTIA) [50,51]; and COSTATIS (COINERTIA AND STATIS) [49] work with pairs of tables in different situations or

times. Power-STATIS [52] assigns weights when creating the consensus matrix; whilst t+1 STATIS [53], STATIS-4 [54] and CANOSTATIS [55] take into account external information. The latest contributions have been INTER-STATIS [56] for interval data and CLUSTATIS [57] to group datasets measured in the same individuals.

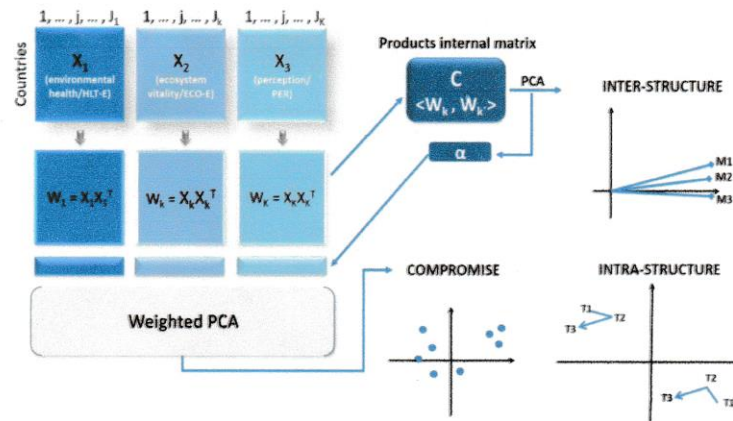


Figure 2. Figure 2. STATIS scheme: structure of the matrices and phases of the STATIS process.

These methods have been applied in different fields; we can mention the recent works of Fu et al. [58] in data on the user interfaces of mobile shopping applications; Feki-Sahoun et al. [59] evaluated the biogeography of phytoplankton; Rundle et al. [60] applied it in neurosciences; Amor-Esteban, Galindo-Villardón and García-Sánchez [34] evaluated corporate social responsibility; Marcondes Filho, Fogliatto and Oliveira [61] and Ramos-Barberán et al. [62] proposed control charts based on the STATIS method; Caballero-Julia, Galindo-Villardón and García [63] applied the method to textual data; Bono and Giacomarra [64] measured the performance of technical efficiency in the photovoltaics sector.

The HJ-biplot and STATIS analysis, as well as in the graphs, were performed using the software Multiplot [65].

3. Results

3.1. HJ-biplot Analysis.

The first part of our analysis is the application of the HJ-biplot method; the objective is to analyse whether there is an association between global corruption indexes (CPI), e-participation (EPI-P), e-government (E-GOV) and environmental performance (EPI-E).

We denote with X the matrix with dimensions n (country), J (variables), where $n = 116$, number of rows, and J , the number of columns or variables, which in this case are the four global indices; therefore, we have a matrix (116×4) .

The first axis explains 81.6% of the variability; we will use the 1-2 factorial plane that explains 92.4% of the total variability since it is enough to explore the main characteristics of the data (see Figure 3).

In terms of the indicators, the groups by e-government level are placed on the graph by configuring a gradient from left to right in order: low (red), medium-low (yellow), medium-high (blue) and high e-government (green). This means that the main reason for the separation of the countries is the level of e-government since it is the characteristic that is represented in the abscissa axis, which is the axis that retains the greatest variation because it is associated with the greater eigenvalue. It should be noted that most countries achieved a good quality of representation at this level, although this was somewhat lower for Iran, Venezuela, Sri Lanka, Kuwait, Argentina, and Albania.

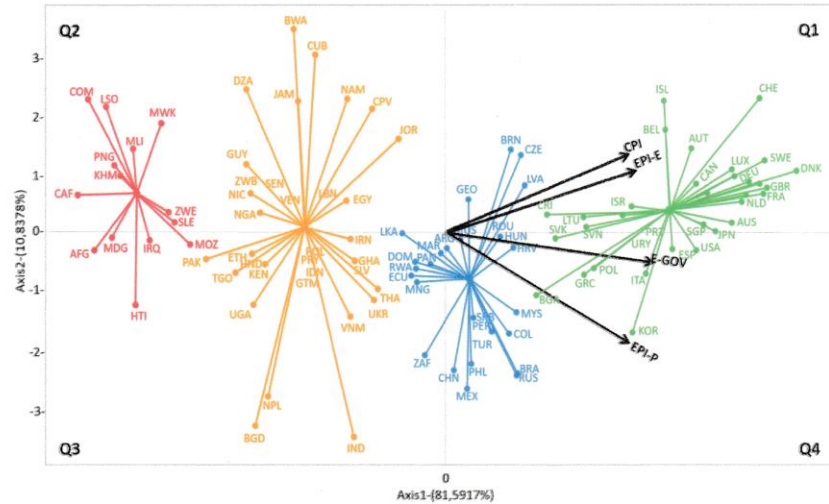


Figure 3. HJ-biplot representation: the position of all countries in relation to the indices. Mauritius, Georgia, Panama, Rwanda, Dominican Republic and Morocco.

The four indicators correlate with each other; however, we observe that there is a closer correlation between the corruption perception indicator (CPI) and the environmental performance indicator (EPI-E); the lower the level of corruption, the higher the index of environmental performance (EPI-E). On the other hand, e-government (E-GOV) and e-participation (EPI-P) indices demonstrate an association, which is expected, since they measure the levels of e-government and e-participation.

The countries located in the first quadrant (Q1) have the lowest rates of corruption (CPI) and the highest rates of environmental performance (EPI-E), meaning that their priorities are to improve environmental performance and reduce corruption. Among the countries that make up this group, we can mention Switzerland, Denmark, Iceland, Austria, Belgium and Sweden, among others, which possess a high economic income. In the fourth quadrant (Q4) are the countries that give higher priority to electronic government (E-GOV) and citizen electronic participation (EPI-P), which have high economic income (for example, Korea, USA, Spain and Japan). Between the second (Q2) and third (Q3) quadrants, there are the countries with the lowest rates, among which we can mention Central African Republic, Afghanistan, Haiti, Egypt, Nicaragua and India, possessing low and medium-low economic income.

3.2. STATIS Analysis

The second part of our analysis is the implementation of the STATIS method, which will allow us to analyse the importance of the dimensions, in addition to studying the similarities and differences between the indicators of each dimension. Thus, we denote with X_k data matrices with I equal to 116 rows (countries) and J_k variables, where K (studies/dimensions) is the number of tables or matrices. Therefore, we obtain three matrices ($I \times J_k$): the first matrix has three indicators of the HLT-E; the second has seven indicators of the ECO-E, and the third with has three indicators of the perception dimension.

As a first step, we will obtain the inter-structure, consisting of the comparison of the structures of the three matrices of the study. The vector correlation matrix is calculated, which allows us to observe the relationship between the matrices. When analysing the vector correlation coefficients (RV) and the distances between the matrices, two similar dimensions are observed, HLT-E and perception (PER) with an RV coefficient of 0.687, while the ECO-E is slightly different. The Euclidean representation in Figure 4 identifies the proximity between the matrices. Factorial plane 1-2 explains 89.6% of the total information.

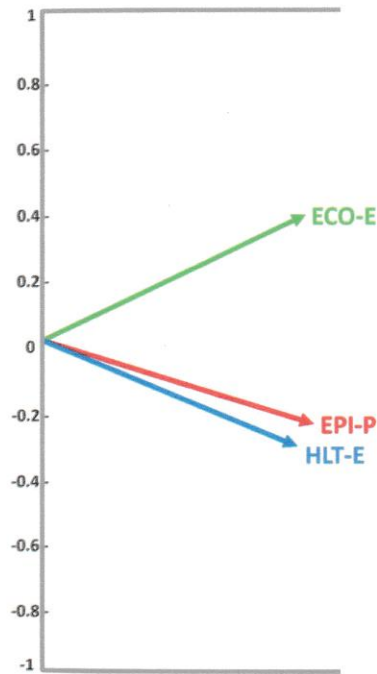


Figure 4. Inter-structure of the three matrices.

In the second step, a matrix is constructed that summarises and synthesises the common structure of all the original matrices: the compromise matrix. With this matrix, the behaviour of countries in the three matrices as a whole is compared and evaluated. The information collected by the first two axes is 59.5%, enough for our analysis. The first axis carries most of the information with 50.1%. Indicators and countries on compromise are represented in Figure 5a.

The indicators are represented by the vectors; the green ones correspond to the vitality dimension of the ecosystem (ECO-E), the red ones to the perception dimension (PER) and the blue ones to the HLT-E. All are located in quadrants 1 (Q1) and 4 (Q4) (right half-plane). The countries that are located in this semi-plane have the highest indicators across all dimensions. Strong correlations are observed between the variables of the HLT-E dimension to the PER dimension, except two indicators: air quality (AIR) and fisheries management (FSH). The strongest relationship is between e-participation (EPI-P) and lead exposure (HMT). We can form three dimensions: the first being the relationship with “ecological issues”; the second with “transparent government”; the third we will call “air and fishing”.

Regarding Environmental Health (HLT-E), all indicators are located between quadrants 1 (Q1) and 4 (Q4) (right half-plane). The countries that position themselves in this semi-plane give greater relevance to lowering levels of heavy metal pollution (HMT), water care and sanitation (H2O) and improving air quality (AIR). On the other hand, they show a high value to e-government (E-GOV), e-participation (EPI-P) and very low values the corruption index (CPI) (see Figure 5b).

In Figure 5c, the ECO-E scenario is shown; the indicators are located in the right half-plane. The countries that are located in quadrant 1 (Q1) give greater importance to improving the indicators we call “ecological issues”; these are forests (FOR), agriculture (AGR), climate and energy (CCE) and air pollution (APE), while those in quadrant IV place greater importance on fisheries management (FISH) and water sources (WRS).

To evaluate the effects of the level of income on environmental performance, we use the level income of the countries. In this way, we visualise the behaviour patterns of the countries. For this reason, and to obtain a better understanding, we project the countries in the commitment subspace, dividing them into the four levels of income (see Figure 6).

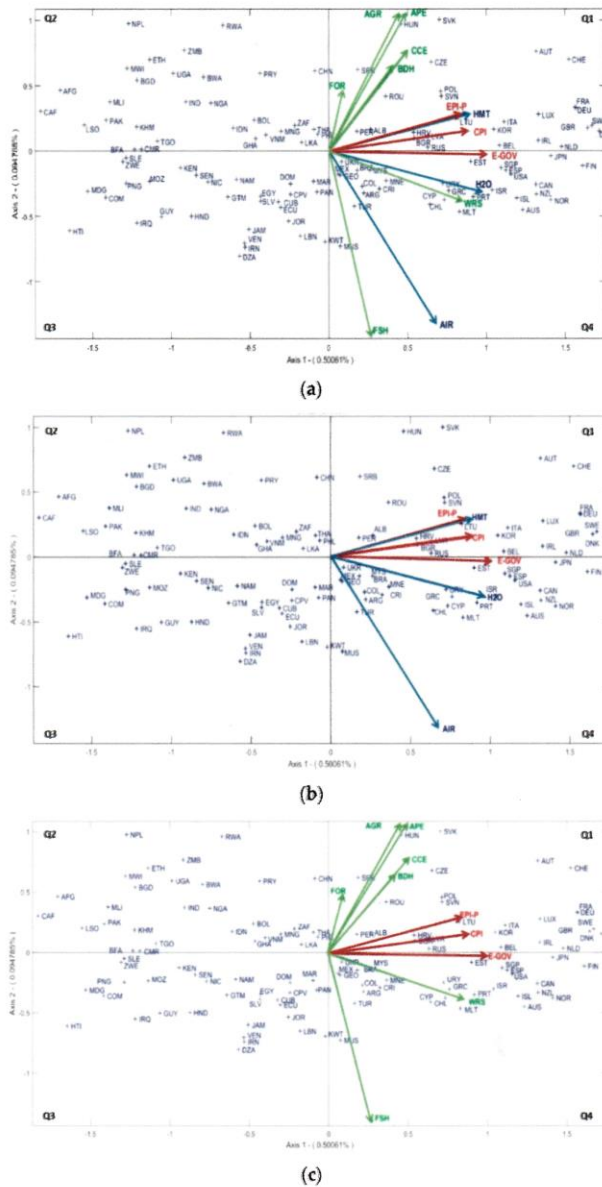


Figure 5. STATIS compromise subspace: (a) the position of 116 countries with respect to the 13 indicators; (b) indicators of the environmental health (HLT-E) and e-participation (EPI-E) matrices; (c) indicators of the ecosystem vitality (ECO-E) and perception (PER) matrices.

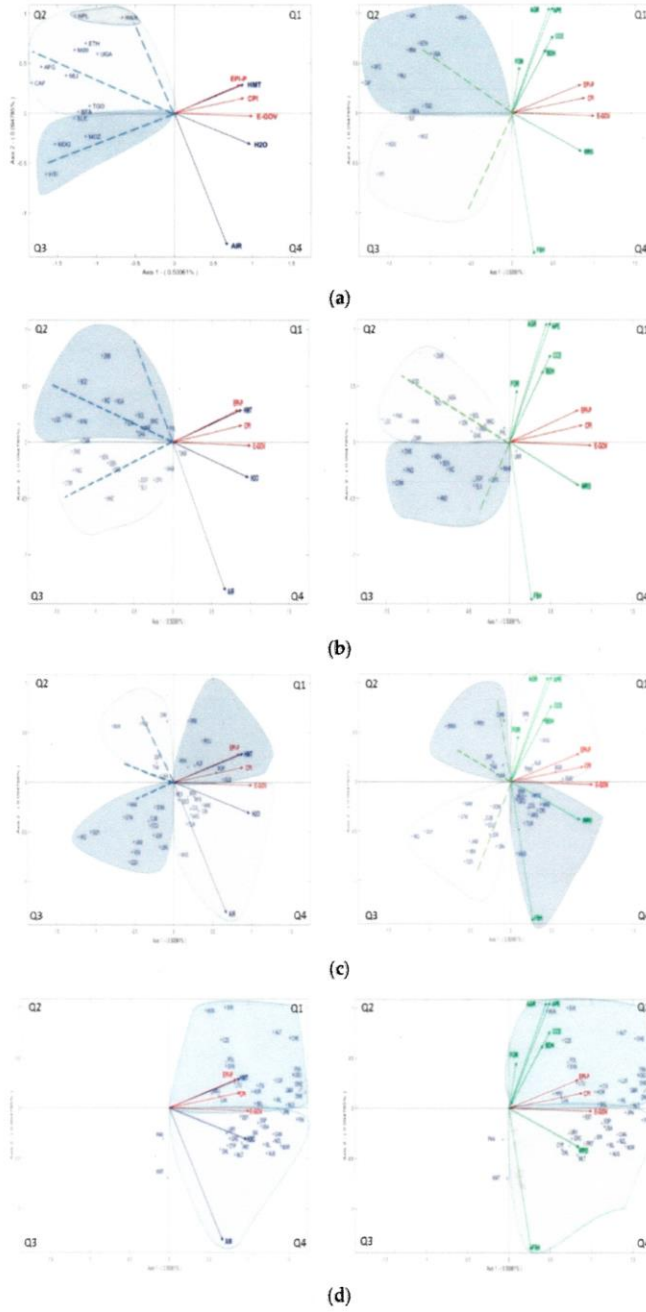


Figure 6. Factorial plane 1–2 of the compromise subspace, divided into the 4 levels of economic income: (a) low level; (b) medium–low level; (c) medium–high level; (d) high level.

Countries with low economic income show very low values to environmental performance and on perception indices; located at the left end of quadrants 2 (Q2) and 3 (Q3) (left half-plane), they have the lowest values in all their indicators, in relation to the other countries (see Figure 6a). These are notably different with respect to those of the right half-plane. Most of these countries are located in quadrant 2 (Q2), show low levels of air quality (AIR) and low levels of drinking water and sanitation (H2O), as well as low levels of wastewater treatment (WRS). Countries that are located in quadrant 3 (Q3) are characterised as having high levels of exposure to lead (HMT) and in general, very low levels of all indicators of ECO-E. This group includes Nepal (NPL) and Rwanda (RWA) with high levels of air pollution, Haiti (HTI) with high levels of lead exposure, and the Central African Republic (CAF) together with Afghanistan (AFG), both with low levels of drinking water and sanitation. On the other hand, they have low levels of e-participation and e-government and very high levels of corruption.

Figure 6b shows the middle-low income countries, which follow a pattern similar to the previous group, albeit with slightly higher indicators. They have low values in all their indicators. They are distributed in quadrants 2 (Q2) and 3 (Q3) and show deficiencies in the indicators of environmental health (HLT-E) and ecosystem vitality (ECO-E).

The countries with medium–high economic income show better indicators and are distributed in across four quadrants (see Figure 6c). Those located in quadrants 2 (Q2) and 3 (Q3) have lower indicators than those located in quadrants 1 (Q1) and 4 (Q4). Within this group, those located in quadrant 2 (Q2) have deficiencies in air quality (AIR), drinking water and sanitation (H2O) and in wastewater treatment (WRS); however, they are better positioned than the two previous groups. On the opposite end (Q4), countries with better values are located in these same indicators. On the other hand, the countries in this group, which are located in quadrant 1 (Q1), they have higher values in all their indicators, especially with regard to the indicators of ecosystem vitality (ECO-E) and e-participation. The countries with slightly lower values are positioned in quadrant 3 (Q3).

Finally, Figure 6d shows the high-income countries, which are characterised by leading all indicators. However, they are distributed in quadrants 1 (Q1) and 4 (Q4), indicating that this group places a higher priority on the indicators of ECO-E and others than environmental health (HLT-E). Thus we have a small group of countries leading in decreasing SO₂ emissions (APE), sustainable nitrogen management (AGR), climate and energy (CCE) and biodiversity and habitat (BDH); the countries that are located near axis 1 lead as regards e-government (EGOV), e-participation (EPI-P) and low levels of corruption (CPI); while those located in the fourth quadrant focus on better quality drinking water and adequate sanitation (H2O) and on wastewater treatment (WRS).

Countries with the best environmental performance and transparent governments are Switzerland (CHE), France (FRA), Germany (DEU), Sweden (SWE), Denmark (DNK) and Finland (FIN). On the other hand, countries with lower levels of environmental performance and higher levels of corruption are Central African Republic (CAF), Afghanistan (AFG), Haiti (HTI) and Madagascar (MDG).

4. Discussion

It has been shown that the combination of the HJ-biplot and STATIS methods allows us to easily observe the environmental performance index (EPI-E) and the indicators of e-government, e-participation and corruption in a single structure.

The application of the HJ-biplot method facilitated the representation of the 116 countries and the four global indicators in a graph, plotting the countries and indicators of the data matrix as points in a small vector space. The STATIS method showed the underlying structures of the data concerning relations between the different indicators and how they comprehensively explained the similarities and differences between countries.

E-participation has increased in recent decades; this is due to the popularisation of social networks; however, in our findings, we observe that e-participation is very low in low and middle-income countries. As noted by Bicking et al. [66] and Koussouris et al. [67], there are many challenges to converting existing users of e-participation platforms into more active users, in addition to encouraging

those who still do not use them. The lagging development of technological infrastructures diminishes the e-participation. Improvement in this factor will help to enrich the quality of the debate, legitimising decision-making, particularly in environmental matters.

From the point of view of government institutions, these are thought to be inadequate to meet the challenges related to environmental performance [68]. Some researchers such as Beierle and Cayford [69] and Newig and Fritsch [70] point out that citizen participation has a significant influence on environmental governance and sustainability. In this sense, our findings show that those countries with very high e-government and e-participation levels show the best indicators of environmental performance, especially in those related to environmental health.

Alió and Gallego [71], on the one hand, and Royo and Yetano [72], on the other, point out that in matters of sustainability, citizen participation plays a critical role. As Royo, Yetano and Acerete [68] identified, the global effort for environmental protection goes hand-in-hand with citizens well-informed of environmental policies and initiatives. Our results are along this vein; as we have discovered, there is a correlation between high levels of e-participation and environmental performance indicators.

Another of our results shows that countries with low levels of income have low levels of environmental performance and those countries with high levels of income also have high levels of environmental performance.

