VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO DE FIN DE GRADO

EL IMPACTO DEL COVID-19 EN LOS DELITOS DE ODIO: UN ANÁLISIS ESTADÍSTICO

THE IMPACT OF COVID-19 ON HATE CRIMES:

A STATISTICAL ANALYSIS

GRADO EN ESTADÍSTICA

Facultad de Ciencias

Curso 2023/2024

Autora: Azahara Manuel Azcaray

Tutora: Mercedes Sánchez Barba

Salamanca, 2024

TRABAJO DE FIN DE GRADO

EL IMPACTO DEL COVID-19 EN LOS DELITOS DE ODIO: UN ANÁLISIS ESTADÍSTICO

THE IMPACT OF COVID-19 ON HATE CRIMES:

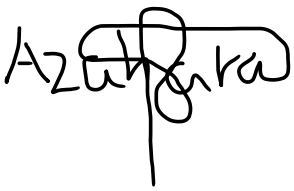
A STATISTICAL ANALYSIS

GRADO EN ESTADÍSTICA

Facultad de Ciencias

Curso 2023/2024

Autora: Azahara Manuel Azcaray



Tutora: Mercedes Sánchez Barba

SANCHEZ BARBA MARIA DE LAS MERCEDES - 07966806C

Firmado digitalmente por SANCHEZ BARBA MARIA DE LAS MERCEDES - 07966806C
Fecha: 2024.07.04 08:03:44 +02'00'

Salamanca, 2024

Certificado del/los tutor/es TFG

D./Dña. Meredes Sánchez Barba, profesor/a del Departamento de Estadística de la Universidad de Salamanca,

HACE/N CONSTAR:

Que el trabajo titulado "El impacto del Covid -19 en los Delitos de Odio: Un análisis Estadístico" , que se presenta, ha sido realizado por Azahara Manuel Azcaray, con DNI ****1033J y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Grado en Estadística en esta Universidad.

Salamanca , 4 de julio de 2024

Fdo.: SANCHEZ BARBA MARIA DE LAS MERCEDES - 07966806C

Firmado digitalmente por SANCHEZ BARBA MARIA DE LAS MERCEDES - 07966806C
Fecha: 2024.07.04 08:02:47 +02'00'

Resumen

Durante la última década, los delitos de odio han generado una creciente preocupación mundial, exacerbada por la pandemia del COVID-19, que aumentó la intolerancia y discriminación en la sociedad. Estos delitos, motivados por prejuicios hacia características personales de la víctima, incluyen actos de racismo, xenofobia, homofobia, entre otros. La investigación analiza la evolución de los delitos de odio en España entre 2018 y 2022, examina las estructuras de covariación de variables relacionadas y agrupa las comunidades autónomas según su perfil de victimización, detenciones y hechos esclarecidos, para determinar si la pandemia ha influido significativamente en estos delitos.

Para llevar a cabo la investigación se han utilizado el análisis de correlaciones, el método de STATIS y el análisis de clúster k-means. Este estudio revela una tendencia a la baja en delitos de odio por aporofobia, pero aumentos en los relacionados con identidad de género y racismo/xenofobia. El análisis STATIS indica una continuidad en los patrones subyacentes con algunas variaciones posiblemente influenciadas por la pandemia. A nivel regional, Cataluña destaca por una alta incidencia, mientras que otras como Andalucía, Comunidad Valenciana, Madrid y el País Vasco muestran perfiles variados en la severidad de los delitos. La crisis del COVID-19 ha dejado un impacto notable, siendo el País Vasco particularmente destacado por la frecuencia y gravedad de los incidentes.

Finalmente, la investigación subraya la importancia de estrategias colaborativas entre comunidades, fuerzas de seguridad y organizaciones civiles para abordar estos desafíos.

Abstract

During the last decade, hate crimes have generated increasing global concern, exacerbated by the COVID-19 pandemic, which heightened intolerance and discrimination in society. These crimes, motivated by prejudice against the personal characteristics of the victim, include acts of racism, xenophobia, homophobia, among others. The research analyzes the evolution of hate crimes in Spain between 2018 and 2022, examines the covariation structures of related variables, and groups the autonomous communities according to their profiles of victimization, arrests, and solved cases, to determine whether the pandemic has significantly influenced these crimes.