The results obtained allow us to affirm that there is a certain articulation between the e-government, e-participation and the environmental performance indices; this suggests that the issues of environmental performance are applied by governments that have a style of public administration more inclined towards citizen participation, which supports the evidence revealed by García-Sánchez, Rodríguez-Domínguez and Gallego-Álvarez [73].

5. Conclusions

The results of the HJ-biplot show us that there is a closer correlation between the CPI and the EPI-E, which indicates that the lower the level of corruption, the higher the EPI-E. Countries are grouped according to their e-government level.

A common behaviour is presented between environmental health dimension (HLT-E) and perception dimension (PER). Those countries with high levels of e-participation (EPI-P) also show lower levels of corruption (CPI).

Concerning income level, countries with low and medium-low economic income are characterised by lagging in all environmental performance indicators, as well as reflecting very low levels of e-government (E-GOV) and e-participation (EPI-P) and very high levels of corruption (CPI). While those with medium-high incomes reflect better indicators, a large number of these countries struggle to keep up. We could say that they are in a transition to better states of environmental performance. As for high-income countries, all of them reflect high levels of environmental performance, citizen participation and the reduction of corruption.

Countries that exhibit greater e-participation, lower levels of corruption and better level income are more likely to follow policies and programmes aimed at achieving better environmental performance. Our results suggest that e-participation is important in improving environmental performance.

In future studies, we recommend placing further emphasis on the real impact of e-participation in national environmental policies and programmes, expanding the qualitative analysis of specific actions in national organisations, and changes in government actions resulting from the use of e-participation.

Author Contributions: Conceptualization, C.C.R.-M., P.G.-V. and I.M.G.-S.; Data curation, C.C.R.-M.; Formal analysis, C.C.R.-M., P.V.-G. and I.M.G.-S.; Funding acquisition, C.C.R.-M.; Methodology, P.V.-G. and P.G.-V.; Project administration, I.M.G.-S.; Supervision, P.G.-V. and I.M.G.-S.; Writing—original draft, C.C.R.M.; Writing—review & editing, C.C.R.-M., P.V.-G., P.G.-V. and I.M.G.-S.

Funding: This research was funded by the National Secretary of Science and Technology of Panama (SENACYT), doctoral scholarship programme.

Acknowledgments: The author thanks Mitzi I. Cubilla M., PhD, for her support in editing this document; and to the anonymous reviewers for their comments, which helped improve the document.

Conflicts of Interest: The authors declare no conflict of interest in the results.

Appendix A

Table A1. Classification of countries by level income.

Low Income \$1035 or less	Low–Middle Income \$1036–4085	High–Middle Income \$4086–12,615	High Income \$12,616 or more
Afghanistan (AFG)	Bangladesh (BGD)	Albania (ALB)	Australia (AUS)
Burkina Faso (BFA)	Bolivia (BOL)	Algeria (DZA)	Austria (AUT)
Central African Republic (CAF)	Cameroon (CMR)	Argentina (ARG)	Belgium (BEL)
Ethiopia (ETH)	Comoros (COM)	Botswana (BWA)	Canada (CAN)
Haiti (HTI)	Cabo verde (CPV)	Brazil (BRA)	Chile (CHL)
Madagascar (MDG)	Cambodia (KHM)	Bulgaria (BGR)	Croatia (HRV)
Malawi (MWI)	Egypt, Arab Rep. (EGY)	China (CHN)	Cyprus (CYP)
Mali (MLI)	El Salvador (SLV)	Colombia (COL)	Czech Republic (CZE)
Mozambique (MOZ)	Ghana (GHA)	Costa Rica (CRI)	Denmark (DNK)
Nepal (NPL)	Honduras (HND)	Cuba (CUB)	Estonia (EST)
Rwanda (RWA)	India (IND)	Dominican Republic (DOM)	Finland (FIN)
Sierra Leone (SLE)	Indonesia (IDN)	Ecuador (ECU)	France (FRA)
Togo (TGO)	Kenya (KEN)	Georgia (GEO)	Germany (DEU)
Uganda (UGA)	Lesotho (LSO)	Guatemala (GTM)	Greece (GRC)
	Mongolia (MNG)	Guyana (GUY)	Hungary (HUN)
	Morocco (MAR)	Irán, Islamic Rep. (IRN)	Iceland (ISL)
	Nicaragua (NIC)	Iraq (IRQ)	Ireland (IRL)
	Nigeria (NGA)	Jamaica (JAM)	Israel (ISR)
	Pakistan (PAK)	Jordan (JOR)	Italy (ITA)
	Papúa Nueva Guinea (PNG)	Lebanon (LBN)	Japan (JPN)
	Philippines (PHL)	Malaysia (MYS)	Korea, Rep. (KOR)
	Senegal (SEN)	Mauritius (MUS)	Kuwait (KWT)
	Ukraine (UKR)	Mexico (MEX)	Latvia (LVA)
	Vietnam (VNM)	Montenegro (MNE)	Lithuania (LTU)
	Zambia (ZMB)	Namibia (NAM)	Luxembourg (LUX)
	Zimbabwe (ZWE)	Paraguay (PRY)	Malta (MLT)
		Perú (PER)	Netherlands (NLD)
		Romania (ROU)	New Zealand (NZL)
		Russian Federation (RUS)	Norway (NOR)
		Serbia (SRB)	Panama (PAN)
		Sri Lanka (LKA)	Poland (POL)
		South Africa (ZAF)	Portugal (PRT)
		Thailand (THA)	Singapore (SGP)
		Turkey (TUR)	Slovak Republic (SVK)
		Venezuela, RB (VEN)	Slovenia (SVN)
			Spain (ESP)
			Sweden (SWE)
			Switzerland (CHE)
			United Kingdom (GBR)
			United States (USA)
			Uruguay (URY)

Note: Bold indicates a change of classification. Source: World Bank, © The World Bank Group [74].

References

1. Wang, Z.; Zhang, B.; Wang, B. The moderating role of corruption between economic growth and CO₂ emissions: Evidence from BRICS economies. *Energy* **2018**, *148*, 506–513. [CrossRef]
2. Burakov, D.; Bass, A. Institutional determinants of environmental pollution in Russia: A non-linear ARDL approach. *Entrep. Sustain. Issues* **2019**, *7*, 510–524. [CrossRef]
3. Anisencenko, Z.; Robalino-López, A.; Rodríguez, T.E.; Pérez, B.E. Regional cooperation in dealing with environmental protection. E-government and sustainable development in Andean countries. In Proceedings of the 11th International Scientific and Practical Conference, Rezekne, Latvia, 15–17 June 2017.
4. Flores, O.M. Gestión ambiental urbana y desarrollo sustentable. Consideraciones desde un enfoque social sobre nuestro hábitat urbano. *Rev. Electrón. Ambient. Total* **2009**, *1*, 1–8.
5. Di Vaio, A.; Varriale, L. Management innovation for environmental sustainability in seaports: Managerial accounting instruments and training for competitive green ports beyond the regulations. *Sustainability* **2018**, *10*, 783. [CrossRef]

6. Wendling, Z.A.; Emerson, J.W.; Esty, D.C.; Levy, M.A.; de Sherbinin, A. *Environmental Performance Index*; Yale Center for Environmental Law & Policy: New Haven, CT, USA, 2018.
7. Justice, J.B.; Melitski, J.; Smith, D.L. E-government as an instrument of fiscal accountability and responsiveness: Do the best practitioners employ the best practices? *Am. Rev. Public Adm.* **2006**, *36*, 301–322. [CrossRef]
8. Torres, L.; Pina, V.; Acerete, B. E-government developments on delivering public services among EU cities. *Gov. Inf. Q.* **2005**, *22*, 217–238. [CrossRef]
9. Siau, K.; Long, Y. Using social development lenses to understand e-government development. *J. Glob. Inf. Manag.* **2006**, *14*, 47–62. [CrossRef]
10. Tolbert, C.J.; Mossberger, K. The effects of e-government on trust and confidence in government. *Public Adm. Rev.* **2006**, *66*, 354–369. [CrossRef]
11. Navarro-Galera, A.; Alcaraz-Quiles, F.J.; Ortiz-Rodriguez, D. Enhancing sustainability transparency in local governments—An empirical research in Europe. *Sustainability* **2018**, *10*, 2161. [CrossRef]
12. Macintosh, A.; Whyte, A. Towards an evaluation framework for eParticipation. *Transform. Gov. People Process Policy* **2008**, *2*, 16–30. [CrossRef]
13. Serov, I.; Leitner, M. An experimental approach to reputation in e-participation. In Proceedings of the 2016 International Conference on Software Security and Assurance (ICSSA), St. Polten, Austria, 24–25 August 2016; pp. 37–42.
14. Armah, F.A.; Yawson, D.O.; Pappoe, A.N.M.; Afrifa, E.K.A. Participation and sustainable management of coastal lagoon ecosystems: The case of the Fosu Lagoon in Ghana. *Sustainability* **2010**, *2*, 383–399. [CrossRef]
15. United Nations. *E-Government Survey 2012: E-Government for the People*; United Nations: New York, NY, USA, 2012.
16. Voß, J.-P.; Newig, J.; Kastens, B.; Monstadt, J.; Nölting, B. Steering for sustainable development: A typology of problems and strategies with respect to ambivalence, uncertainty and distributed power. *J. Environ. Policy Plan.* **2007**, *9*, 193–212. [CrossRef]
17. Renn, O.; Schweizer, P.-J. Inclusive risk governance: Concepts and application to environmental policy making. *Environ. Policy Gov.* **2009**, *19*, 174–185. [CrossRef]
18. Wellalage, N.H.; Locke, S.; Samujh, H. Corruption, gender and credit constraints: Evidence from South Asian SMEs. *J. Bus. Ethics* **2019**, *159*, 267–280. [CrossRef]
19. Masud, M.A.K.; Bae, S.M.; Manzanares, J.; Kim, J.D. Board directors' expertise and corporate corruption disclosure: The moderating role of political connections. *Sustainability* **2019**, *11*, 4491. [CrossRef]
20. Windsor, D. International business, corruption, and bribery. In *Advances in Sustainability and Environmental Justice*; Emerald Group Publishing Limited: Bingley, UK, 2013; pp. 65–95. ISBN 9781781906255.
21. Kim, E.; Ha, Y.; Kim, S. Public debt, corruption and sustainable economic growth. *Sustainability* **2017**, *9*, 433. [CrossRef]
22. Previtali, P.; Cerchiello, P. The prevention of corruption as an unavoidable way to ensure healthcare system sustainability. *Sustainability* **2018**, *10*, 3071. [CrossRef]
23. Lisciandra, M.; Migliardo, C. An empirical study of the impact of corruption on environmental performance: Evidence from panel data. *Environ. Resour. Econ.* **2017**, *68*, 297–318. [CrossRef]
24. Yale Center for Environmental Law and Policy—YCELP—Yale University; Yale Data-Driven Environmental Solutions Group—Yale University; Center for International Earth Science Information Network—CIESIN—Columbia University; World Economic Forum—WEF. 2018 Environmental Performance Index (EPI). Available online: <https://sedac.ciesin.columbia.edu/data/set/epi-environmental-performance-index-2018> (accessed on 1 May 2018).
25. United Nations (UN). E-Government Knowledge Database (UNeGovDD). Available online: <https://publicadministration.un.org/egovkb/en-us/> (accessed on 10 February 2019).
26. TI-Transparency International Corruption Perception Index. 2018. Available online: <https://www.transparency.org/cpi2018> (accessed on 10 February 2019).
27. World Bank. World Bank's Indicators Database. 2018. Available online: <https://datos.bancomundial.org/indicador> (accessed on 4 April 2018).
28. Galindo-Villardón, M.P. An alternative for simultaneous representation: HJ-biplot. *Questiò* **1986**, *10*, 12–23.
29. Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **1971**, *58*, 453–467. [CrossRef]

30. Mendes, S.; Fernández-Gómez, M.; Galindo, M.; Morgado, F.; Maranhão, P.; Azeiteiro, U.; Bacelar-Nicolau, P. The study of bacterioplankton dynamics in the Berlengas Archipelago (West coast of Portugal) by applying the HJ-biplot method. *Arquipel. Life Mar. Sci.* **2009**, *26*, 25–35.
31. Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218. [[CrossRef](#)]
32. Carrasco, G.; Molina, J.-L.; Patino-Alonso, M.-C.; Castillo, M.D.C.; Vicente-Galindo, M.-P.; Galindo-Villardón, M.-P. Water quality evaluation through a multivariate statistical HJ-biplot approach. *J. Hydrol.* **2019**, *577*, 123993. [[CrossRef](#)]
33. Cubilla-Montilla, M.; Nieto-Librero, A.-B.; Galindo-Villardón, M.P.; Vicente Galindo, M.P.; Garcia-Sanchez, I.-M. Are cultural values sufficient to improve stakeholder engagement human and labour rights issues? *Corp. Soc. Responsib. Environ. Manag.* **2019**, *26*, 938–955. [[CrossRef](#)]
34. Amor-Esteban, V.; Galindo-Villardón, M.-P.; García-Sánchez, I.-M. Industry mimetic isomorphism and sustainable development based on the X-STATIS and HJ-biplot methods. *Environ. Sci. Pollut. Res.* **2018**, *25*, 26192–26208. [[CrossRef](#)]
35. García-Sánchez, I.-M.; Frías-Aceituno, J.-V.; Rodríguez-Domínguez, L. Determinants of corporate social disclosure in Spanish local governments. *J. Clean. Prod.* **2013**, *39*, 60–72. [[CrossRef](#)]
36. Castellanos-Martín, A.; Castillo-Lluva, S.; Sáez-Freire, M.D.M.; Blanco-Gómez, A.; Hontecillas-Prieto, L.; Patino-Alonso, C.; Galindo-Villardón, P.; del Villar, L.P.; Martín-Seisdedos, C.; Isidoro-García, M.; et al. Unraveling heterogeneous susceptibility and the evolution of breast cancer using a systems biology approach. *Genome Biol.* **2015**, *16*, 40. [[CrossRef](#)]
37. Gallego-Álvarez, I.; Galindo-Villardón, M.P.; Rodríguez-Rosa, M. Analysis of the sustainable society index worldwide: A study from the biplot perspective. *Soc. Indic. Res.* **2015**, *120*, 29–65. [[CrossRef](#)]
38. Tejedor-Flores, N.; Vicente-Galindo, P.; Galindo-Villardón, P. Sustainability multivariate analysis of the energy consumption of Ecuador using MuSIASEM and BIPLLOT approach. *Sustainability* **2017**, *9*, 984. [[CrossRef](#)]
39. Nieto-Librero, A.B.; Sierra, C.; Vicente-Galindo, M.P.; Ruiz-Barzola, O.; Galindo-Villardón, M.P. Clustering Disjoint HJ-biplot: A new tool for identifying pollution patterns in geochemical studies. *Chemosphere* **2017**, *176*, 389–396. [[CrossRef](#)]
40. Hernández, M.; Espinosa, F.; Galindo, P. Tomato fruit quality as influenced by the interactions between agricultural techniques and harvesting period. *J. Plant Nutr. Soil Sci.* **2014**, *177*, 443–448. [[CrossRef](#)]
41. Garcia-Talegon, J.; Vicente, M.A.; Molina-Ballesteros, E.; Vicente-Tavera, S. Determination of the origin and evolution of building stones as a function of their chemical composition using the inertia criterion based on an HJ-biplot. *Chem. Geol.* **1999**, *153*, 37–51. [[CrossRef](#)]
42. Orfao, A.; Gonzalez, M.; san Miguel, J.F.; Cañizo, M.C.; Galindo, P.; Caballero, M.D.; Jimenez, R.; Borrasca, A.L. Clinical and immunological findings in large B-cell chronic lymphocytic leukemia. *Clin. Immunol. Immunopathol.* **1988**, *46*, 177–185. [[CrossRef](#)]
43. Des Plantes, H.L. Structuration des Tableaux à Trois Indices de la Statistique. Ph.D. Thesis, Université de Montpellier II, Montpellier, France, 1976.
44. Escoufier, Y. Le traitement des variables vectorielles. *Biometrics* **1973**, *29*, 751–760. [[CrossRef](#)]
45. Des Plantes, H.L.; Thiébaud, B. Evue de statistique appliquée. *Rev. Stat. Appl.* **1977**, *25*, 57–81.
46. Lavit, C. *Analyse Conjointe de Tableaux Quantitatifs*; Masson: Paris, France, 1988; ISBN 2225814783.
47. Jaffrenou, P.A. Sur L'Analyse des Familles Finies des Variables Vectorielles: Bases Algébrique et Application à la Description Statistique. Ph.D. Thesis, l'Université de Sainte-Etienne, Saint-Étienne, France, 1978.
48. Abdi, H.; Valentin, D.; Chollet, S.; Chrea, C. Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Qual. Prefer.* **2007**, *18*, 627–640. [[CrossRef](#)]
49. Thioulouse, J. Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *Ann. Appl. Stat.* **2011**, *5*, 2300–2325. [[CrossRef](#)]
50. Simier, M.; Blanc, L.; Pellegrin, F.; Nandris, D. Approche simultanee de k couples de tableaux: Application à l'étude des relations pathologie végétale-environnement. *Rev. Stat. Appl.* **1999**, *47*, 31–46.
51. Thioulouse, J.; Simier, M.; Chessel, D. Simultaneous analysis of a sequence of paired ecological tables. *Ecology* **2004**, *85*, 272–283. [[CrossRef](#)]
52. Bénasséni, J.; Bennani-Dosse, M. Analyzing multiset data by the Power STATIS-ACT method. *Adv. Data Anal. Classif.* **2012**, *6*, 49–65. [[CrossRef](#)]

53. Sauzay, L.; Hanafi, M.; Qannari, E.M.; Schlich, P. Analyse de K + 1 tableaux a l'aide de la méthode STATIS: Application en evaluation sensorielle. In *9ième Journées Européennes Agro-industrie et Méthodes Statistiques*; Faculté de Pharmacie: Montpellier, France, 2006; pp. 1–23.
54. Sabatier, R.; Vivien, M. A new linear method for analyzing four-way multiblocks tables: STATIS-4. *J. Chemom.* **2008**, *22*, 399–407. [[CrossRef](#)]
55. Vallejo-Arboleda, A.; Vicente-Villardón, J.L.; Galindo-Villardón, M.P.P. Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Comput. Stat. Data Anal.* **2007**, *51*, 4193–4205. [[CrossRef](#)]
56. Corrales, D.; Rodríguez, O. Interstatis: El método stasis para datos de tipo intervalo. *Rev. Mat. Teor. Apl.* **2014**. [[CrossRef](#)]
57. Llobell, F.; Cariou, V.; Vigneau, E.; Labenne, A.; Qannari, E.M. Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Qual. Prefer.* **2018**, *79*, 103520. [[CrossRef](#)]
58. Fu, Y.; Jiang, H.; Zhang, D.; Zhang, X. Comparison of perceptual differences between users and designers in mobile shopping app interface design: Implications for evaluation practice. *IEEE Access* **2019**, *7*, 23459–23470. [[CrossRef](#)]
59. Feki-Sahnoun, W.; Hamza, A.; Béjaoui, B.; Mahfoudi, M.; Rebai, A.; Hassen, M.B. Multi-table approach to assess the biogeography of phytoplankton: Ecological and management implications. *Hydrobiologia* **2018**, *815*, 229–251. [[CrossRef](#)]
60. Rundle, M.M.; Coch, D.; Connolly, A.C.; Granger, R.H. Dissociating frequency and animacy effects in visual word processing: An fMRI study. *Brain Lang.* **2018**, *183*, 54–63. [[CrossRef](#)]
61. Filho, D.M.; Fogliatto, F.S.; de Oliveira, L.P.L. Monitoring nonlinear batch process using stasis-based method. In *Applied Mechanics and Materials*; Trans Tech Publications: Stafa-Zurich, Switzerland, 2014.
62. Ramos-Barberán, M.; Hinojosa-Ramos, M.V.; Ascencio-Moreno, J.; Vera, F.; Ruiz-Barzola, O.; Galindo-Villardón, M.P. Batch process control and monitoring: A Dual STATIS and Parallel Coordinates (DS-PC) approach. *Prod. Manuf. Res.* **2018**, *6*, 470–493. [[CrossRef](#)]
63. Caballero-Juliá, D.; Galindo-Villardón, M.P.; García, M.-C. JK-Meta-Biplot y STATIS Dual como herramientas de análisis de tablas textuales múltiples. *RISTI* **2017**, *25*, 18–33. [[CrossRef](#)]
64. Bono, F.; Giacomarra, M. The photovoltaic growth in the European Union requires stronger RES support. *J. Policy Model.* **2016**, *38*, 324–339. [[CrossRef](#)]
65. Vicente-Villardón, J.L. *MULTIPILOT: A Package for Multivariate Analysis Using Biplots*; Universidad de Salamanca: Salamanca, Spain, 2010.
66. Bicking, M.; Triantafyllou, A.; Henderson, F.; Koussouris, S.; Wimmer, M.A. 2011 IST-Africa conference proceedings. In *Proceedings of the IST-Africa 2011 Conference*, Gaborone, Botswana, 11–13 May 2011; pp. 1–8.
67. Koussouris, S.; Charalabidis, Y.; Askounis, D. A review of the European Union eParticipation action pilot projects. *Transform. Gov. People Process Policy* **2011**, *5*, 8–19. [[CrossRef](#)]
68. Royo, S.; Yetano, A.; Acerete, B. E-participation and environmental protection: Are local governments really committed? *Public Adm. Rev.* **2014**, *74*, 87–98. [[CrossRef](#)]
69. Beierle, T.C.; Cayford, J. *Democracy in Practice: Public Participation in Environmental Decisions*; Resources for the Future: Washington, DC, USA, 2002; ISBN 9781891853548.
70. Newig, J.; Fritsch, O. Environmental governance: Participatory, multi-level—And effective? *Environ. Policy Gov.* **2009**, *19*, 197–214. [[CrossRef](#)]
71. Alió, M.Á.; Gallego, A. Civic entities in environmental local planning. A contribution from a participative research in the metropolitan area of Barcelona. *Geojournal* **2002**, *56*, 123–134. [[CrossRef](#)]
72. Royo, S.; Yetano, A. “Crowdsourcing” as a tool for e-participation: Two experiences regarding CO₂ emissions at municipal level. *Electron. Commer. Res.* **2015**, *15*, 323–348. [[CrossRef](#)]
73. García-Sánchez, I.M.; Rodríguez-Domínguez, L.; Gallego-Álvarez, I. The relationship between political factors and the development of e-participatory government. *Inf. Soc.* **2011**, *27*, 233–251. [[CrossRef](#)]
74. World Bank. World Bank Country and Lending Groups. © The World Bank Group. 2019. Available online: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519> (accessed on 12 February 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

ARTÍCULOS SOMETIDOS

Tabla 7. Trabajos de Investigación en proceso de Publicación

Título	Autores	Revista
SPARSE-STATIS dual: A NEW METHODOLOGY FOR ENVIRONMENTAL DATA ANALYSIS	Carmen C. Rodríguez-Martínez Mitzi I. Cubilla-Montilla Purificación Galindo-Villardón Purificación Vicente-Galindo	Ecological Indicators Indexada en: WOS Métrica JCR Cuartil: Q1 (SCIE) JIF: 4.229
THREE-MODE METHODS: A REVIEW AND RECENT DEVELOPMENTS OF THE STATIS FAMILY	Carmen C. Rodríguez-Martínez Mitzi I. Cubilla-Montilla Purificación Galindo-Villardón Purificación Vicente-Galindo	International Statistical Review Indexada en: WOS Métrica JCR Cuartil: Q1 (SCIE) JIF: 3.275
WHAT DETERMINES THE EVOLUTION OF E-PARTICIPATION?	Carmen C. Rodríguez-Martínez Purificación Galindo-Villardón Purificación Vicente-Galindo Isabel García-Sánchez	Online Information Review Indexada en: WOS Métrica JCR Cuartil: Q2 (SSCI) JIF: 1.805
CRIME STUDY IN LATIN AMERICA BASED ON NEW MULTIVARIATE DATA ANALYSIS TOOL: COSTATIS	Carmen C. Rodríguez-Martínez Alexander A. Alleyne B. Purificación Galindo-Villardón Purificación Vicente-Galindo	Crime & Delinquency Indexada en: WOS Métrica JCR Cuartil: Q1 (SSCI) JIF: 2.188

PARTICIPACIÓN EN CONGRESOS

Tabla 8. Participación en Congresos Nacionales e Internacionales

Año	Congreso	Fecha	Título de la Comunicación o Póster	Modalidad
2015	XXX Foro Internacional de Estadística. Acapulco, México	14 al 18 de septiembre-2015	Visión METABILOT de la delincuencia en América Latina: 2003-2012	Comunicación Oral
2016	XXVI Simposio Internacional de Estadística. Sucre, Colombia	8 al 12 de agosto-2016.	Tablas de Tres Vías: una aplicación con datos sobre la delincuencia.	Comunicación Oral
2017	IV Conferencia Española de Biometría. CEB 2017. Sevilla, España	13 al 15 de septiembre-2017	Multivariate profile of crime in Latin American countries	Póster
2018	I Congreso Internacional "Políticas Públicas en defensa de la inclusión, la diversidad y el género". Salamanca, España	22 al 24 de julio- 2018	Visión Multivariante de los indicadores de Género en países europeos	Comunicación Oral
2019	IV International Workshop on Proximity Data, Multivariate Analysis and Classification. USAL. España	25 y 26 de abril-2019	Multivariate profile of the E-participation based in X-STATIS method	Póster
2020	I Foro Internacional de Charlas Multidisciplinarias de Análisis de Datos	08 al 19 de Junio-2020	Aplicación del método COSTATIS a datos sobre delincuencia en América Latina.	Comunicación Oral

ANEXO

Anexo 1. Revisión bibliográfica de publicaciones con aplicación y desarrollos teóricos sobre los métodos de la Familia STATIS (1990-2020)

Numero	Repositorio	Autores	Título	Area	Año	Método
1	NUM	(L'Hermier des Plantes, 1976)	Structuration des tableaux à trois indices de la statistique	Mathematics	1976	STATIS
2	NUM	(L'Hermier des Plantes & Thiébaud, 1977)	Étude de la pluviosité au moyen de la méthode STATIS	Mathematics	1977	STATIS
3	NUM	(Escoufier, 1980)	L'analyse conjointe de plusieurs matrices de données.	Mathematics	1980	STATIS
4	NUM	(Foucart, 1982)	Programmation de la méthode STATIS sur micro-ordinateurs	Mathematics	1982	STATIS
5	NUM	(Faucart & Foucart, 1983)	Une nouvelle approche de la méthode STATIS	Mathematics	1983	STATIS
6	NUM	(Lavit, 1985)	Application de la méthode STATIS Statistique	Mathematics	1985	STATIS
7	WOS	(Kroonenberg, 1989)	The Analysis of multiple tables in factorial ecology .3. 3-Mode principal component analysis - analyse triadique complete	Environmental Sciences & Ecology	1989	X-STATIS (PTA)
8	WOS	(Vera & Carmona, 1991)	Statistical and Computational Analysis of data tables	Mathematics	1991	STATIS
9	WOS	(Génard, Souty, Holmes, Reich, & Breuils, 1994)	Correlations among quality parameters of peach fruit	Agriculture	1994	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
10	WOS	(Lavit et al., 1994)	The ACT (STATIS method)	Computer Science	1994	STATIS
11	WOS	(Pellegrin, Nandris, Westrelin, & Kohler, 1995)	Pathological status of arabica in New-Caledonia, correlations between epidemics and environment	Agriculture	1995	X-STATIS (PTA)
12	WOS	(Qannari, Wakeling, & MacFie, 1995)	A hierarchy of models for analysing sensory data	Food Science & Technology	1995	STATIS
13	WOS	(Coquet, Troxler, & Wipff, 1996)	The STATIS method: Characterization of conformational states of flexible molecules from molecular dynamics simulations in solution	Biochemistry & Molecular Biology	1996	STATIS
14	NUM	(D. Chessel & Hanafi, 1996)	Analyses de la co-inertie de K nuages de points	Mathematics	1996	STATIS
15	NUM	(Pagès, 1996)	Éléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS	Mathematics	1996	STATIS
16	WOS	(Chiarello & Chassery, 1997)	Typology of spatial structures of images having the same color set	Computer Science	1997	X-STATIS (PTA)
17	WOS	(Issanchou, Schlich, & Lesschaeve, 1997)	Sensory analysis: Methodological aspects relevant to the study of cheese	Food Science & Technology	1997	STATIS
18	NUM	(Qannari, Courcoux, Lejeune, & Maystre, 1997)	Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle	Mathematics	1997	STATIS
19	WOS	(Blanc, Chessel, & Doledec, 1998)	Study of the temporal stability of spatial structures by analysing series of totally paired faunistic tables	Fisheries	1998	X-STATIS (PTA)

Numero	Repositorio	Autores	Título	Area	Año	Método
20	WOS	(Blanc & Beaudou, 1998)	Temporal stability of spatial structures of fish community in Languedoc-Roussillon and Provence-Alpes-Cote d'Azur regions.	Fisheries	1998	X-STATIS (PTA)
21	WOS	(Gaertner et al., 1998)	Stability of spatial structures of demersal assemblages: a multitable approach	Fisheries	1998	STATIS-CoA
22	WOS	(Grossi, Raymond, Sanlaville-Boisson, & Jay, 1999)	Rosa taxonomy and hierarchy of markers defined by ACT STATIS	Biochemistry & Molecular Biology	1999	STATIS
23	WOS	(Raymond, Fiasson, & Jay, 2000)	Synthetic taxonomy of Rosa races using ACT-STATIS	Biochemistry & Molecular Biology	2000	STATIS
24	WOS	(Meyners, Kunert, & Qannari, 2000)	Comparing generalized procrustes analysis and STATIS	Food Science & Technology	2000	STATIS
25	WOS	(Gaertner, 2000)	Seasonal organization patterns of demersal assemblages in the Gulf of Lions (north-western Mediterranean Sea)	Marine & Freshwater Biology	2000	X-STATIS (PTA)
26	WOS	(Licandro & Ibanez, 2000)	Changes of zooplankton communities in the Gulf of Tigullio (Ligurian Sea, Western Mediterranean) from 1985 to 1995. Influence of hydroclimatic factors	Marine & Freshwater Biology	2000	STATIS
27	WOS	(Sicard, Sabatier, Niel, & Cadier, 2002)	A new approach in space-time analysis of multivariate hydrological data: Application to Brazil's Nordeste region rainfall	Environmental Sciences & Ecology	2002	STATIS
28	WOS	(Lekve, 2002)	Species richness and environmental conditions of fish along the Norwegian Skagerrak coast	Fisheries	2002	STATIS
29	WOS	(Anneville, Ginot, Druart, & Angeli, 2002)	Long-term study (1974-1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach	Marine & Freshwater Biology	2002	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
30	WOS	(Gaertner, Bertrand, & Souplet, 2002)	STATIS-CoA: A methodological solution to assess the spatio-temporal organization of species assemblages. Application to the demersal assemblages of the French Mediterranean Sea	Marine & Freshwater Biology	2002	STATIS-CoA
31	WOS	(Enachescu & Postelnicu, 2003)	Patterns in journal citation data revealed by exploratory multivariate analysis	Computer Science	2003	STATIS; STATIS-dual
32	WOS	(Rossi, 2003)	The spatiotemporal pattern of a tropical earthworm species assemblage and its relationship with soil structure	Environmental Sciences & Ecology	2003	X-STATIS (PTA)
33	WOS	(Muiño et al., 2003)	Consistency in the correlation of school parameters across years and stocks	Fisheries	2003	STATIS
34	WOS	(Meyners, 2003)	Methods to analyse sensory profiling data - a comparison	Food Science & Technology	2003	STATIS
35	WOS	(Gailhard, Durbec, Beliaeff, & Sabatier, 2003)	Phytoplankton ecology along French coasts: inter-site comparison	Life Sciences & Biomedicine	2003	STATIS
36	WOS	(Stanimirova et al., 2004)	STATIS, a three-way method for data analysis. Application to environmental data	Automation & Control Systems	2004	STATIS
37	WOS	(Perez-Hugalde, Judez, Litago, Yuste, & Fuentes-Pila, 2004)	Statistical procedure for clonal preselection of <i>Vitis vinifera</i> L. cv. Tempranillo in the Duero Valley, Spain	Biotechnology & Applied Microbiology	2004	STATIS-dual
38	WOS	(Oliveira & Mexia, 2004)	AIDS in Portugal: endemic versus epidemic forecasting scenarios for mortality	Business & Economics	2004	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
39	WOS	(Vivien & Sabatier, 2004)	A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables	Computer Science	2004	STATIS
40	WOS	Thioulouse, Jean(Thioulouse et al., 2004)	Simultaneous analysis of a sequence of paired ecological tables	Environmental Sciences & Ecology	2004	STATICO
41	WOS	(Chaya, Perez-Hugalde, Judez, Wee, & Guinard, 2003)	Use of the STATIS method to analyze time-intensity profiling data	Food Science & Technology	2004	STATIS
42	WOS	(Simier, Blanc, Aliaume, Diouf, & Albaret, 2004)	Spatial and temporal structure of fish assemblages in an "inverse estuary", the Sine Saloum system (Senegal)	Marine & Freshwater Biology	2004	STATIS-CoA
43	WOS	(Gourvéneq et al., 2005)	Monitoring batch processes with the STATIS approach	Automation & Control Systems	2005	STATIS
44	WOS	(Gaertner, Bertrand, Samani, & Souplet, 2005)	Spatio-temporal organization patterns of demersal assemblages of the east coast of Corsica (Mediterranean Sea)	Environmental Sciences & Ecology	2005	STATIS-CoA
45	Scopus	(Abdi et al., 2005)	DISTATIS: The Analysis of Multiple Distance Matrices	Mathematics	2005	DiSTATIS
46	WOS	(Ernault, Freiré-Diaz, Langlois, & Alard, 2006)	Are similar landscapes the result of similar histories?	Environmental Sciences & Ecology	2006	X-STATIS (PTA)
47	WOS	(Jiménez, Decaëns, & Rossi, 2006)	Stability of the spatio-temporal distribution and niche overlap in neotropical earthworm assemblages	Environmental Sciences & Ecology	2006	X-STATIS (PTA)
48	WOS	(Simier, Laurent, Ecoutin, & Albaret, 2006)	The Gambia River estuary: A reference point for estuarine fish assemblages studies in West Africa	Marine & Freshwater Biology	2006	STATICO; STATIS-CoA

Numero	Repositorio	Autores	Título	Area	Año	Método
49	WOS	(Lobry, Lepage, & Rochard, 2006)	From seasonal patterns to a reference situation in an estuarine environment: Example of the small fish and shrimp fauna of the Gironde estuary (SW France)	Marine & Freshwater Biology;	2006	STATIS
50	WOS	(Zambrano et al., 2007)	Genetic diversity of Venezuelan cassava collection	Agriculture	2007	STATIS
51	WOS	(Pérez Aparicio, Ángeles Toledano Medina, & Lafuente Rosales, 2007)	Descriptive sensory analysis in different classes of orange juice by a robust free-choice profile method	Chemistry	2007	STATIS
52	WOS	(Stanimirova et al., 2007)	Chemometric analysis of the water purification process data	Chemistry	2007	STATIS
53	WOS	(Oliveira & Mexia, 2007b)	Modelling series of studies with a common structure	Computer Science	2007	STATIS
54	WOS	(Vallejo-Arboleda et al., 2007)	Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology	Computer Science	2007	CanoSTATIS
55	WOS	(Pavoine, Blondel, Baguette, & Chessel, 2007)	A new technique for ordering asymmetrical three-dimensional data sets in ecology	Environmental Sciences & Ecology	2007	X-STATIS (PTA)
56	WOS	(Pavoine & Bailly, 2007)	New analysis for consistency among markers in the study of genetic diversity: development and application to the description of bacterial diversity	Evolutionary Biology	2007	STATIS
57	U (University of Texas at Dallas)	(Abdi & Valentin, 2007)	Some new and easy ways to describe, compare and evaluate products and assessors	Food Science & Technology	2007	DiSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
58	WOS	(Abdi et al., 2007)	Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications	Food Science & Technology	2007	DiSTATIS
59	WOS	(Carassou & Ponton, 2007)	Spatio-temporal structure of pelagic larval and juvenile fish assemblages in coastal areas of New Caledonia, southwest Pacific	Marine & Freshwater Biology	2007	STATICo
60	WOS	(Oliveira & Mexia, 2007a)	ANOVA-like analysis of matched series of studies with a common structure	Mathematics	2007	STATIS
61	WOS	(Lera M. et al., 2007)	Multivariate study of the evolution of body dimensions in elderly Chilean women	Nutrition & Dietetics	2007	STATIS
62	WOS	(Sabatier & Vivien, 2008)	A new linear method for analyzing four-way multiblocks tables: STATIS-4	Automation & Control Systems	2008	STATIS-4
63	WOS-SciELO	(Fogliatto & Niang, 2008)	Multivariate statistical control of unsynchronized batch processes	Engineering	2008	STATIS
64	WOS	(Serghini et al., 2008)	Stability of the spatial structures of demersal assemblage in the Moroccan southern Atlantic zone	Environmental Sciences & Ecology	2008	STATIS
65	WOS	(Lelièvre, Chollet, Abdi, & Valentin, 2008)	What is the validity of the sorting task for describing beers? A study using trained and untrained assessors	Food Science & Technology	2008	DiSTATIS
66	NUM	(Lazraq, Hanafi, Cléroux, Allaire, & Lepage, 2008)	Une approche inférentielle pour la validation du compromis de la méthode STATIS	Mathematics	2008	STATIS
67	WOS	(Vallejo-Arboleda et al., 2008)	Analysis of Time Evolution for Group Structured Data: Canonical Dual STATIS and Doubly Multivariate Repeated Measures Model	Mathematics	2008	CanoSTATIS-dual