To carry out the research, correlation analysis, the STATIS method, and k-means cluster analysis were used. This study reveals a downward trend in hate crimes related to aporophobia, but increases in those related to gender identity and racism/xenophobia. The STATIS analysis indicates a continuity in underlying patterns with some variations possibly influenced by the pandemic. At the regional level, Cataluña stands out for its high incidence, while others like Andalucia, the Valencian Community, Madrid, and País Vasco show varied profiles in the severity of the crimes.

Finally, the research underscores the importance of collaborative strategies between communities, security forces, and civil organizations to address these challenges.

ÍNDICE

1. INTRODUCCIÓN

Durante esta última década, el fenómeno de los delitos de odio ha suscitado una creciente preocupación en todo el mundo, generando una profunda inquietud de intolerancia y discriminación en nuestra sociedad. Además, el estallido de la pandemia ocasionada por el virus SARS-CoV-2 en 2020, provocó un cambio en la normal convivencia entre ciudadanos, generando la aparición de diferentes tipos de conductas y comportamientos perjudiciales, entre los que se encuentran los delitos de odio (Giraldo Pérez, 2022).

Los delitos de Odio, *bias crimes* o *hate crimes* comúnmente conocidos, son una expresión para definir cualquier delito cometido por un sujeto impulsado por prejuicios hacia un estereotipo basado en una característica personal de la víctima, independientemente de cuál sea esa característica física, como puede ser su ideología, su raza, su orientación sexual.... Este concepto está estrechamente relacionado con la intolerancia y el motivo discriminatorio o prejuicioso del perpetrador (Díaz López, 2020). Para otros, este concepto es simplemente la pertenencia de la víctima a un grupo social específico odiado por el autor, más que al odio en sí mismo.

Además, el término "delito de odio" va más allá de simplemente violar el principio de igualdad (por ejemplo, un delito contra los derechos de los trabajadores), sino que, por su naturaleza también afectan la dignidad de la persona (como el caso de una agresión violenta contra un individuo por su orientación sexual). Sin embargo, es difícil obtener una definición clara de este concepto debido a la complejidad de establecer los grupos o características que deben ser considerados protegidos bajo la categoría de delitos de odio (Güerri Ferrández, 2015).

Un término muy relacionado con los "hate crimes" son los "hate speech" o discurso de odio, que se definen como expresiones o discursos difundidos a través de diversos medios sociales con la intención de promover el odio, los prejuicios y la violencia contra individuos específicos. Además, también pueden tener el propósito de degradar, intimidad o discriminar a las personas por motivos como su raza, etnia, religión, nacionalidad, género u otros aspectos (Rodríguez Expósito, 2016).

Este término ha tomado gran relevancia y su incidencia ha aumentado, frente a los delitos de odio, y esto se debe principalmente a un par de factores. Lo primero, el surgimiento de partidos de extrema derecha cuyos mítines provocan la incitación al odio en varios países europeos, junto con la popularización del Internet que facilita la difusión y complica la persecución de este tipo de declaraciones, han ayudado a la necesidad de combatir el discurso de odio. Por consiguiente, la intención de combatir el discurso del odio ha generado un amplio debate académico y doctrinal, ya que va en contra de otros derechos básicos como la libertad de expresión (Güerri Ferrández, 2015).

Sin embargo, aunque el discurso de odio pueda conducir al delito de odio, en muchas ocasiones estos dos términos no pueden considerarse equivalentes ni intercambiables. Esta equivalencia sólo será válida en el caso de delitos de odio en forma de "delitos expresivos": amenazas, delitos contra la integridad moral, glorificación y humillación de las víctimas del terrorismo, apología…(Cámara Arroyo, 2017).
A lo largo de nuestra historia, actos similares han ocurrido y, en ciertos períodos, el odio ha aumentado significativamente. Europa tiene una larga tradición de intolerancia, marcada por periodos de esclavitud, colonialismo e inmigración. Esta historia

compartida puede explicar, en parte, la ausencia de una definición legal y social unificada de los delitos de odio en Europa hasta la fecha. Cada país aborda los crímenes de odio según su contexto social y temporal. Aunque a principios de los años ochenta, algunos países anglosajones comenzaron a introducir legislación específica para combatir lo que denominaron "hate crimes", años más tarde en ciertos países aún se observaba un marcado desinterés político, social, legislativo y jurídico en detener no solo este tipo de delitos, sino también el discurso del odio que los precede ("hate speech") (Pérez Álvarez et al., 2016).