Numero	Repositorio	Autores	Título	Area	Año	Método
68	WOS	(Arcidiacono, Sarnacchiaro, & Velleman, 2008)	Testing fidelity to a new psychological intervention for family members of substance misusers during implementation in Italy	Substance Abuse	2008	STATIS
69	WOS	(Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009)	A structured overview of simultaneous component based data integration	Biochemistry & Molecular Biology	2009	STATIS
70	WOS	(Barros, Girard, Kubrusly, & Saboia, 2009)	Recent Labor Market Patterns in Brazilian Metropolises: Changes and Differentiations	Business & Economics	2009	STATIS
71	WOS	(Rudnitskaya et al., 2009)	Instrumental measurement of beer taste attributes using an electronic tongue	Chemistry	2009	STATIS
72	WOS	(Fogliatto & Niang, 2009)	Multivariate statistical control of batch processes with variable duration	Engineering	2009	STATIS
73	WOS	(Rolland, Bertrand, Maumy, & Jacquet, 2009)	Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis	Engineering	2009	X-STATIS (PTA)
74	WOS	(Cadoret, Lê, & Pagès, 2009)	A Factorial Approach for Sorting Task data (FAST)	Food Science & Technology	2009	DiSTATIS
75	WOS	(Lelièvre, Chollet, Abdi, & Valentin, 2009)	Beer-Trained and Untrained Assessors Rely More on Vision than on Taste When They Categorize Beers	Food Science & Technology	2009	DiSTATIS
76	WOS	(Vivien & Sune, 2009)	Two four-way multiblock methods used for comparing two consumer panels of children	Food Science & Technology	2009	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
77	WOS	(Ballester, Abdi, Langlois, Peyron, & Valentin, 2009)	The Odor of Colors: Can Wine Experts and Novices Distinguish the Odors of White, Red, and Ros, Wines?	Food Science & Technology;	2009	DiSTATIS
78	WOS	(Fournier et al., 2009)	Investigation of Transport Processes inside Karst Aquifer by Means of STATIS	Geology	2009	STATIS
79	WOS	(Mendes et al., 2009)	Spatio-temporal structure of diatom assemblages in a temperate estuary. A STATICO analysis	Marine & Freshwater Biology	2009	STATICO
80	U (Conservatoire National des Arts et Métiers)	(Niang et al., 2009)	Contrôle multivarié de procédés par lots à l'aide de Statis	Mathematics	2009	STATIS
81	WOS-SciELO	(González, Ávila, & Olaya, 2009)	Analysis of the AFM and STATIS Representation Spaces in the Bogota's 1993 Biographical Mobility Survey	Mathematics	2009	STATIS
82	WOS	(Abdi, Dunlop, & Williams, 2009)	How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS)	Neurosciences & Neurology;	2009	DiSTATIS
83	WOS	(Regnault et al., 2009)Re	Satisfactory cross-cultural validity of the ACTG symptom distress module in HIV-1-infected antiretroviral-naive patients	Research & Experimental Medicine	2009	STATIS
84	WOS	(Stefanescu, Serban, Busu, & Ferrara, 2010)	Portfolio optimization using classification and functional data analysis techniques	Business & Economics	2010	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
85	WOS	(Santosa, Abdi, & Guinard, 2010)	A modified sorting task to investigate consumer perceptions of extra virgin olive oils	Food Science & Technology	2010	DiSTATIS
86	WOS	(Shieh & Chi, 2010)	Factors influencing macroinvertebrate assemblages in artificial subtropical ponds of Taiwan	Marine & Freshwater Biology	2010	STATICO
87	WOS	(Daszykowski & Walczak, 2011)	Methods for the exploratory analysis of two-dimensional chromatographic signals	Chemistry	2011	STATIS
88	WOS	(Stanimirova, Boucon, & Walczak, 2011)	Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction	Chemistry	2011	STATIS
89	WOS-SciELO	(Marcondes Filho et al., 2011)	Multivariate control charts for monitoring non-linear batch processes	Engineering	2011	Kernel-STATIS
90	WOS	(Marques, Pardal, Mendes, & Azeiteiro, 2011)	Using multitable techniques for assessing the temporal variability of species-environment relationship in a copepod community from a temperate estuarine ecosystem	Environmental Sciences & Ecology	2011	STATICO
91	WOS	(Mendes et al., 2011)	Zooplankton distribution in a marine protected area: The berlengas natural reserve (western coast of Portugal)	Environmental Sciences & Ecology	2011	X-STATIS (PTA)
92	WOS	(Oszwald, Gond, Dolédec, & Lavelle, 2011)	Identification of land-use indicators to assess changes in landscape mosaics	Forestry	2011	STATIS
93	WOS	(Thioulouse, 2011)	Simultaneous analysis of a sequence of paired ecological tables: a comparison of several methods	Mathematics	2011	STATICO; CoSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
94	WOS	(Gudmundsson et al., 2011)	Spatial cross-correlation patterns of European low, mean and high flows	Water Resources	2011	STATIS
95	WOS	(Baumberger et al., 2012)	Plant community changes as ecological indicator of seabird colonies' impacts on Mediterranean Islands	Biodiversity & Conservation	2012	STATICO
96	Scopus	(Figueiredo et al., 2012)	Restructuring in privatised firms: A Statis approach	Business & Economics	2012	STATIS
97	WOS	(Eros et al., 2012)	Temporal variability in the spatial and environmental determinants of functional metacommunity organization - stream fish in a human-modified landscape	Environmental Sciences & Ecology	2012	X-STATIS (PTA)
98	WOS	(Blancher et al., 2012)	A method to investigate the stability of a sorting map	Food Science & Technology	2012	DiSTATIS
99	WOS	(Napoléon et al., 2012)	Spatiotemporal dynamics of physicochemical and photosynthetic parameters in the central English Channel	Marine & Freshwater Biology	2012	X-STATIS (PTA)
100	WOS	(Abdi et al., 2012)	STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling	Mathematics	2012	STATIS; STATIS-dual; DiSTATIS; X-STATIS(PTA); CovSTATIS; CanoSTATIS; (k+1)STATIS; STATIS4; ANISOSTATIS; Power STATIS; Double STATIS
101	WOS	(Areia & Oliveira, 2012)	Transversal Analysis in Matched Series of Studies	Mathematics	2012	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
102	WOS	(Bénasséni & Bennani-Dosse, 2012)	Analyzing multiset data by the Power STATIS-ACT method	Mathematics	2012	Power-STATIS
103	WOS	(Mendes et al., 2012)	An empirical comparison of Canonical Correspondence Analysis and STATICO in the identification of spatio-temporal ecological relationships	Mathematics	2012	STATICO
104	WOS	(Churchill et al., 2012)W.	Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods	Neurosciences & Neurology	2012	DiSTATIS
105	WOS	(Shinkareva et al., 2012)	Exploring commonalities across participants in the neural representation of objects	Neurosciences & Neurology	2012	STATIS
106	WOS	(Mazzocchi et al., 2012)	Stability and resilience in coastal copepod assemblages: The case of the Mediterranean long-term ecological research at Station MC (LTER-MC)	Oceanography	2012	STATICO
107	WOS	(Klie et al., 2012)	Compromise of multiple time-resolved transcriptomics experiments identifies tightly regulated functions	Plant Sciences	2012	STATIS; STATIS-dual
108	WOS	(Gonçalves et al., 2012)	Responses of Copepoda Life-History Stages to Climatic Variability in a Southern-European Temperate Estuary	Zoology	2012	STATICO
109	WOS	(Godefroid et al., 2013)	Spatial stability of a plant-feeding nematode community in relation to macro-scale soil properties	Agriculture	2013	STATIS-CoA
110	WOS	(Pizarro et al., 2013)	A sensory approach for the monitoring of accelerated red wine aging processes using multi-block methods	Food Science & Technology	2013	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
111	WOS	(Gourdol et al., 2013)	On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry	Geochemistry & Geophysics	2013	X-STATIS (PTA)
112	WOS	(Areia & Oliveira, 2013)	Longitudinal Analysis for Matched Series of Studies	Mathematics	2013	STATIS
113	WOS	(Kissita et al., 2013)	Canonical correlation analysis between two vertical multi-block tables and applications to qualitative variables	Mathematics	2013	STATICO
114	WOS	(Krishnan et al., 2013)	Distance-Based Partial Least Squares Analysis	Mathematics	2013	DiSTATIS
115	WOS	(Niang et al., 2013)	Non parametric on-line control of batch processes based on STATIS and clustering	Mathematics	2013	STATIS
116	WOS	(Sabatier et al., 2013)	A new proposal, the Multitableaux Discriminating analysis: STATIS-LDA	Mathematics	2013	STATIS-LDA
117	WOS	(Ferraro et al., 2013)	Multi-marker network in ST-elevation myocardial infarction patients undergoing primary percutaneous coronary intervention: When and what to measure	Medical Laboratory Technology	2013	STATIS
118	WOS	(Derek Beaton et al., 2014)	An ExPosition of multivariate analysis with the singular value decomposition in R	Computer Science	2014	STATIS; STATIS-dual; DiSTATIS; X-STATIS(PTA); CovSTATIS; CanoSTATIS; (k+1)STATIS; STATIS4; ANISOSTATIS; Power STATIS; Double STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
119	WOS	(Pagès, 2014)	Multiple Factor Analysis: General Presentation and Comparison with STATIS	Computer Science	2014	STATIS
120	WOS	(Marcondes Filho et al., 2014)	Monitoring Nonlinear Batch Process Using Statis-based Method	Engineering; Mechanics	2014	STATIS
121	WOS	(Hernández-Fariñas et al., 2014)	Temporal changes in the phytoplankton community along the French coast of the eastern English Channel and the southern Bight of the North Sea	Fisheries	2014	X-STATIS (PTA)
122	WOS	(Mielby et al., 2014)	Comparison of descriptive analysis, projective mapping and sorting performed on pictures of fruit and vegetable mixes	Food Science & Technology	2014	DiSTATIS
123	WOS	(da Silva & Ramos, 2014a)	On the rate of convergence of uniform approximations for sequences of distribution functions	Mathematics	2014	STATIS
124	WOS	(da Silva & Ramos, 2014b)	Uniform approximations for distributions of continuous random variables with application in dual STATIS method	Mathematics	2014	STATIS-dual
125	WOS-SciELO	(Corrales & Rodríguez, 2014)	Interstatis: the statis method for interval valued data	Mathematics	2014	InterSTATIS
126	WOS	(Shinkareva et al., 2014)	Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data	Neurosciences & Neurology	2014	STATIS
127	WOS	(Viswanathan et al., 2014)	Quantitative comparison between crowd models for evacuation planning and evaluation	Physics	2014	DiSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
128	WOS	(Klie et al., 2014)	Conserved Changes in the Dynamics of Metabolic Processes during Fruit Development and Ripening across Species	Plant Sciences	2014	STATIS
129	WOS	(Rossi et al., 2014)	Dissecting the Space-Time Structure of Tree-Ring Datasets Using the Partial Triadic Analysis	Science & Technology	2014	X-STATIS (PTA)
130	WOS	(Schwarz et al., 2015)	A Read-Aloud Storybook Selection System for Prereaders at the Preschool Language Level: A Pilot Study	Audiology & Speech-Language Pathology	2015	DiSTATIS
131	WOS	(Gallego-Álvarez et al., 2015)	Evolution of sustainability indicator worldwide: A study from the economic perspective based on the X-STATICO method	Biodiversity & Conservation	2015	X-STATIS (PTA); X-STATICO
132	WOS	(Darwiche-Criado et al., 2015)	Identifying spatial and seasonal patterns of river water quality in a semiarid irrigated agricultural Mediterranean basin	Environmental Sciences & Ecology	2015	X-STATIS (PTA)
133	WOS	(Jiménez et al., 2015)	A Methodological Approach for Spatiotemporally Analyzing Water-Polluting Effluents in Agricultural Landscapes Using Partial Triadic Analysis	Environmental Sciences & Ecology	2015	X-STATIS (PTA)
134	WOS	(Lemesle et al., 2015)	Impact of seaweed beachings on dynamics of delta N-15 isotopic signatures in marine macroalgae	Environmental Sciences & Ecology	2015	X-STATIS (PTA)
135	WOS	(Masmoudi et al., 2015)	Salinity gradient and nutrients as major structuring factors of the phytoplankton communities in salt marshes	Environmental Sciences & Ecology	2015	STATICO
136	WOS	(Ladhar et al., 2015)	Strong and stable environmental structuring of the zooplankton communities in interconnected salt ponds	Marine & Freshwater Biology	2015	CoSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
137	WOS	(Boumaza et al., 2015)	Interpreting the Principal Component Analysis of Multivariate Density Functions	Mathematics	2015	STATIS-dual
138	WOS	(Kidé et al., 2015)	Spatio-Temporal Dynamics of Exploited Groundfish Species Assemblages Faced to Environmental and Fishing Forcings: Insights from the Mauritanian Exclusive Economic Zone	Science & Technology	2015	STATICO
139	WOS	(Bono & Giacomarra, 2016)	The photovoltaic growth in the European Union requires stronger RES support	Business & Economics	2016	STATIS
140	WOS	(Lahne et al., 2016)	Replication Improves Sorting-Task Results Analyzed by DISTATIS in a Consumer Study of American Bourbon and Rye Whiskeys	Food Science & Technology	2016	DISTATIS
141	WOS	(Ren et al., 2016)	Spatial and temporal assessment of the initial pattern of phytoplankton population in a newly built coastal reservoir	Geology	2016	X-STATIS (PTA)
142	WOS	(Guesdon et al., 2016)	Effect of local hydroclimate on phytoplankton groups in the Charente estuary	Marine & Freshwater Biology	2016	X-STATIS;STATICO
143	U (Toulouse University)	(Villa-Vialaneix et al., 2016)	Multiway-SIR for Longitudinal Multi-table Data Integration Sommaire	Mathematics	2016	SIR-STATIS
144	WOS	(Amerise & Tarsitano, 2016)	Combining dissimilarity matrices by using rank correlations	Mathematics	2016	DiSTATIS
145	WOS	(Connolly et al., 2016)	How the Human Brain Represents Perceived Dangerousness or "Predacity" of Animals	Neurosciences & Neurology	2016	STATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
146	WOS	(Rivadeneira et al., 2016)	Analysis of well-being in OECD countries through STATIS methodology	Science & Technology	2016	STATIS
147	WOS	(Marcondes Filho et al., 2017)	Multivariate quality control of batch processes using STATIS	Automation & Control Systems	2017	STATIS
148	WOS	(Nguyen & Holmes, 2017)	Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations	Biochemistry & Molecular Biology	2017	DiSTATIS
149	WOS	(Amor-Esteban et al., 2017)	Cultural values on CSR patterns and evolution: A study from the biplot representation	Biodiversity & Conservation	2017	X-STATIS
150	WOS	(Antonio Duarte Santos et al., 2017)	The Dynamics between Economic Growth and Living Standards in EU Countries: A STATICO Approach for the Period 2006-2014	Business & Economics	2017	STATICO
151	WOS-SciELO	(Caballero-Juliá et al., 2017)	JK-Meta-Biplot and STATIS-dual as multiple textual tables tools	Computer Science	2017	STATIS-dual
152	WOS	(Lopes et al., 2017)	Knowledge transfer through journals	Education & Educational Research	2017	STATIS
153	WOS-SciELO	(Olivares et al., 2017)	The STATIS-ACT method applied to rainfall in the Venezuelan Eastern Region	Education & Educational Research	2017	STATIS
154	WOS	(Slimani et al., 2017)	The use of STATICO and COSTATIS, two exploratory three-ways analysis methods: an application to the ecology of aquatic heteroptera in the Medjerda watershed (Tunisia)	Environmental Sciences & Ecology	2017	STATICO; CoSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
155	WOS	(Gómez-Corona et al., 2017)	The role of gender and product consumption in the mental representation of industrial and craft beers: An exploratory study with Mexican consumers	Food Science & Technology	2017	DiSTATIS
156	WOS	(Castillo-Escrivà et al., 2017)	Metacommunity dynamics of Ostracoda in temporary lakes: Overall strong niche effects except at the onset of the flooding period	Marine & Freshwater Biology	2017	X-STATIS (PTA)
157	WOS	(Mendes et al., 2017)	CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables	Mathematics	2017	STATICO
158	WOS	(Rodríguez-Rosa et al., 2017)	Are Social, Economic and Environmental Well-Being Equally Important in all Countries Around the World? A Study by Income Levels	Social Sciences	2017	X-STATIS (PTA); CoSTATIS
159	WOS	(Rundle et al., 2018)	Dissociating frequency and animacy effects in visual word processing: An fMRI study	Audiology & Speech-Language Pathology	2018	STATIS
160	WOS	(Amor-Esteban, Galindo-Villardón, & David, 2018)	Study of the Importance of National Identity in the Development of Corporate Social Responsibility Practices: A Multivariate Vision	Business & Economics	2018	X-STATICO
161	WOS	(Ramos-Barberán et al., 2018)	Batch process control and monitoring: a Dual STATIS and Parallel Coordinates (DS-PC) approach	Engineering	2018	STATIS-dual
162	WOS	(Amor-Esteban, Galindo-Villardón, & García-Sánchez, 2018)	Industry mimetic isomorphism and sustainable development based on the X-STATIS and HJ-biplot methods	Environmental Sciences & Ecology	2018	X-STATIS (PTA)

Numero	Repositorio	Autores	Título	Area	Año	Método
163	WOS	(Brand et al., 2018)	Sorting in Combination with Quality Scoring: A Tool for Industry Professionals to Identify Drivers of Wine Quality Rapidly	Food Science & Technology	2018	DiSTATIS
164	WOS	(Lahne et al., 2018)	Rapid sensory profiles with DISTATIS and Barycentric Text Projection: An example with amari, bitter herbal liqueurs	Food Science & Technology	2018	DiSTATIS
165	WOS	(Feki-Sahnoun et al., 2018)	Multi-table approach to assess the biogeography of phytoplankton: ecological and management implications	Marine & Freshwater Biology	2018	STATIS
166	WOS	(Rodrigues et al., 2018)	Use of phytoplankton functional groups as a model of spatial and temporal patterns in reservoirs: a case study in a reservoir of central Brazil	Marine & Freshwater Biology	2018	STATICO
167	U (University of Texas at Dallas)	(Kriegsman, 2018a)	Discriminant DISTATIS. A Multi.Way Discriminant Analysis for Distance Matrices.	Mathematics	2018	DiDiSTATIS;HiDiSTATIS
168	WOS	(Corduas & Ragozini, 2018)	Comparing Multistep Ahead Forecasting Functions for Time Series Clustering	Mathematics	2018	DiSTATIS
169	WOS	(Thabet et al., 2018)	Interannual and seasonal effects of environmental factors on the zooplankton distribution in the solar saltern of Sfax (south-western Mediterranean sea)	Oceanography	2018	STATICO
170	WOS	(Gallego-Álvarez et al., 2018)	An analysis of the environmental information in international companies according to the new GRI standards	Science & Technology	2018	X-STATIS (PTA)
171	WOS	(Amor-Esteban, García-Sánchez, et al., 2018)	Analysing the Effect of Legal System on Corporate Social Responsibility (CSR) at the Country Level, from a Multivariate Perspective	Social Sciences	2018	X-STATIS (PTA)

Numero	Repositorio	Autores	Título	Area	Año	Método
172	WOS	(Giordano et al., 2019)	Analyzing multiplex networks using factorial methods	Anthropology	2019	DiSTATIS
173	WOS	(Nguyen & Holmes, 2019)	Ten quick tips for effective dimensionality reduction	Biochemistry & Molecular Biology	2019	STATIS; DiSTATIS
174	U (Salamanca University)	(Amor-Esteban et al., 2019)	A Multivariate vision of the industry Specificity as an explanatory factor of companies' social performance	Business & Economics	2019	X-STATIS (PTA)
175	WOS	(Volkova, 2019)	Russian and European Population's Quality of Life Analysis with the Instruments of Common Principal Components (CPC)	Business & Economics	2019	STATIS
176	WOS	(Fu et al., 2019)	Comparison of Perceptual Differences Between Users and Designers in Mobile Shopping App Interface Design: Implications for Evaluation Practice	Computer Science	2019	DiSTATIS
177	WOS	(António Duarte Santos et al., 2019)	A COSTATIS approach to business sustainability in turbulent environments from 2008 to 2014	Engineering	2019	CoSTATIS
178	WOS	(Mabrouki et al., 2019)	Biotypology of stream macroinvertebrates from North African and semi arid catchment: Oued Za (Morocco)	Fisheries	2019	X-STATIS (PTA)
179	WOS	(Llobell, Cariou, et al., 2019)	A new approach for the analysis of data and the clustering of subjects in a CATA experiment	Food Science & Technology	2019	STATIS
180	WOS	(Llobell, Vigneau, & Qannari, 2019)	Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics	Food Science & Technology	2019	STATIS; CluSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
181	WOS	(Liu et al., 2019)	Quantifying the Spatio-Temporal Process of Township Urbanization: A Large-Scale Data-Driven Approach	Physical Geography	2019	X-STATIS (PTA)
182	WOS	(Rodríguez-Martínez et al., 2019)	Exploring Relationships between Environmental Performance, E-Government and Corruption: A Multivariate Perspective	Science & Technology	2019	STATIS
183	WOS	(Sáinz et al., 2019)	Influence of cultivar, irrigation, ripening stage, and annual variability on the oxidant/antioxidant systems of olives as determined by MDS-PTA	Science & Technology	2019	MDS-X-STATIS (MDS-PTA)
184	WOS	(Mérigot et al., 2019)	Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern mediterranean sea	Marine & Freshwater Biology	2019	STATICO-CoA
185	WOS-MedLine	(Zingaretti et al., 2020)	Link-HD: a versatile framework to explore and integrate heterogeneous microbial communities.	Bioinformática	2020	Link-HD (STATIS Generalización)
186	WOS	(Ceron et al., 2020)	Climatic variables influence the temporal dynamics of an anuran metacommunity in a nonstationary way	Environmental Sciences & Ecology	2020	STATICO
187	WOS	(Honoré-Chedozeau et al., 2020)	From perceptual to conceptual categorization of wines: What is the effect of expertise?	Food Science & Technology	2020	DiSTATIS
188	WOS	(Llobell et al., 2020)	Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics	Food Science & Technology	2020	STATIS; CluSTATIS

Numero	Repositorio	Autores	Título	Area	Año	Método
189	WOS	(Lahne, 2020)	Assessment of instructions on panelist cognitive framework and free sorting task results: A case study of cold brew coffee	Food Science & Technology	2020	DiSTATIS;HiDiSTATIS
190	WOS	(Lahne, 2020)	Sorting backbone analysis: A network-based method of extracting key actionable information from free-sorting task results	Food Science & Technology	2020	DiSTATIS
191	WOS	(Raman et al., 2020)	Bach, Mozart, and Beethoven: Sorting piano excerpts based on perceived similarity using DiSTATIS	Psychology	2020	DiSTATIS
192	WOS	(Mérillet et al., 2020)	Environment outweighs the effects of fishing in regulating demersal community structure in an exploited marine ecosystem	Environmental Sciences & Ecology	2020	STATICO
193	WOS	(Schwarz et al., 2020)	Storybook selection criteria used by teachers of d/Deaf and hard-of-hearing prereaders communicating in English	Education & Educational Research	2020	DiSTATIS

BIBLIOGRAFÍA

- Abascal, E., & Landaluce, M. I. (2000). Métodos factoriales de análisis de tablas múltiples como técnicas de validación de los resultados de un análisis de componentes principales. *XIV Reunión ASEPELT*, 21. <https://doi.org/84-699-2357-9>
- Abdi, H., Dunlop, J. P., & Williams, L. J. (2009). How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage*, 45(1), 89-95. <https://doi.org/10.1016/j.neuroimage.2008.11.008>
- Abdi, H., O'Toole, A. J., Valentin, D., & Edelman, B. (2005). DISTATIS: The Analysis of Multiple Distance Matrices. En *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops* (Vol. 3, pp. 42-42). IEEE. <https://doi.org/10.1109/CVPR.2005.445>
- Abdi, H., & Valentin, D. (2007). Some New and Easy Ways to Describe, Compare and Evaluate Products and Assessors. En D. Valentin, D. Z. Nguyen, & L. Pelletier (Eds.), *SPICE 2007* (pp. 3-16). Ho Chi Minh, Vietnam: Vietnam National University-Ho Chi Minh City Publishing House.
- Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, 18(4), 627-640. <https://doi.org/10.1016/j.foodqual.2006.09.003>
- Abdi, H., & Williams, L. J. (2010). Barycentric Discriminant analysis (BADIA). En N. J. Salkind, D. M. Dougherty, & B. Frey (Eds.), *Encyclopedia of Research Design* (Thousand O, pp. 64-75). SAGE Publications Inc.
- Abdi, H., Williams, L. J., & Béra, M. (2018). Barycentric Discriminant Analysis. En R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining* (pp. 121-140). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-7131-2_110192
- Abdi, H., Williams, L. J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 124-167. <https://doi.org/10.1002/wics.198>
- Acar, E., Çamtepe, S. A., Krishnamoorthy, M. S., & Yener, B. (2005). Modeling and Multiway Analysis of Chatroom Tensors. En *Intelligence and Security Informatics* (pp. 256-268). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11427995_21
- Amerise, I. L., & Tarsitano, A. (2016). Combining dissimilarity matrices by using rank correlations. *Computational Statistics*, 31(1), 353-367. <https://doi.org/10.1007/s00180-015-0590-x>
- Amor-Esteban, V., Galindo-Villardón, M. P., & David, F. (2018). Study of the Importance of

- National Identity in the Development of Corporate Social Responsibility Practices: A Multivariate Vision. *Administrative Sciences*, 8(3), 50.
<https://doi.org/10.3390/admsci8030050>
- Amor-Esteban, V., Galindo-Villardón, M. P., & David, F. (2019). A Multivariate Vision of the industry specificity as an explanatory factor of companies' social performance. En I.-M. García-Sánchez & J. Martínez-Ferrero (Eds.), *The disclosure and assurance of corporate social responsibility : a growing market* (pp. 1-36). Cambridge Scholars Publishing.
- Amor-Esteban, V., Galindo-Villardón, M. P., & García-Sánchez, I.-M. (2018). Industry mimetic isomorphism and sustainable development based on the X-STATIS and HJ-biplot methods. *Environmental Science and Pollution Research*, 25(26), 26192-26208.
<https://doi.org/10.1007/s11356-018-2663-1>
- Amor-Esteban, V., Galindo-Villardón, M. P., & Sánchez, I. M. G. (2017). Cultural values on CSR patterns and evolution: A study from the biplot representation. *Ecological Indicators*, 81, 18-29. <https://doi.org/10.1016/j.ecolind.2017.05.051>
- Amor-Esteban, V., García-Sánchez, I.-M., & Galindo-Villardón, M. P. (2018). Analysing the Effect of Legal System on Corporate Social Responsibility (CSR) at the Country Level, from a Multivariate Perspective. *Social Indicators Research*, 140(1), 435-452.
<https://doi.org/10.1007/s11205-017-1782-2>
- Anneville, O., Ginot, V., Druart, J. C., & Angeli, N. (2002). Long-term study (1974-1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach. *Journal of Plankton Research*, 24(10), 993-1008. <https://doi.org/10.1093/plankt/24.10.993>
- Arcidiacono, C., Sarnacchiaro, P., & Velleman, R. (2008). Testing fidelity to a new psychological intervention for family members of substance misusers during implementation in Italy. *Journal of Substance Use*, 13(6), 361-381. <https://doi.org/10.1080/14659890802211291>
- Areia, A., & Oliveira, M. M. (2012). Transversal analysis in matched series of studies. En *AIP Conference Proceedings* (pp. 1670-1673). <https://doi.org/10.1063/1.4756490>
- Areia, A., & Oliveira, M. M. (2013). Longitudinal analysis for matched series of studies. En *AIP Conference Proceedings*. <https://doi.org/10.1063/1.4825622>
- Baccalá, N. (2004). *Contribuciones al análisis de matrices de datos multivía: Tipología de las variables*. *Contribuciones al análisis de matrices de datos Multivía*. Universidad de Salamanca. Recuperado de <https://scholar.google.es/scholar?hl=es&q=contribuciones+al+análisis+de+matrices+de+datos+multivía&btnG=&lr=#0>
- Bader, B. W., & Kolda, T. G. (2006). Algorithm 862: MATLAB Tensor Classes for Fast Algorithm Prototyping. *ACM Transactions on Mathematical Software*, 32(4), 635-653. Recuperado

de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.9005&rep=rep1&type=pdf>

Bahram, M., & Bro, R. (2007). A novel strategy for solving matrix effect in three-way data using parallel profiles with linear dependencies. *Analytica Chimica Acta*, 584(2), 397-402.

<https://doi.org/10.1016/j.aca.2006.11.070>

Ballester, J., Abdi, H., Langlois, J., Peyron, D., & Valentin, D. (2009). The Odor of Colors: Can Wine Experts and Novices Distinguish the Odors of White, Red, and Rosé Wines?

Chemosensory Perception, 2(4), 203-213. <https://doi.org/10.1007/s12078-009-9058-0>

Barros, A. C., Girard, J.-L., Kubrusly, L. S., & Saboia, J. (2009). Les trajectoires récentes du marché du travail dans les métropoles du Brésil : changements et différenciations. *Revue d'Économie Régionale & Urbaine*, novembre(4), 635.

<https://doi.org/10.3917/reru.094.0635>

Basso Luiz, C. (2006). *Análisis conjunto de varias matrices de datos: contribuciones a la tipología de los individuos*. Universidad de Salamanca. Recuperado de

<https://dialnet.unirioja.es/servlet/tesis?codigo=19861>

Baumberger, T., Affre, L., Torre, F., Vidal, E., Dumas, P.-J., & Tatoni, T. (2012). Plant community changes as ecological indicator of seabird colonies' impacts on Mediterranean Islands. *Ecological Indicators*, 15(1), 76-84. <https://doi.org/10.1016/j.ecolind.2011.09.009>

Beaton, D., Fatt, C. C., & Abdi, H. (2013). DistatisR: DiSTATIS three way metric multidimensional scaling.

Beaton, Derek, Chin Fatt, C. R., & Abdi, H. (2014). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics and Data Analysis*, 72,

176-189. <https://doi.org/10.1016/j.csda.2013.11.006>

Bénasséni, J., & Bennani-Dosse, M. (2012). Analyzing multiset data by the Power STATIS-ACT method. *Advances in Data Analysis and Classification*, 6(1), 49-65.

<https://doi.org/10.1007/s11634-011-0085-8>

Benzécri, J. P. (1973). *L'Analyse des données. vol. 2, l'analyse des correspondances*. Paris, France.

Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 39(227), 357-365.

<https://doi.org/10.1080/01621459.1944.10500699>

Bioconductor. (2003). Bioinformatics: Open source software for Bioinformatics. Recuperado de

<https://www.bioconductor.org/>

- Blanc, L., & Beaudou, D. (1998). Stabilité temporelle des structures spatiales des peuplements piscicoles des régions Languedoc-Roussillon et Provence-Alpes-Côte d'Azur. *Bulletin Français de la Pêche et de la Pisciculture*, (348), 23-45.
<https://doi.org/10.1051/kmae:1998050>
- Blanc, L., Chessel, D., & Doledec, S. (1998). Etude de la stabilité temporelle des structures spatiales par analyses d'une série de tableaux de relevés faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture*, (348), 1-21.
<https://doi.org/10.1051/kmae:1998049>
- Blancher, G., Clavier, B., Egoroff, C., Duineveld, K., & Parcon, J. (2012). A method to investigate the stability of a sorting map. *Food Quality and Preference*, 23(1), 36-43.
<https://doi.org/10.1016/j.foodqual.2011.06.010>
- Bono, F., & Giacomarra, M. (2016). The photovoltaic growth in the European Union requires stronger RES support. *Journal of Policy Modeling*, 38(2), 324-339.
<https://doi.org/10.1016/j.jpolmod.2016.01.003>
- Bougeard, S., & Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software*, 86, 1-17. <https://doi.org/10.18637/jss.v086.i01>
- Boumaza, R., Santagostini, P., Yousfi, S., Hunault, G., Bourbeillon, J., Pumo, B., & Demotes-Mainard, S. (2019). Package 'dad'. Three-Way/Multigroup Data Analysis Through Densities.
- Boumaza, R., Yousfi, S., & Demotes-Mainard, S. (2015). Interpreting the Principal Component Analysis of Multivariate Density Functions. *Communications in Statistics - Theory and Methods*, 44(16), 3321-3339. <https://doi.org/10.1080/03610926.2013.824103>
- Brand, J., Kidd, M. c., van Antwerpen, L., Valentin, D., Naes, T., & Nieuwoudt, H. H. (2018). Sorting in Combination with Quality Scoring: A Tool for Industry Professionals to Identify Drivers of Wine Quality Rapidly. *South African Journal of Enology and Viticulture*, 39(2), 163-175. <https://doi.org/10.21548/39-2-3203>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. *Biometrics*, 40(3), 874. <https://doi.org/10.2307/2530946>
- Caballero-Juliá, D., Galindo-Villardón, M. P., & García, M.-C. (2017). JK-Meta-Biplot y STATIS Dual como herramientas de análisis de tablas textuales múltiples. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 25, 18-33. <https://doi.org/10.17013/risti.25.18-33>
- Cadima, J., & Jolliffe, I. T. (1995). Loading and correlations in the interpretation of principle components. *Journal of Applied Statistics*, 22(2), 203-214.
<https://doi.org/10.1080/757584614>

- Cadoret, M., Lê, S., & Pagès, J. (2009). A Factorial Approach for Sorting Task data (FAST). *Food Quality and Preference*, 20(6), 410-417.
<https://doi.org/10.1016/j.foodqual.2009.02.010>
- Carassou, L., & Ponton, D. (2007). Spatio-temporal structure of pelagic larval and juvenile fish assemblages in coastal areas of New Caledonia, southwest Pacific. *Marine Biology*.
<https://doi.org/10.1007/s00227-006-0389-y>
- Carlier, A., & Kroonenberg, P. M. (1996). Decompositions and biplots in three-way correspondence analysis. *Psychometrika*, 61(2), 355-373.
<https://doi.org/10.1007/BF02294344>
- Carroll, J. D. (1968). A generalization of canonical correlation analysis to three or more sets of variables. En *Proceedings of 76th annual convention of the American Psychological Associations*. (pp. 227-228).
- Carroll, J. D., & Arabie, P. (1980). Multidimensional Scaling. *Annual Review of Psychology*, 31(1), 607-649. <https://doi.org/10.1146/annurev.ps.31.020180.003135>
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3), 283-319. <https://doi.org/10.1007/BF02310791>
- Carroll, J. D., & Chang, J. J. (1972). IDIOSCAL (Individual Difference In Orientation SCALing): A generalization of INDSCAL allowing IDIOSyncratic reference systems as well as analytic approximation to INDSCAL. En *Spring Meeting of the Psychometric Society*. Princeton, New Jersey.
- Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1980). Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45(1), 3-24. <https://doi.org/10.1007/BF02293596>
- Castillo-Escrivà, A., Valls, L., Rochera, C., Camacho, A., & Mesquita-Joanes, F. (2017). Metacommunity dynamics of Ostracoda in temporary lakes: Overall strong niche effects except at the onset of the flooding period. *Limnologica*, 62, 104-110.
<https://doi.org/10.1016/j.limno.2016.11.005>
- Cazes, P., Chouakria, A., Diday, E., & Schektman, Y. (1997). Extension de l'analyse en composantes principales à des données de type intervalle. *Revue de statistique appliquée*, 45(3), 5-24.
- Ceron, K., Santana, D. J., Lucas, E. M., Zocche, J. J., & Provete, D. B. (2020). Climatic variables influence the temporal dynamics of an anuran metacommunity in a nonstationary way. *Ecology and Evolution*, 10(11), 4630-4639.
<https://doi.org/10.1002/ece3.6217>

- Ceulemans, E, Mechelen, I. Van, & Leenen, I. (2003). Tucker3 hierarchical classes analysis. *Psychometrika*, 68(3), 413-433. Recuperado de <http://link.springer.com/article/10.1007/BF02294735>
- Ceulemans, Eva, & Van Mechelen, I. (2004). Tucker2 hierarchical classes analysis. *Psychometrika*, 69(3), 375-399. <https://doi.org/10.1007/BF02295642>
- Ceulemans, Eva, & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: Interrelations and model selection. *Psychometrika*, 70(3), 461-480. <https://doi.org/10.1007/s11336-003-1067-3>
- Chaya, C., Perez-Hugalde, C., Judez, L., Wee, C. S., & Guinard, J.-X. (2003). Use of the STATIS method to analyze time-intensity profiling data. *Food Quality and Preference*, 15(1), 3-12. [https://doi.org/10.1016/S0950-3293\(02\)00219-7](https://doi.org/10.1016/S0950-3293(02)00219-7)
- Chessel, D., & Hanafi, M. (1996). Analyses de la co-inertie de K nuages de points. *Revue Statistique Appliquée*, 44(2), 36-60.
- Chessel, Daniel, Dufour, A.-B., & Thioulouse, J. (2004). The ade4 Package – I: One-Table Methods. *R News*, 4(1), 5-10. Recuperado de <https://cran.r-project.org/doc/Rnews/>.
- Chew, P., Bader, B. W., Kolda, T., & Abdelali, A. (2007). *Cross-language information retrieval using PARAFAC2*. Albuquerque, NM, and Livermore, CA. <https://doi.org/10.2172/908061>
- Chiarello, E., & Chassery, J. M. (1997). Typology of Spatial Structures of Images having the Same Color Set. *Journal of Mathematical Imaging and Vision*, 7, 359-374. <https://doi.org/10.1023/A:1008211329008>
- Chin Fatt, C. R., Beaton, D., & Abdi, H. (2015). Package ‘ MExPosition ’. Multi-table ExPosition.
- Churchill, N. W., Oder, A., Abdi, H., Tam, F., Lee, W., Thomas, C., ... Strother, S. C. (2012). Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. *Human Brain Mapping*, 33(3), 609-627. <https://doi.org/10.1002/hbm.21238>
- CISIA. (1989). Centre international de statistique et d'informatique appliquées. Recuperado 11 de septiembre de 2019, de https://data.bnf.fr/fr/12070008/centre_international_de_statistique_et_d_informatique_appliquees/
- Connolly, A. C., Sha, L., Guntupalli, J. S., Oosterhof, N., Halchenko, Y. O., Nastase, S. A., ... Haxby, J. V. (2016). How the Human Brain Represents Perceived Dangerousness or “Predacity” of Animals. *The Journal of Neuroscience*, 36(19), 5373-5384. <https://doi.org/10.1523/JNEUROSCI.3395-15.2016>

- Coquet, R., Troxler, L., & Wipff, G. (1996). The STATIS method: Characterization of conformational states of flexible molecules from molecular dynamics simulations in solution. *Journal of Molecular Graphics*, *14*(4), 206-212. [https://doi.org/10.1016/S0263-7855\(96\)00064-1](https://doi.org/10.1016/S0263-7855(96)00064-1)
- Corduas, M., & Ragozini, G. (2018). Comparing Multistep Ahead Forecasting Functions for Time Series Clustering. En *10th Scientific Meeting of the Classification-and-Data-Analysis-Group of the Italian-Statistical-Society (CLADAG)* (pp. 191-199). https://doi.org/10.1007/978-3-319-55708-3_21
- Corrales, D., & Rodríguez, O. (2014). Interstatis: el método statis para datos de tipo intervalo. *Revista de Matemática: Teoría y Aplicaciones*, *21*(1), 73. <https://doi.org/10.15517/rmta.v21i1.14139>
- Cortés Saud, A. D. C. (2005). *Contribuciones al Análisis de Tablas de Tres Vías Restringido*. Universidad de Salamanca.
- d'Aspremont, A., El Ghaoui, L., Jordan, M. I., & Lanckriet, G. R. G. (2007). A Direct Formulation for Sparse PCA Using Semidefinite Programming. *SIAM Review*, *49*(3), 434-448. <https://doi.org/10.1137/050645506>
- da Silva, J. L., & Ramos, L. P. (2014a). On the rate of convergence of uniform approximations for sequences of distribution functions. *Journal of the Korean Statistical Society*, *43*(1), 47-65. <https://doi.org/10.1016/j.jkss.2013.06.001>
- da Silva, J. L., & Ramos, L. P. (2014b). Uniform Approximations for Distributions of Continuous Random Variables with Application in Dual STATIS Method. *REVSTAT – Statistical Journal*, *12*(2), 101-118.
- Darwiche-Criado, N., Jiménez, J.-J., Comín, F. A., Sorando, R., & Sánchez-Pérez, J. M. (2015). Identifying spatial and seasonal patterns of river water quality in a semiarid irrigated agricultural Mediterranean basin. *Environmental Science and Pollution Research*, *22*(23), 18626-18636. <https://doi.org/10.1007/s11356-015-5484-5>
- Daszykowski, M., & Walczak, B. (2011). Methods for the exploratory analysis of two-dimensional chromatographic signals. *Talanta*, *83*(4), 1088-1097. <https://doi.org/10.1016/j.talanta.2010.08.032>
- de Almeida, A. L. F., Favier, G., & Mota, J. C. M. (2008). A Constrained Factor Decomposition With Application to MIMO Antenna Systems. *IEEE Transactions on Signal Processing*, *56*(6), 2429-2442. <https://doi.org/10.1109/TSP.2008.917026>
- De Boeck, P., & Rosenberg, S. (1988). Hierarchical Classes: Model and Data Analysis. *Psychometrika*, *53*(3), 361-381. Recuperado de [//ejournals.ebsco.com/direct.asp?ArticleID=483F92FE40C2008B910D](http://ejournals.ebsco.com/direct.asp?ArticleID=483F92FE40C2008B910D)