En España, hasta 2014, diversos organismos internacionales afirmaban que el país era uno de los pocos países de Europa que no registraba datos sobre delitos de odio. Sin embargo, el primer informe sobre delitos de odio en España, publicado en diciembre de 2013, reconocía que España había logrado avances significativos en la lucha contra dichos delitos desde 2010. A partir de ese momento, España pasó de ser uno de los países que no registraban apenas datos, a uno de los cinco primeros en recopilar más información sobre este fenómeno. Aunque, como discutiremos más adelante, estos actos calificados como racistas o xenófobos presentan muchas limitaciones en cuanto a la fiabilidad de los datos, debido principalmente a la falta de formación específica y el desconocimiento de estos tipos de actos por parte de los miembros de las fuerzas y cuerpos de seguridad del Estado (López Ortega, 2017).

Aunque este progreso en combatir estos actos es evidente, actualmente Europa se enfrenta a desafíos cada vez mayores. La globalización ha provocado un aumento de la movilidad humana y, por tanto, de la migración, lo que ha dado lugar a sociedades multiculturales. Sin embargo, este proceso va acompañado de una creciente sensación de inseguridad en las comunidades de riesgo, lo que aumenta el desprecio por quienes son percibidos como diferentes. Este choque de culturas ha desembocado en conflictos sociales y jurídicos, llevando a invocar el derecho penal como herramienta para abordar las conductas discriminatorias contra minorías. Asimismo, organismos nacionales e internacionales han reafirmado su compromiso con la prevención y el procesamiento de los delitos de odio, impulsadas por estudios que muestran la necesidad de combatir la indiferencia y la impunidad que alimentan la intolerancia (García Domínguez, 2020) . El racismo, la xenofobia y los delitos de odio en general representan, en la era de la globalización, el gran peligro de las sociedades abiertas (Giménez-Salinas Colomer et al., 2003).

Lo verdaderamente preocupante es la falta de sensibilización sobre estas acciones. Al igual que sucedió con la violencia de género, mucha parte de la sociedad no reconoce plenamente la gravedad de esta problemática, ni contamos con las suficientes herramientas necesarias para detectar, controlar y abordar esta otra forma de perjuicio que ha estado presente siempre en cualquier sociedad: el rechazo a lo diferente (Pérez Álvarez et al., 2016).

Con respecto a la víctima directa, muchos estudios han demostrado que el sufrimiento psicológico experimentado es mayor en las víctimas de delitos de odio en comparación con las víctimas de delitos violentos que no tienen una motivación prejuiciosa. Esto se debe a que el delito no solo afecta a la víctima directa, sino también a la comunidad a la que pertenece. Pero en muchas ocasiones, este riesgo aumenta cuando la víctima no se identifica con el grupo al que pertenece y se culpa así misma por la victimización causando la internalización de evaluaciones negativas sobre sí misma. Por otro lado, las víctimas que se identifican fuertemente con su grupo buscarán apoyo en él y fortalecerán

su identidad como mecanismo de afrontamiento. Sin embargo, esto también puede llevarlas a temer la revictimización, ya que han sido atacadas por aspectos que forman parte intrínseca de su identidad y que no pueden cambiar (Güerri Ferrández, 2015).

Estos delitos pueden tener distintas consecuencias emocionales, psicológicas y sociales dependiendo de la trascendencia del hecho y el carácter de la víctima, así como en función del tiempo transcurrido. Entre las principales consecuencias psicológicas que pueden presentarse luego de ser víctima de un delito son: depresión, ansiedad y sobre todo el trastorno de estrés postraumático (Güerri Ferrández, 2015).