- Dill, J. (2019). Big Data. En *Advanced Information and Knowledge Processing*.
https://doi.org/10.1007/978-3-030-24367-8_2
- Doledec, S., & Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, 31(3), 277-294.
<https://doi.org/10.1111/j.1365-2427.1994.tb01741.x>
- Dray, S., & Dufour, A.-B. (2007). The **ade4** Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4). <https://doi.org/10.18637/jss.v022.i04>
- Dray, S., Dufour, A.-B., & Chessel, D. (2007). The ade4 Package – II: Two-Table and K-Table Methods. *R News*, 7(2), 47-52. Recuperado de <http://cran.r-project.org/doc/Rnews/>
- Dray, S., Dufour, A. B., & Thioulouse, J. (2004). Ade4. *Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences*.
<https://doi.org/10.18637/jss.v022.i04>>.License
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2), 407-499.
- Egido Miguélez, J. (2015). *Biplot dinámico*. Universidad de Salamanca. Recuperado de <https://gedos.usal.es/handle/10366/125245>
- Egido Miguélez, J., & Galindo-Villardón, M. P. (2015). Dynamic Biplot. Evolution of the Economic Freedom in the European Union. *British Journal of Applied Science & Technology*, 11(3), 1-13. <https://doi.org/10.9734/BJAST/2015/20289>
- Enachescu, C., & Postelnicu, T. (2003). Patterns in journal citation data revealed by exploratory multivariate analysis. *SCIENTOMETRICS*, 56(1), 43-59. Recuperado de https://apps-webofknowledge-com.ezproxy.usal.es/full_record.do?product=UA&search_mode=MarkedList&qid=26&SID=C4TeLITr6nZ7Czgdqxz&page=1&doc=7&colName=WOS
- Erichson, N., Zheng, P., Manohar, K., Brunton, S., Kuetz, J., & Aravkin, A. (2018). Sparse Principal Component Analysis via Variable Projection. *arXiv preprint arXiv:1804.00341*.
- Ernault, A., Freiré-Díaz, S., Langlois, E., & Alard, D. (2006). Are similar landscapes the result of similar histories? *Landscape Ecology*, 21(5), 631-639. <https://doi.org/10.1007/s10980-005-5321-1>
- Eros, T., Sály, P., Takács, P., Specziár, A., & Biró, P. (2012). Temporal variability in the spatial and environmental determinants of functional metacommunity organization - stream fish in

- a human-modified landscape. *Freshwater Biology*, 57(9), 1914-1928.
<https://doi.org/10.1111/j.1365-2427.2012.02842.x>
- Escofier, B., & Pagès, J. (1984). L'analyse factorielle multiple. *Cahiers du Bureau universitaire de recherche opérationnelle Série Recherche*, 42, 3-68.
- Escofier, B., & Pagès, J. (1990). *Analyses Factorielles Simples et Multiples: Objectifs, méthodes et interpretation* (Vol. 1).
- Escoufier, Y. (1973). Le Traitement des Variables Vectorielles. *Biometrics*, 29(4), 751-760.
<https://doi.org/10.2307/2529140>
- Escoufier, Y. (1976). Opérateur associé à un tableau de données. *Annales de l'inséé*, (22/23), 165. <https://doi.org/10.2307/20075217>
- Escoufier, Y. (1980). L'analyse conjointe de plusieurs matrices de données. En M. Jolivet (Ed.), *Biométrie et Temps* (pp. 59-76). Paris: Société Française de Biométrie.
- Escoufier, Y. (1985). Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données. *Statistique et analyse des données*, 10(1), 1-10.
- Eslami, A., Qannari, E. M., Bougeard, S., & Sanchez, G. (2015). Multigroup: methods for multigroup data analysis.
- Estienne, F., Matthijs, N., Massart, D. ., Ricoux, P., & Leibovici, D. (2001). Multi-way modelling of high-dimensionality electroencephalographic data. *Chemometrics and Intelligent Laboratory Systems*, 58(1), 59-72. [https://doi.org/10.1016/S0169-7439\(01\)00140-X](https://doi.org/10.1016/S0169-7439(01)00140-X)
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470977811>
- Farcomeni, A. (2009). An exact approach to sparse principal component analysis. *Computational Statistics*, 24(4), 583-604. <https://doi.org/10.1007/s00180-008-0147-3>
- Faucart, T., & Foucart, T. (1983). Une nouvelle approche de la méthode STATIS. *Revue de statistique appliquée*, 31(2), 61-75.
- Favier, G., & de Almeida, A. L. F. (2014). Overview of constrained PARAFAC models. *Eurasip Journal on Advances in Signal Processing*. <https://doi.org/10.1186/1687-6180-2014-142>
- Feki-Sahnoun, W., Hamza, A., Béjaoui, B., Mahfoudi, M., Rebai, A., & Bel Hassen, M. (2018). Multi-table approach to assess the biogeography of phytoplankton: ecological and management implications. *Hydrobiologia*, 815(1), 229-251.
<https://doi.org/10.1007/s10750-018-3566-3>
- Ferraro, S., Ardoino, I., Bassani, N., Santagostino, M., Rossi, L., Biganzoli, E., ... Panteghini, M.

- (2013). Multi-marker network in ST-elevation myocardial infarction patients undergoing primary percutaneous coronary intervention: When and what to measure. *Clinica Chimica Acta*, 417, 1-7. <https://doi.org/10.1016/J.CCA.2012.12.001>
- Figueiredo, A., Figueiredo, F., Monteiro, N. P., & Straume, O. R. (2012). Restructuring in privatised firms: A Statis approach. *Structural Change and Economic Dynamics*, 23(1), 108-116. <https://doi.org/10.1016/j.strueco.2011.10.005>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*. <https://doi.org/10.1017/CBO9781107415324.004>
- Flury, B. (1984). Common Principal Components in K Groups. *Journal of the American Statistical Association*, 79(388), 892-898. <https://doi.org/10.2307/2288721>
- Flury, B. (1988). *Common Principal Components and Related Multivariate Models* (1 edition). New York, New York, USA: John Wiley & Sons, Inc.
- Fogliatto, F. S., & Niang, N. (2008). Multivariate statistical control of unsynchronized batch processes. *Produção*, 18(2), 240-259.
- Fogliatto, F. S., & Niang, N. (2009). Multivariate statistical control of batch processes with variable duration. En *2009 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 434-438). IEEE. <https://doi.org/10.1109/IEEM.2009.5373316>
- Foucart, T. (1982). Programmation de la méthode STATIS sur micro-ordinateurs. *Statistique et analyse des données*, 7(3), 68-82.
- Fournier, M., Motelay-Massei, A., Massei, N., Aubert, M., Bakalowicz, M., & Dupont, J. P. (2009). Investigation of transport processes inside karst aquifer by means of STATIS. *Ground Water*, 47(3), 391-400. <https://doi.org/10.1111/j.1745-6584.2008.00532.x>
- Franquet, E., Dolédec, S., & Chessel, D. (1995). Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia*, 300-301(1), 425-431. <https://doi.org/10.1007/BF00024484>
- Frutos Bernal, E. (2014). *Análisis de Datos Acoplados: Método T3-PCA*. Universidad de Salamanca.
- Fu, Y., Jiang, H., Zhang, D., & Zhang, X. (2019). Comparison of Perceptual Differences Between Users and Designers in Mobile Shopping App Interface Design: Implications for Evaluation Practice. *IEEE Access*, 7, 23459-23470. <https://doi.org/10.1109/ACCESS.2019.2899671>
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal

- component analysis. *Biometrika*, 58(3), 453-467. <https://doi.org/10.1093/biomet/58.3.453>
- Gaertner, J. C. (2000). Seasonal organization patterns of demersal assemblages in the Gulf of Lions (north-western Mediterranean Sea). *Journal of the Marine Biological Association of the United Kingdom*, 80(5), 777-783. <https://doi.org/10.1017/S0025315400002745>
- Gaertner, J. C., Bertrand, J. A., Samani, D., & Souplet, A. (2005). Spatio-temporal organization patterns of demersal assemblages of the east coast of Corsica (Mediterranean Sea). *Vie et Milieu*, 55, 81-89.
- Gaertner, J. C., Bertrand, J. A., & Souplet, A. (2002). STATIS-CoA: A methodological solution to assess the spatio-temporal organization of species assemblages. Application to the demersal assemblages of the French Mediterranean Sea. *Scientia Marina*, 66(S2), 221. <https://doi.org/10.3989/scimar.2002.66s2221>
- Gaertner, J. C., Chessel, D., & Bertrand, J. (1998). Stability of spatial structures of demersal assemblages: a multitable approach. *Aquatic Living Resources*, 11(2), 75-85. [https://doi.org/10.1016/S0990-7440\(98\)80063-6](https://doi.org/10.1016/S0990-7440(98)80063-6)
- Gailhard, I., Durbec, J. P., Beliaeff, B., & Sabatier, R. (2003). Phytoplankton ecology along French coasts: Inter-site comparison. *Comptes Rendus - Biologies*, 326(9), 853-863. <https://doi.org/10.1016/j.crv.2003.09.002>
- Galindo-Villardón, M. P. (1986). An alternative for simultaneous representation: HJ-Biplot. *Questiío: Quaderns d'Estadística, Sistemes, Informàtica i Investigació Operativa*.
- Galindo-Villardón, M. P. (2007). Formulación biplot de técnicas de análisis de datos de tres modos. En *XII Reunión Científica del GAB; I Enxuentro Argentino-Chileno de Biometría* (pp. 1-8). San Martín de Los Andes, Patagonia Argentina.
- Gallego-Álvarez, I., Galindo-Villardón, M. P., & Rodríguez-Rosa, M. (2015). Evolution of sustainability indicator worldwide: A study from the economic perspective based on the X-STATICO method. *Ecological Indicators*, 58, 139-151. <https://doi.org/10.1016/J.ECOLIND.2015.05.025>
- Gallego-Álvarez, I., Lozano, M. B., & Rodríguez-Rosa, M. (2018). An analysis of the environmental information in international companies according to the new GRI standards. *Journal of Cleaner Production*, 182, 57-66. <https://doi.org/10.1016/j.jclepro.2018.01.240>
- Gauss, C. F. (1855). *Méthode des Moindres Carrés: Mémoires sur La Combinaison des Observations*. Mallet-Bachelier.
- Génard, M., Souty, M., Holmes, S., Reich, M., & Breuils, L. (1994). Correlations among quality parameters of peach fruit. *Journal of the Science of Food and Agriculture*, 66(2), 241-245. <https://doi.org/10.1002/jsfa.2740660220>

- Giordani, P., & Rocci, R. (2013a). Candecomp/Parafac with ridge regularization. *Chemometrics and Intelligent Laboratory Systems*, 129, 3-9.
<https://doi.org/10.1016/j.chemolab.2013.08.002>
- Giordani, P., & Rocci, R. (2013b). Constrained CANDECOMP/PARAFAC via the Lasso. *Psychometrika*, 78(4), 669-685.
- Giordano, G., Ragozini, G., & Vitale, M. P. (2019). Analyzing multiplex networks using factorial methods. *Social Networks*, 59, 154-170. <https://doi.org/10.1016/j.socnet.2019.07.005>
- GitHub. (2019). Recuperado de <https://github.com>
- Glacon, F. (1981). *Analyse conjointe de plusieurs matrices de données : comparaison de différentes méthodes*. Université Scientifique et Médicale de Grenoble.
- Godefroid, M., Delaville, L., Marie-Luce, S., & Quénéhervé, P. (2013). Spatial stability of a plant-feeding nematode community in relation to macro-scale soil properties. *Soil Biology and Biochemistry*, 57, 173-181. <https://doi.org/10.1016/J.SOILBIO.2012.06.019>
- Gómez-Corona, C., Valentin, D., Escalona-Buendía, H. B., & Chollet, S. (2017). The role of gender and product consumption in the mental representation of industrial and craft beers: An exploratory study with Mexican consumers. *Food Quality and Preference*, 60, 31-39.
<https://doi.org/10.1016/j.foodqual.2017.03.008>
- Gonçalves, A. M. M., Pardal, M. Â., Marques, S. C., Mendes, S., Fernández-Gómez, M. J., Galindo-Villardón, M. P., & Azeiteiro, U. M. (2012). Responses of copepoda life-history stages to climatic variability in a southern-European temperate estuary. *Zoological Studies*, 51(3), 321-335.
- González-García, N. (2019). *Análisis Sparse de Tensores Multidimensionales*. Universidad de Salamanca.
- González, V. M., Ávila, E., & Olaya, J. (2009). Análisis de los espacios de representación del statis y del afm en el estudio de movilidad biográfica en Bogotá 1993. *Revista Colombiana de Estadística*, 32(1), 1-15.
- Gourdol, L., Hissler, C., Hoffmann, L., & Pfister, L. (2013). On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry. *Applied Geochemistry*, 39, 93-107. <https://doi.org/10.1016/j.apgeochem.2013.10.002>
- Gourvéneq, S., Stanimirova, I., Saby, C.-A. A., Airiau, C. Y., & Massart, D. L. (2005). Monitoring batch processes with the STATIS approach. *Journal of Chemometrics*, 19(5-7), 288-300.
<https://doi.org/10.1002/cem.931>
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*.

<https://doi.org/10.1007/BF02291478>

- Gower, J. C., & Hand, D. J. (David J. . (1996). *Biplots*. London, UK: Chapman & Hall.
- Grossi, C., Raymond, O., Sanlaville-Boisson, C., & Jay, M. (1999). Rosa taxonomy and hierarchy of markers defined by ACT STATIS. *ZEITSCHRIFT FUR NATURFORSCHUNG SECTION C-A JOURNAL OF BIOSCIENCES*, *54c*(1-2), 25-34.
<https://doi.org/10.1515/znc-1999-1-206>
- Gudmundsson, L., Tallaksen, L. M., & Stahl, K. (2011). Spatial cross-correlation patterns of European low, mean and high flows. *Hydrological Processes*.
<https://doi.org/10.1002/hyp.7807>
- Guesdon, S., Stachowski-Haberkorn, S., Lambert, C., Beker, B., Brach-Papa, C., Auger, D., & Béchemin, C. (2016). Effect of local hydroclimate on phytoplankton groups in the Charente estuary. *Estuarine, Coastal and Shelf Science*, *181*, 325-337.
<https://doi.org/10.1016/j.ecss.2016.08.035>
- Harshman, R. A. (1970). Models and conditions for an «explanatory» multimodal factor analysis, (10), 1-84.
- Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*.
- Harshman, R. A. (1978). Models for analysis of asymmetrical relationships among N objects or stimuli. En *First Joint Meeting of the Psychometric Society and the Society of Mathematical Psychology*. Hamilton, Ontario. Recuperado de <http://ci.nii.ac.jp/naid/10030119135/en/>
- Harshman, R. A., & Lundy, M. E. (1996). Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/candecomp. *Psychometrika*, *61*(1), 133-154. <https://doi.org/10.1007/BF02296963>
- Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3), 297-310. <https://doi.org/10.1214/ss/1177013604>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of Supervised Learning (pp. 9-41). https://doi.org/10.1007/978-0-387-84858-7_2
- Hausman, R. E. (1982). Constrained multivariate analysis. In: Zanakis SH, Rustagi JS (eds). *Optimisation in Statistics*. North-Holland, Amsterdam, 137-151.
- Hernández-Fariñas, T., Soudant, D., Barillé, L., Belin, C., Lefebvre, A., & Bacher, C. (2014). Temporal changes in the phytoplankton community along the French coast of the eastern English Channel and the southern Bight of the North Sea. *ICES Journal of Marine*

- Science*, 71(4), 821-833. <https://doi.org/10.1093/icesjms/fst192>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67. <https://doi.org/10.1080/00401706.1970.10488634>
- Honoré-Chedozeau, C., Chollet, S., Lelièvre-Desmas, M., Ballester, J., & Valentin, D. (2020). From perceptual to conceptual categorization of wines: What is the effect of expertise? *Food Quality and Preference*, 80, 103806. <https://doi.org/10.1016/j.foodqual.2019.103806>
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321-377.
- Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299. <https://doi.org/10.2307/1390807>
- Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer*. <https://doi.org/10.1007/BF01898350>
- Israelsson, A. (1969). Three-way (or second order) component analysis. Nonlinear iterative partial least-squares (NIPALS) estimation procedures. *Bulletin of the International Statistical Institute*, 43, 29-51.
- Issanchou, S., Schlich, P., & Lesschaeve, I. (1997). Sensory analysis: Methodological aspects relevant to the study of cheese. *Lait*, 77(1), 5-12. <https://doi.org/10.1051/lait:199711>
- Jaffrenou, P. A. (1978). *Sur L'Analyse des familles finies des variables vectorielles: Bases algébrique et application à la description statistique*. l'Université de Sainte-Etienne, France.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jiménez, J.-J., Darwiche-Criado, N., Sorando, R., Comín, F. A., & Sánchez-Pérez, J. M. (2015). A Methodological Approach for Spatiotemporally Analyzing Water-Polluting Effluents in Agricultural Landscapes Using Partial Triadic Analysis. *Journal of Environment Quality*, 44(5), 1617. <https://doi.org/10.2134/jeq2014.09.0377>
- Jiménez, J.-J., Decaëns, T., & Rossi, J.-P. (2006). Stability of the spatio-temporal distribution and niche overlap in neotropical earthworm assemblages. *Acta Oecologica*, 30(3), 299-311. <https://doi.org/10.1016/j.actao.2006.06.008>
- Jolliffe, I. T. (1995). Rotation of principal components: choice of normalization constraints. *Journal of Applied Statistics*, 22(1), 29-35. <https://doi.org/10.1080/757584395>
- Jolliffe, I. T., Trendafilov, N. T., & Uddin, M. (2003). A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3), 531-547. <https://doi.org/10.1198/1061860032148>

- Jolliffe, I. T., & Uddin, M. (2000). The Simplified Component Technique: An Alternative to Rotated Principal Components. *Journal of Computational and Graphical Statistics*, 9(4), 689-710. <https://doi.org/10.1080/10618600.2000.10474908>
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58, 433-460.
- Kidé, S. O., Manté, C., Dubroca, L., Demarcq, H., & Mérigot, B. (2015). Spatio-Temporal Dynamics of Exploited Groundfish Species Assemblages Faced to Environmental and Fishing Forcings: Insights from the Mauritanian Exclusive Economic Zone. *PLOS ONE*, 10(10), e0141566. <https://doi.org/10.1371/journal.pone.0141566>
- Kiers, H. A. L. (1988). Comparison of "anglo-saxon" and "french" three-mode methods. *Statistique et analyse des données*, 13(3), 14-32.
- Kiers, H. A. L. (1991). Hierarchical relations among three-way methods. *Psychometrika*. <https://doi.org/10.1007/BF02294485>
- Kissita, G., Niere, L., Makany, R. A., & Mizere, D. (2013). Canonical correlation analysis between two vertical multi-block tables and applications to qualitative variables. *Advances and Applications in Statistics*, 34(1), 11. Recuperado de https://www.researchgate.net/publication/265994847_Canonical_correlation_analysis_between_two_vertical_multi-block_tables_and_applications_to_qualitative_variables
- Klie, S., Caldana, C., & Nikoloski, Z. (2012). Compromise of Multiple Time-Resolved Transcriptomics Experiments Identifies Tightly Regulated Functions. *Frontiers in Plant Science*, 3(November), 1-12. <https://doi.org/10.3389/fpls.2012.00249>
- Klie, S., Osorio, S., Tohge, T., Drincovich, M. F., Fait, A., Giovannoni, J. J., ... Nikoloski, Z. (2014). Conserved Changes in the Dynamics of Metabolic Processes during Fruit Development and Ripening across Species. *Plant physiology*, 164(1), 55-68. <https://doi.org/10.1104/pp.113.226142>
- Kolda, T. G., & Bader, B. W. (2009). Tensor Decompositions and Applications. *SIAM Review*, 51(3), 455-500. <https://doi.org/10.1137/07070111X>
- Kriegsman, M. A. (2018a). *Discriminant DiSTATIS: A Multi-Way Discriminant Analysis for Distance Matrices, Illustrations with the Sorting Task*. The University of Texas at Dallas. Recuperado de <https://utd-ir.tdl.org/handle/10735.1/6392?show=full>
- Kriegsman, M. A. (2018b). Paquete en R DiDiSTATIS: Discriminant DiSTATIS. Recuperado de <https://github.com/michaelkriegsman?tab=repositories>
- Krishnan, A., Kriegeskorte, N., & Abdi, H. (2013). Distance-Based Partial Least Squares Analysis. En Texas University (Ed.), *New Perspectives in Partial Least Squares and*

- Related Methods* (pp. 131-145). Houston, TX. https://doi.org/10.1007/978-1-4614-8283-3_8
- Kroonenberg, P. M. (1989). The analysis of multiple tables in factorial ecology. III.-three-mode principle component analyses: «Analyse triadique complète». *ACTA OECOL; OECOL GEN*, 10(3), 245-256.
- Kroonenberg, P. M. (2008). *Applied Multiway data Analysis*. Wiley series in probability and statistics. John Wiley & Sons, Inc.
- Kroonenberg, P. M. (2016). My Multiway Analysis: From Jan de Leeuw to TWPack and Back. *Journal of Statistical Software*, 73(3). <https://doi.org/10.18637/jss.v073.i03>
- Kroonenberg, P. M., & de Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1), 69-97. <https://doi.org/10.1007/BF02293599>
- Krzanowski, W. J. (1979). Between-Groups Comparison of Principal Components. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2286995>
- Krzanowski, W. J. (1982). Between-group comparison of principal components — some sampling results. *Journal of Statistical Computation and Simulation*, 15(2-3), 141-154. <https://doi.org/10.1080/00949658208810577>
- L'Hermier des Plantes, H. (1976). *Structuration des tableaux à trois indices de la statistique*. Université de Montpellier II.
- L'Hermier des Plantes, H., & Thiébaud, B. (1977). Étude de la pluviosité au moyen de la méthode STATIS. *Revue de statistique appliquée*, 25(2), 57-81.
- Ladhar, C., Tastard, E., Casse, N., Denis, F., & Ayadi, H. (2015). Strong and stable environmental structuring of the zooplankton communities in interconnected salt ponds. *Hydrobiologia*, 743(1), 1-13. <https://doi.org/10.1007/s10750-014-1998-y>
- Lahne, J. (2020). Sorting backbone analysis: A network-based method of extracting key actionable information from free-sorting task results. *Food Quality and Preference*, 82. <https://doi.org/10.1016/j.foodqual.2020.103870>
- Lahne, J., Abdi, H., & Heymann, H. (2018). Rapid sensory profiles with DISTATIS and Barycentric Text Projection: An example with amari , bitter herbal liqueurs. *Food Quality and Preference*, 66, 36-43. <https://doi.org/10.1016/j.foodqual.2018.01.003>
- Lahne, J., Collins, T. S., & Heymann, H. (2016). Replication Improves Sorting-Task Results Analyzed by DISTATIS in a Consumer Study of American Bourbon and Rye Whiskeys. *Journal of Food Science*, 81(5), S1263-S1271. <https://doi.org/10.1111/1750-3841.13301>

- Lavit, C. (1985). Application de la méthode STATIS. *Statistique et analyse des données*, 10(1), 103-116.
- Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs* (Masson). Paris, France: Masson.
Recuperado de
https://books.google.es/books/about/Analyse_conjointe_de_tableaux_quantitati.html?id=d1oAAAACAAJ&redir_esc=y
- Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18(1), 97-119. [https://doi.org/10.1016/0167-9473\(94\)90134-1](https://doi.org/10.1016/0167-9473(94)90134-1)
- Lazraq, A., Hanafi, M., Cléroux, R., Allaire, J., & Lepage, Y. (2008). Une approche inférentielle pour la validation du compromis de la méthode STATIS. *Journal de la Société Française de Statistique*, 149(1), 97-109.
- Le Dien, S., & Pagès, J. (2003). Hierarchical Multiple Factor Analysis: application to the comparison of sensory profiles. *Food Quality and Preference*, 14(5-6), 397-403. [https://doi.org/10.1016/S0950-3293\(03\)00027-2](https://doi.org/10.1016/S0950-3293(03)00027-2)
- Leenen, I., Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1999). Indclas: A three-way hierarchical classes model. *Psychometrika*, 64(1), 9-24. <https://doi.org/10.1007/BF02294316>
- Legendre, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot.
- Lekve, K. (2002). Species richness and environmental conditions of fish along the Norwegian Skagerrak coast. *ICES Journal of Marine Science*, 59(4), 757-769. <https://doi.org/10.1006/jmsc.2002.1247>
- Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 19(8), 697-703. <https://doi.org/10.1016/j.foodqual.2008.05.001>
- Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2009). Beer-Trained and Untrained Assessors Rely More on Vision than on Taste When They Categorize Beers. *Chemosensory Perception*, 2(3), 143-153. <https://doi.org/10.1007/s12078-009-9050-8>
- Lemesle, S., Mussio, I., Rusig, A.-M., Menet-Nédélec, F., & Claquin, P. (2015). Impact of seaweed beachings on dynamics of $\delta^{15}\text{N}$ isotopic signatures in marine macroalgae. *Marine Pollution Bulletin*, 97(1-2), 241-254. <https://doi.org/10.1016/j.marpolbul.2015.06.010>
- Lera M., L., Albala B., C., Santos M., J. L., Arroyo A., P., García G., C., & Bunout B., D. (2007).

- Multivariate study of the evolution of body dimensions in elderly Chilean women. *Archivos latinoamericanos de nutrición.*, 57(3), 219-223. Recuperado de https://apps-webofknowledge-com.ezproxy.usal.es/full_record.do?product=UA&search_mode=MarkedList&qid=61&SID=C4TeLITr6nZ7Czgdqxz&page=1&doc=3&colName=WOS
- Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1991.10475035>
- Licandro, P., & Ibanez, F. (2000). Changes of zooplankton communities in the Gulf of Tigullio (Ligurian Sea, Western Mediterranean) from 1985 to 1995. Influence of hydroclimatic factors. *Journal of Plankton Research*, 22(12), 2225-2253. <https://doi.org/10.1093/plankt/22.12.2225>
- Liu, X., Wang, Y., Li, Y., & Wu, J. (2019). Quantifying the Spatio-Temporal Process of Township Urbanization: A Large-Scale Data-Driven Approach. *ISPRS International Journal of Geo-Information*, 8(9), 389. <https://doi.org/10.3390/ijgi8090389>
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference*, 72, 31-39. <https://doi.org/10.1016/j.foodqual.2018.09.006>
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2020). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, 79, 103520. <https://doi.org/10.1016/j.foodqual.2018.05.013>
- Llobell, F., Vigneau, E., Cariou, V., & Qannari, E. M. (2019). ClustBlock: Clustering of Datasets. Recuperado de <https://cran.r-project.org/package=ClustBlock>
- Llobell, F., Vigneau, E., & Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104. <https://doi.org/10.1016/j.foodqual.2019.02.017>
- Lloyd, S. P. (1982). Least squares quantization in {PCM}. Special issue on quantization. *IEEE Transactions on Information Theory*, 28(2), 129-137.
- Lobry, J., Lepage, M., & Rochard, E. (2006). From seasonal patterns to a reference situation in an estuarine environment: Example of the small fish and shrimp fauna of the Gironde estuary (SW France). *Estuarine, Coastal and Shelf Science*, 70(1-2), 239-250. <https://doi.org/10.1016/j.ecss.2006.06.014>
- Lopes, I., Silva, J. A., Castela, G., & Rebelo, E. (2017). Knowledge Transfer Through Journals. En *Bridging Tourism Theory and Practice* (pp. 271-288). <https://doi.org/10.1108/S2042-144320170000008018>

- Mabrouki, Y., Taybi, A. F., El Alami, M., & Berrahou, A. (2019). Biotypology of stream macroinvertebrates from North African and semi arid catchment: Oued Za (Morocco). *Knowledge & Management of Aquatic Ecosystems*, (420), 17.
<https://doi.org/10.1051/kmae/2019009>
- Marcondes Filho, D., de Oliveira, L. P. L., & Fogliatto, F. S. (2017). Erratum to: Multivariate quality control of batch processes using STATIS. *The International Journal of Advanced Manufacturing Technology*, 88(5-8), 2355-2355. <https://doi.org/10.1007/s00170-016-8960-2>
- Marcondes Filho, D., Fogliatto, F. S., & de Oliveira, L. P. L. (2011). Multivariate control charts for monitoring non-linear batch processes. *Producao*, 21(1), 132-148.
<https://doi.org/10.1590/S0103-65132011005000006>
- Marcondes Filho, D., Fogliatto, F. S., & de Oliveira, L. P. L. (2014). Monitoring Nonlinear Batch Process Using Statis-Based Method. *Applied Mechanics and Materials*, 518, 350-355.
<https://doi.org/10.4028/www.scientific.net/AMM.518.350>
- Marques, S. C., Pardal, M. Â., Mendes, S., & Azeiteiro, U. M. (2011). Using multitable techniques for assessing the temporal variability of species-environment relationship in a copepod community from a temperate estuarine ecosystem. *Journal of Experimental Marine Biology and Ecology*. <https://doi.org/10.1016/j.jembe.2011.05.015>
- Martín-Rodríguez, J. (1996). *Contribuciones a la Integración de Subespacios desde una Perspectiva Biplot*. Universidad de Salamanca.
- Martín-Rodríguez, J., Galindo-Villardón, M. P., & Vicente-Villardón, J. L. (2002). Comparison and integration of subspaces from a biplot perspective. *Journal of Statistical Planning and Inference*, 102(2), 411-423. [https://doi.org/10.1016/S0378-3758\(01\)00101-X](https://doi.org/10.1016/S0378-3758(01)00101-X)
- Masmoudi, S., Tastard, E., Guermazi, W., Caruso, A., Morant-Manceau, A., & Ayadi, H. (2015). Salinity gradient and nutrients as major structuring factors of the phytoplankton communities in salt marshes. *Aquatic Ecology*, 49(1), 1-19.
<https://doi.org/10.1007/s10452-014-9500-5>
- Mazzocchi, M. G., Dubroca, L., García-Comas, C., Capua, I. Di, & Ribera d'Alcalà, M. (2012). Stability and resilience in coastal copepod assemblages: The case of the Mediterranean long-term ecological research at Station MC (LTER-MC). *Progress in Oceanography*, 97-100, 135-151. <https://doi.org/10.1016/j.pocean.2011.11.003>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*.
- McCabe, G. P. (1984). Principal Variables. *Technometrics*, 26(2), 137-144.
<https://doi.org/10.1080/00401706.1984.10487939>

- Mendes, S. (2011). *Métodos Multivariantes para evaluar patrones de estabilidad y cambio desde una perspectiva BIPLLOT*. Universidad de Salamanca.
- Mendes, S., Fernández-Gómez, M. J., Marques, S. C., Pardal, M. Â., Azeiteiro, U. M., & Galindo-Villardón, M. P. (2017). CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables. *Journal of Applied Statistics*, *44*(15), 2729-2755. <https://doi.org/10.1080/02664763.2016.1261815>
- Mendes, S., Fernández-Gómez, M. J., Pereira, M. J., Azeiteiro, U. M., & Galindo-Villardón, M. P. (2012). An empirical comparison of Canonical Correspondence Analysis and STATICO in the identification of spatio-temporal ecological relationships. *Journal of Applied Statistics*, *39*(5), 979-994. <https://doi.org/10.1080/02664763.2011.634393>
- Mendes, S., Fernández-Gómez, M. J., Resende, P., Jorge Pereira, M., Galindo-Villardón, M. P., & Azeiteiro, U. M. (2009). Spatio-temporal structure of diatom assemblages in a temperate estuary. A STATICO analysis. *Estuarine, Coastal and Shelf Science*, *84*(4), 637-644. <https://doi.org/10.1016/j.ecss.2009.08.003>
- Mendes, S., Marques, S. C., Azeiteiro, U. M., Fernández-Gómez, M. J., Galindo-Villardón, M. P., Maranhão, P., ... Leandro, S. M. (2011). Zooplankton distribution in a Marine protected area: The berlingas natural reserve (Western Coast of Portugal). *Fresenius Environmental Bulletin*.
- Mérigot, B., Gaertner, J. C., Brind'amour, A., Carbonara, P., Esteban, A., Garcia-Ruiz, C., ... Vrgoc, N. (2019). Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern mediterranean sea. *Scientia Marina*, *83*(S1), 153-163. <https://doi.org/10.3989/scimar.04954.30A>
- Mérillet, L., Kopp, D., Robert, M., Mouchet, M., & Pavoine, S. (2020). Environment outweighs the effects of fishing in regulating demersal community structure in an exploited marine ecosystem. *Global Change Biology*, *26*(4), 2106-2119. <https://doi.org/10.1111/gcb.14969>
- Meyners, M. (2003). Methods to analyse sensory profiling data—a comparison. *Food Quality and Preference*, *14*(5-6), 507-514. [https://doi.org/10.1016/S0950-3293\(03\)00004-1](https://doi.org/10.1016/S0950-3293(03)00004-1)
- Meyners, M., Kunert, J., & Qannari, E. M. (2000). Comparing generalized procrustes analysis and statis. *Food Quality and Preference*, *11*(1-2), 77-83. [https://doi.org/10.1016/S0950-3293\(99\)00038-5](https://doi.org/10.1016/S0950-3293(99)00038-5)
- Mielby, L. H., Hopfer, H., Jensen, S., Thybo, A. K., & Heymann, H. (2014). Comparison of descriptive analysis, projective mapping and sorting performed on pictures of fruit and vegetable mixes. *Food Quality and Preference*, *35*, 86-94. <https://doi.org/10.1016/j.foodqual.2014.02.006>

- Moghaddam, B., Weiss, Y., & Avidan, S. (2006). Spectral bounds for sparse PCA: Exact and greedy algorithms. *Advances in neural information processing systems*.
- Moore, R. E. (1979). *Methods and Applications of Interval Analysis. Methods and Applications of Interval Analysis*. <https://doi.org/10.1137/1.9781611970906>
- Muiño, R., Carrera, P., Petitgas, P., Beare, D. J., Georgakarakos, S., Haralambous, J., ... Reid, D. G. (2003). Consistency in the correlation of school parameters across years and stocks. *ICES Journal of Marine Science*, *60*(1), 164-175. <https://doi.org/10.1006/jmsc.2002.1322>
- Napoléon, C., Raimbault, V., Fiant, L., Riou, P., Lefebvre, S., Lampert, L., & Claquin, P. (2012). Spatiotemporal dynamics of physicochemical and photosynthetic parameters in the central English Channel. *Journal of Sea Research*, *69*, 43-52. <https://doi.org/10.1016/j.seares.2012.01.005>
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, *135*(3), 370. <https://doi.org/10.2307/2344614>
- Nguyen, L. H., & Holmes, S. (2017). Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations. *BMC Bioinformatics*, *18*(S10), 394. <https://doi.org/10.1186/s12859-017-1790-x>
- Nguyen, L. H., & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology*, *15*(6), e1006907. <https://doi.org/10.1371/journal.pcbi.1006907>
- Niang, N., Fogliatto, F. S., & Saporta, G. (2013). Non parametric on-line control of batch processes based on STATIS and clustering. *Journal de la Société Française de Statistique*, *154*(3), 124-142.
- Niang, N., Fogliatto, F., & Saporta, G. (2009). Contrôle multivarié de procédés par lots à l'aide de Statist. En *41èmes Journée de Statistique*.
- Olivares, B. O., Zingaretti, M. L., Demey Zambrano, J. A., & Demey, J. R. (2017). The STATIS-ACT method applied to rainfall in the Venezuelan Eastern Region. *Cuadernos de Investigación UNED*, *9*(1), 97-106.
- Oliveira, M. M., & Mexia, J. (2007a). ANOVA-like analysis of matched series of studies with a common structure. *Journal of Statistical Planning and Inference*, *137*, 1862-1870. <https://doi.org/10.1016/j.jspi.2006.05.007>
- Oliveira, M. M., & Mexia, J. T. (2004). AIDS in Portugal: endemic versus epidemic forecasting scenarios for mortality. *International Journal of Forecasting*, *20*(1), 131-135. [https://doi.org/10.1016/S0169-2070\(03\)00011-6](https://doi.org/10.1016/S0169-2070(03)00011-6)

- Oliveira, M. M., & Mexia, J. T. (2007b). Modelling series of studies with a common structure. *Computational Statistics and Data Analysis*, 51, 5876-5885.
<https://doi.org/10.1016/j.csda.2006.11.003>
- Oszwald, J., Gond, V., Dolédec, S., & Lavelle, P. (2011). Identification of land-use indicators to assess changes in landscape mosaics. *Bois et Forêts des Tropiques*, 307, 7-21.
- Pagès, J. (1996). Éléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS. *Revue Statistique Appliquée*, 44(4), 81-95. Recuperado de http://www.numdam.org/item?id=RSA_1996__44_4_81_0
- Pagès, J. (2014). Multiple Factor Analysis: General Presentation and Comparison with STATIS. En J. Blasius & M. Greenacre (Eds.), *Visualization and Verbalization of Data* (pp. 223-237). CRC PRESS-TAYLOR & FRANCIS GROUP,. Recuperado de <https://www.taylorfrancis.com/books/e/9780429167980/chapters/10.1201/b16741-20>
- Pavoine, S., & Bailly, X. (2007). New analysis for consistency among markers in the study of genetic diversity: development and application to the description of bacterial diversity. *BMC Evolutionary Biology*, 7(1), 156. <https://doi.org/10.1186/1471-2148-7-156>
- Pavoine, S., Blondel, J., Baguette, M., & Chessel, D. (2007). A new technique for ordering asymmetrical three-dimensional data sets in ecology. *Ecology*, 88(2), 512-523.
<https://doi.org/10.1890/05-1806>
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572. <https://doi.org/10.1080/14786440109462720>
- Pellegrin, F., Nandris, D., Westrelin, S., & Kohler, F. (1995). Pathological status of arabica in New-Caledonia, correlations between epidemics and environment. En *16TH International Scientific Colloquium on Coffee* (pp. 690-698). Kyoto, Japan.
- Perez-Hugalde, C., Judez, L., Litago, J., Yuste, J., & Fuentes-Pila, J. (2004). Statistical procedure for clonal preselection of *Vitis vinifera* L. cv. Tempranillo in the Duero Valley, Spain. *American journal of enology and viticulture.*, 55(4), 335-345. Recuperado de https://apps- webofknowledge-com.ezproxy.usal.es/full_record.do?product=UA&search_mode=MarkedList&qid=17&SID=C4TeLITr6nZ7Czgdqxz&page=1&doc=5&colName=WOS
- Pérez Aparicio, J., Ángeles Toledano Medina, M., & Lafuente Rosales, V. (2007). Descriptive sensory analysis in different classes of orange juice by a robust free-choice profile method. *Analytica Chimica Acta*, 595(1-2), 238-247.
<https://doi.org/10.1016/j.aca.2007.02.054>
- Pernin, M.-O. (1986). *Contribution à la méthodologie d'analyse de données longitudinales :*

- exemple de la croissance chez l'être humain (Auxologie)*. <http://www.theses.fr>. Université Lyon I. Recuperado de <https://www.theses.fr/1986LYO10023>
- Pinzón, L. M. (2011). *Biplot Consenso para Análisis de Tablas Múltiples*. Universidad de Salamanca. Recuperado de <http://hdl.handle.net/10366/115645>
- Pinzón, L. M., & Vicente-Villardón, J. L. (2012). *Biplot Consenso para Análisis de Tablas Múltiples*. Salamanca: Universidad de Salamanca.
- Pizarro, C., Esteban-Díez, I., Rodríguez-Tecedor, S., & González-Sáiz, J. M. (2013). A sensory approach for the monitoring of accelerated red wine aging processes using multi-block methods. *Food Quality and Preference*, 28(2), 519-530. <https://doi.org/10.1016/j.foodqual.2013.01.002>
- Qannari, E. M., Courcoux, P., Lejeune, M., & Maystre, O. (1997). Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle. *Revue de statistique appliquée*, 45(1), 61-74.
- Qannari, E. M., Wakeling, I., & MacFie, H. J. H. (1995). A hierarchy of models for analysing sensory data. *Food Quality and Preference*, 6(4), 309-314. [https://doi.org/10.1016/0950-3293\(95\)00033-X](https://doi.org/10.1016/0950-3293(95)00033-X)
- Qi, X., Luo, R., & Zhao, H. (2013). Sparse principal component analysis by choice of norm. *Journal of Multivariate Analysis*, 114:127-150. <https://doi.org/10.1016/j.jmva.2012.07.004>
- R-Forge. (2019). Recuperado de <https://r-forge.r-project.org/>
- R Development Core Team. (2019). R Software. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing . Vienna, Austria. Recuperado de <http://www.r-project.org/>
- Raman, R., Kriegsman, M. A., Abdi, H., Tillmann, B., & Dowling, W. J. (2020). Bach, Mozart, and Beethoven: Sorting piano excerpts based on perceived similarity using DiSTATIS. *New Ideas in Psychology*, 57, 100757. <https://doi.org/10.1016/j.newideapsych.2019.100757>
- Ramos-Barberán, M., Hinojosa-Ramos, M. V., Ascencio-Moreno, J., Vera, F., Ruiz-Barzola, O., & Galindo-Villardón, M. P. (2018). Batch process control and monitoring: a Dual STATIS and Parallel Coordinates (DS-PC) approach. *Production and Manufacturing Research*, 6(1), 470-493. <https://doi.org/10.1080/21693277.2018.1547228>
- Raymond, O., Fiasson, J. L., & Jay, M. (2000). Synthetic taxonomy of Rosa races using ACT-STATIS. *ZEITSCHRIFT FÜR NATURFORSCHUNG C-A JOURNAL OF BIOSCIENCES*, 55(5-6), 399-409.

- Regnault, A., Marfatia, S., Louie, M., Mear, I., Meunier, J., & Viala-Danten, M. (2009). Satisfactory cross-cultural validity of the ACTG symptom distress module in HIV-1-infected antiretroviral-naive patients. *Clinical Trials: Journal of the Society for Clinical Trials*, 6(6), 574-584. <https://doi.org/10.1177/1740774509352515>
- Ren, X., Yang, K., Che, Y., Wang, M., Zhou, L., & Chen, L. (2016). Spatial and temporal assessment of the initial pattern of phytoplankton population in a newly built coastal reservoir. *Frontiers of Earth Science*, 10(3), 546-559. <https://doi.org/10.1007/s11707-015-0543-2>
- Rivadeneira, F. J., Figueiredo, A. M., Figueiredo, F. O., Carvajal, S. M., & Rivadeneira, R. A. (2016). Analysis of Well-Being in OECD Countries Through STATIS Methodology. *HOLOS*, 7, 335-351. <https://doi.org/10.15628/holos.2016.5003>
- Robert, P., & Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient. *Applied Statistics*, 25(3), 257. <https://doi.org/10.2307/2347233>
- Rodrigues, L. C., Pivato, B. M., Vieira, L. C. G., Bovo-Scomparin, V. M., Bortolini, J. C., Pineda, A., & Train, S. (2018). Use of phytoplankton functional groups as a model of spatial and temporal patterns in reservoirs: a case study in a reservoir of central Brazil. *Hydrobiologia*, 805(1), 147-161. <https://doi.org/10.1007/s10750-017-3289-x>
- Rodríguez-Martínez, C. C., García-Sánchez, I.-M., Vicente-Galindo, P., & Galindo-Villardón, M. P. (2019). Exploring Relationships between Environmental Performance, E-Government and Corruption: A Multivariate Perspective. *Sustainability*, 11(22), 6497. <https://doi.org/10.3390/su11226497>
- Rodríguez-Rosa, M., Gallego-Álvarez, I., Vicente-Galindo, P., & Galindo-Villardón, M. P. (2017). Are Social, Economic and Environmental Well-Being Equally Important in all Countries Around the World? A Study by Income Levels. *Social Indicators Research*, 131(2), 543-565. <https://doi.org/10.1007/s11205-016-1257-x>
- Rodríguez-Rosa, M., Gallego-Álvarez, I., & Galindo-Villardón, M. P. (2019). Spatio-temporal analysis of economic, social, and environmental issues in the framework of sustainable development in worldwide countries. *Sustainable Development*, 27(3), 429-447. <https://doi.org/10.1002/sd.1916>
- Rodríguez Rosa, M. (2016). *Contribuciones al Análisis de la Sostenibilidad Internacional , desde una Perspectiva Algebraica Multivariante Comparada*. Universidad de Salamanca.
- Rolland, A., Bertrand, F., Maumy, M., & Jacquet, S. (2009). Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis. *Water Research*, 43(13), 3155-3168. <https://doi.org/10.1016/j.watres.2009.03.049>

- Rossi, J.-P. (2003). The spatiotemporal pattern of a tropical earthworm species assemblage and its relationship with soil structure: The 7th international symposium on earthworm ecology · Cardiff · Wales · 2002. *Pedobiologia*, 47(5-6), 497-503.
<https://doi.org/10.1078/0031-4056-00219>
- Rossi, J.-P., Nardin, M., Godefroid, M., Ruiz-Diaz, M., Sergent, A.-S., Martinez-Meier, A., ... Rozenberg, P. (2014). Dissecting the Space-Time Structure of Tree-Ring Datasets Using the Partial Triadic Analysis. *PLoS ONE*, 9(9), e108332.
<https://doi.org/10.1371/journal.pone.0108332>
- Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The Bagplot: A Bivariate Boxplot. *The American Statistician*, 53(4), 382-387. <https://doi.org/10.1080/00031305.1999.10474494>
- Rudnitskaya, A., Polshin, E., Kirsanov, D., Lammertyn, J., Nicolai, B., Saison, D., ... Legin, A. (2009). Instrumental measurement of beer taste attributes using an electronic tongue. *Analytica Chimica Acta*, 646(1-2), 111-118. <https://doi.org/10.1016/j.aca.2009.05.008>
- Rundle, M. M., Coch, D., Connolly, A. C., & Granger, R. H. (2018). Dissociating frequency and animacy effects in visual word processing: An fMRI study. *Brain and Language*, 183, 54-63. <https://doi.org/10.1016/j.bandl.2018.05.005>
- Sabatier, R., Myrtille, V., & Christelle, R. (2013). Une nouvelle proposition , l'Analyse Discriminante Multitableaux : STATIS-LDA. *Journal de la Société Française de Statistique*, 154(3), 31-43.
- Sabatier, R., & Vivien, M. (2008). A new linear method for analyzing four-way multiblocks tables: STATIS-4. *Journal of Chemometrics*, 22(6), 399-407.
<https://doi.org/10.1002/cem.1150>
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. En *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*.
<https://doi.org/10.1109/CTS.2013.6567202>
- Sáinz, J. A., Garrido, I., Hernández, M., Montaña, A., Llerena, J. L., & Espinosa, F. (2019). Influence of cultivar, irrigation, ripening stage, and annual variability on the oxidant/antioxidant systems of olives as determined by MDS-PTA. *PLOS ONE*, 14(4), e0215540. <https://doi.org/10.1371/journal.pone.0215540>
- Santos, António Duarte, da Silva, N. T., & Castela, G. (2019). A COSTATIS approach to business sustainability in turbulent environments from 2008 to 2014. *International Journal for Quality Research*, 13(4), 887-900. <https://doi.org/10.24874/IJQR13.04-10>
- Santos, Antonio Duarte, Ribeiro, S., Castela, G., & Tavares da Silva, N. (2017). The Dynamics between Economic Growth and Living Standards in EU Countries A STATICO Approach for the Period 2006-2014. *Estudios de economía aplicada*, ISSN-e 1133-3197, Vol. 35, Nº

- 3, 2017 (*Ejemplar dedicado a: Crisis, Economía y Finanzas*), págs. 629-652, 35(3), 629-652. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=6131523>
- Santosa, M., Abdi, H., & Guinard, J.-X. (2010). A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality and Preference*, 21(7), 881-892. <https://doi.org/10.1016/j.foodqual.2010.05.011>
- Sauzay, L., Hanafi, M., Qannari, E. M., & Schlich, P. (2006). Analyse de K + 1 tableaux a l'aide de la méthode STATIS: application en evaluation sensorielle. En *9ième Journées Européennes Agro-industrie et Méthodes Statistiques* (pp. 1-23). Faculté de Pharmacie, Montpellier, France.
- Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. En *Data Handling in Science and Technology*. [https://doi.org/10.1016/S0922-3487\(96\)80033-5](https://doi.org/10.1016/S0922-3487(96)80033-5)
- Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Massachusetts: MIT Press.
- Schwarz, A. L., Jurica, M., Matson, C., Stiller, R., Webb-Culver, T., & Abdi, H. (2020). Storybook selection criteria used by teachers of d/Deaf and hard-of-hearing prereaders communicating in English. *Deafness & Education International*, 22(3), 176-211. <https://doi.org/10.1080/14643154.2019.1572284>
- Schwarz, A. L., van Kleeck, A., Beaton, D., Horne, E., MacKenzie, H., & Abdi, H. (2015). A Read-Aloud Storybook Selection System for Prereaders at the Preschool Language Level: A Pilot Study. *Journal of Speech, Language, and Hearing Research*, 58(4), 1273-1291. https://doi.org/10.1044/2015_JSLHR-L-15-0056
- Serghini, M., Boutayeb, A., Boumâaz, A., Srairi, A., Mesfioui, A., Zoubi, A., & Dridi, A. (2008). Stability of the spatial structures of demersal assemblage in the Moroccan southern Atlantic zone. *Applied Ecology and Environmental Research*.
- Shen, H., & Huang, J. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99(6), 1015-1034.
- Shieh, S.-H., & Chi, Y.-S. (2010). Factors influencing macroinvertebrate assemblages in artificial subtropical ponds of Taiwan. *Hydrobiologia*, 649(1), 317-330. <https://doi.org/10.1007/s10750-010-0275-y>
- Shinkareva, S. V., Malave, V. L., Just, M. A., & Mitchell, T. M. (2012). Exploring commonalities across participants in the neural representation of objects. *Human Brain Mapping*, 33(6), 1375-1383. <https://doi.org/10.1002/hbm.21296>
- Shinkareva, S. V., Wang, J., Kim, J., Facciani, M. J., Baucom, L. B., & Wedell, D. H. (2014).

- Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Human Brain Mapping*, 35(7), 3558-3568. <https://doi.org/10.1002/hbm.22421>
- Sicard, E., Sabatier, R., Niel, H., & Cadier, E. (2002). A new approach in space-time analysis of multivariate hydrological data: Application to Brazil's Nordeste region rainfall. *Water Resources Research*, 38(12), 55(1)-55(10). <https://doi.org/10.1029/2002wr001413>
- Simier, M., Blanc, L., Aliaume, C., Diouf, P. ., & Albaret, J.-J. (2004). Spatial and temporal structure of fish assemblages in an "inverse estuary", the Sine Saloum system (Senegal). *Estuarine, Coastal and Shelf Science*, 59(1), 69-86. <https://doi.org/10.1016/j.ecss.2003.08.002>
- Simier, M., Blanc, L., Pellegrin, F., & Nandris, D. (1999). Approche simultanée de K couples de tableaux : application à l'étude des relations pathologie végétale-environnement. *Revue Statistique Appliquée*, 47(1), 31-46.
- Simier, M., Laurent, C., Ecoutin, J.-M., & Albaret, J.-J. (2006). The Gambia River estuary: A reference point for estuarine fish assemblages studies in West Africa. *Estuarine, Coastal and Shelf Science*, 69(3-4), 615-628. <https://doi.org/10.1016/j.ecss.2006.05.028>
- Slimani, N., Guilbert, E., Ayni, F. El, Jrad, A., Boumaiza, M., & Thioulouse, J. (2017). The use of STATICO and COSTATIS, two exploratory three-ways analysis methods: an application to the ecology of aquatic heteroptera in the Medjerda watershed (Tunisia). *Environmental and Ecological Statistics*, 24(2), 269-295. <https://doi.org/10.1007/s10651-017-0370-6>
- Sowmya, R., & Suneetha, K. R. (2017). Data Mining with Big Data. En *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*. <https://doi.org/10.1109/ISCO.2017.7855990>
- Stanimirova, I., Boucon, C., & Walczak, B. (2011). Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction. *Talanta*, 83(4), 1239-1246. <https://doi.org/10.1016/j.talanta.2010.09.018>
- Stanimirova, I., Połowniak, M., Skorek, R., Kita, A., John, E., Buhl, F., & Walczak, B. (2007). Chemometric analysis of the water purification process data. *Talanta*, 74(1), 153-162. <https://doi.org/10.1016/j.talanta.2007.05.044>
- Stanimirova, I., Walczak, B., Massart, D. L., Simeonov, V., Saby, C. A., & Di Crescenzo, E. (2004). STATIS, a three-way method for data analysis. Application to environmental data. *Chemometrics and Intelligent Laboratory Systems*, 73(2), 219-233. <https://doi.org/10.1016/j.chemolab.2004.03.005>
- Stefanescu, M. V., Serban, F., Busu, M., & Ferrara, M. (2010). Portfolio optimization using classification and functional data analysis techniques. *Economic Computation and*

- Stolfi, J., & De Figueiredo, L. H. (2003). An Introduction to Affine Arithmetic. *TEMA - Tendências em Matemática Aplicada e Computacional*.
<https://doi.org/10.5540/tema.2003.04.03.0297>
- Thabet, R., Leignel, V., Ayadi, H., & Tastard, E. (2018). Interannual and seasonal effects of environmental factors on the zooplankton distribution in the solar saltern of Sfax (south-western Mediterranean sea). *Continental Shelf Research*, 165, 1-11.
<https://doi.org/10.1016/j.csr.2018.06.002>
- Thioulouse, J. (2011). Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *The Annals of Applied Statistics*, 5(4), 2300-2325.
<https://doi.org/10.1214/10-AOAS372>
- Thioulouse, J., & Chessel, D. (1987). Multi-table analysis of ecological factors. I. From description of ecosystem status to that of ecosystem function using triadic analysis. *ACTA OECOL OECOL GEN*, 8(4), 463-480.
- Thioulouse, J., Chessel, D., Dolédec, S., & Olivier, J. M. (1997). ADE-4: A multivariate analysis and graphical display software. *Statistics and Computing*, 7, 75-83.
- Thioulouse, J., Simier, M., & Chessel, D. (2004). Simultaneous Analysis of a sequence of paired ecological tables. *Ecology*, 85(1), 272-283. <https://doi.org/10.1890/02-0605>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Oxford, England: Wiley.
- Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, 29(3-4), 431-454.
- Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change*, 15, 122-137.
- Tucker, L. R. (1964). The Extension of Factor Analysis to Three Dimensional Matrices. *Contributions to Mathematical Psychology*, 110-182.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279-311. <https://doi.org/10.1007/BF02289464>
- Tucker, L. R. (1972). Relation between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.

- Tucker, L. R. (1975). Three-mode factor analysis applied to multidimensional scaling. En *US-Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques*. La Jolla, CA.
- Vallejo-Arboleda, A. (2004). *Análisis de datos Multivía con estructura de grupos*. Universidad de Salamanca.
- Vallejo-Arboleda, A., Vicente-Villardón, J. L., & Galindo-Villardón, M. P. (2007). Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Computational Statistics and Data Analysis*, 51(9), 4193-4205.
<https://doi.org/10.1016/j.csda.2006.04.032>
- Vallejo-Arboleda, A., Vicente-Villardón, J. L., Galindo-Villardón, M. P., Fernández, M., Fernández, C., & Bécares, E. (2008). Análisis de la evolución en el tiempo para datos con estructura de grupos: STATIS dual canónico y modelo de medidas repetidas doblemente multivariantes. *Revista Colombiana de Estadística*, 31(2), 321-340.
- Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. AL, & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, 10(1), 246. <https://doi.org/10.1186/1471-2105-10-246>
- Vera, J. F. V., & Carmona, A. G. (1991). Statistical and Computational Analysis of data tables. En R. (Gutiérrez J. Gutiérrez & M. J. Valderrama (Eds.), *5th International Symp on Applied stochastic models and data analysis* (p. 656). Granada, Spain: World Scientific.
- Vicente-Galindo, P. (2013). Análisis de tablas de tres vías : recientes desarrollos del STATIS. *Trabajo fin de master*, 1-65.
- Vicente-Villardón, J. L. (2010). MULTBILOT: A package for Multivariate Analysis using Biplots. Spain: Departamento de Estadística. Universidad de Salamanca. Recuperado de <http://biplot.usal.es/ClassicalBiplot/index.html>
- Vicente-Villardón, J. L. (2015). MultBiplotR: Multivariate Analysis using Biplots. R package version 0.1.0.
- Vichi, M., & Saporta, G. (2009). Clustering and disjoint principal component analysis. *Computational Statistics & Data Analysis*, 53(8), 3194-3208.
- Vigneau, E., & Qannari, E. M. (2003). Clustering of Variables Around Latent Components. *Communications in Statistics Part B: Simulation and Computation*.
<https://doi.org/10.1081/SAC-120023882>
- Villa-Vialaneix, N. (2016). SirStatis. R package for longitudinal STATIS analyses. Recuperado de <https://github.com/tuxette/SirStatis>

- Villa-Vialaneix, N., Sautron, V., & Chavent, M. (2016). Multiway-SIR for Longitudinal Multi-table Data Integration Sommaire. En *22nd International Conference on Computational Statistics (COMPSTAT)*. Oviedo. Recuperado de <https://hal-irsrn.archives-ouvertes.fr/INRA/hal-01416735v1>
- Vines, S. K. (2000). Simple principal components. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(4), 441-451. <https://doi.org/10.1111/1467-9876.00204>
- Viswanathan, V., Lee, C. E., Lees, M. H., Cheong, S. A., & Sloot, P. M. A. (2014). Quantitative comparison between crowd models for evacuation planning and evaluation. *European Physical Journal B*, *87*(2), 1-12. <https://doi.org/10.1140/epjb/e2014-40699-x>
- Vivien, M., & Sabatier, R. (2004). A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Computational Statistics & Data Analysis*, *46*(1), 155-171. [https://doi.org/10.1016/S0167-9473\(03\)00146-4](https://doi.org/10.1016/S0167-9473(03)00146-4)
- Vivien, M., & Sune, F. (2009). Two four-way multiblock methods used for comparing two consumer panels of children. *Food Quality and Preference*, *20*(7), 472-481. <https://doi.org/10.1016/j.foodqual.2009.04.004>
- Volkova, M. (2019). Russian and European Population's Quality of Life Analysis with the Instruments of Common Principal Components (CPC). *Economics and the Mathematical Methods*, *55*(3), 34. <https://doi.org/10.31857/S042473880004678-4>
- Wendling, Z. A., Emerson, J. W., de Sherbinin, A., Esty, D. C., Hoving, K., Ospina, C. D., ... Ballesteros-Figueroa, A. (2020). *2020 Environmental Performance Index*. New Haven, CT. Recuperado de epi.yale.edu
- Wilderjans, T., Ceulemans, E., & Van Mechelen, I. (2008). The CHIC Model: A Global Model for Coupled Binary Data. *Psychometrika*, *73*(4), 729-751. <https://doi.org/10.1007/s11336-008-9069-9>
- Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, *10*(3), 515-534.
- Zambrano, A. Y., Fuenmayor, F., Gutiérrez, Z., Rodríguez, A., Ruiz, L., Moreno, R., ... Vicente-Villardón, J. L. (2007). Genetic diversity of venezuelan cassava collection. *Acta Horticulturae*, (738), 729-733. <https://doi.org/10.17660/ActaHortic.2007.738.96>
- Zhang, Z., Xu, Y., Yang, J., Li, X., & Zhang, D. (2015). A survey of sparse representation: algorithms and applications. *IEEE*, 490-530.
- Zingaretti, L. M., Renand, G., Morgavi, D. P., & Ramayo-Caldas, Y. (2020). Link-HD: a versatile framework to explore and integrate heterogeneous microbial communities. *Bioinformatics*,

36(7), 2298-2299. <https://doi.org/10.1093/bioinformatics/btz862>

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*., 67(2), 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2), 265-286. <https://doi.org/10.1198/106186006X113430>