En cuanto al agresor, según estudios que se han llevado a cabo, suelen ser hombres, de raza blanca mayoritariamente, jóvenes (menores de 30 años), desempleados y suelen cometer la agresión en grupo. Por otro lado, entre las motivaciones para llevar a cabo este tipo de actos se encuentran: la búsqueda de emociones, la defensiva, la vengativa o de misión (Pérez et al., 2021).

Como parte de esta investigación, es necesario examinar los diferentes tipos de delitos de odio que son objeto de análisis. Esta clasificación permite comprender la diversidad y complejidad de este fenómeno social. A continuación, se describen los diferentes ámbitos que se computan como "delitos de odio":

- **Racismo/xenofobia:** Se refiere a cualquier incidente considerado racista o xenófobo por la víctima, así como por cualquier otra persona presente, incluyendo agentes policiales u otros testigos, incluso si la víctima no está de acuerdo con esa percepción. Esto abarca actos de odio, violencia, discriminación, fobia y rechazo dirigidos hacia extranjeros o personas pertenecientes a diferentes grupos por su origen o religión racial, étnica, nacional o cultural.
- **Orientación sexual e identidad de género:** cualquier incidente que evidencie la presencia de motivos de odio o discriminación dirigidos hacia la víctima debido a su orientación sexual o identidad de género.
- **Creencias o prácticas religiosas:** cualquier acto que indique la presencia de un motivo de odio o discriminación hacia la víctima debido a sus creencias religiosas. También incluye los actos cometidos contra ateos y agnósticos con este motivo, aunque no se incluyen por motivos antisemitas.
- **Antisemitismo:** cualquier hecho de odio, discriminación, violencia, miedo o rechazo dirigido contra judíos o ciudadanos del Estado de Israel.
- **Discapacidad:** Este ámbito se define como cualquier acto contra una persona discapacitada o con diversidad funcional, en el que el responsable del incidente intervenga contra la víctima y esté motivado por la discriminación o relacionado con un delito de odio.
- **Aporofobia:** El motivo debe ser el odio o el rechazo a los pobres. Implican aquellas expresiones y actos de intolerancia que se relacionan con el "odio, repugnancia o actitud hostil hacia los pobres necesitados e indefensos"
- **Ideología:** Cualquier acto que indique la presencia de un motivo de odio o discriminar a las víctimas en función de su percepción de aspectos relacionados con los sistemas económicos, la sociedad, la cultura o la política.
- **Discriminación por razón de sexo/género:** Cualquier incidente que denote la presencia de motivos de odio o discriminación hacía la víctima debido a su género (hombre o mujer), o dirigido específicamente contra mujeres simplemente por serlo, con la intención de ejercer dominio y mostrar su sentido

de superioridad sobre ellas. Se excluyen de esta categoría los incidentes cometidos contra la orientación sexual e identidad de género.

Por otro lado, a partir del 2018, se comenzó a contabilizar estos tres nuevos ámbitos:

- **Discriminación por razón de enfermedad:** Cualquier acción realizada por motivos discriminatorios contra una persona que padece una enfermedad temporal o permanente que afecta su salud física o mental y se considera un factor distintivo, el tratamiento basado únicamente en la existencia de esa enfermedad o la estigmatización de la persona involucrada, constituye una causa de discriminación.
- **Discriminación generacional:** Trato desigual o degradante hacía una persona o conjunto de personas debido a su edad. Este tipo de discriminación se manifiesta principalmente en forma de gerofobia, que implica sentimientos hostiles y acciones discriminatorias dirigidas hacia las personas de edad avanzada.
- **Antigitanismo:** Todas estas acciones llevadas a cabo sobre la base de la discriminación, el odio y el estigma dirigidas contra las personas gitanas y su entorno (Metodología Estadística, n.d.).

La investigación sobre los delitos de odio y su evolución durante la pandemia COVID-19 se fundamenta en una serie de desafíos actuales y la urgente necesidad de abordarlos mediante programas de mejora en diversas áreas cruciales.

Entre ellas se encuentran la ausencia de estadísticas oficiales, dificultando la comprensión completa del problema y la planificación de intervenciones efectivas. A su vez, la falta de denuncias en estos casos influye en este problema, pues el número de delitos no denunciados es considerablemente alto en el caso de los delitos de odio.
Por otro lado, la limitación de recursos y la capacitación inadecuada del personal responsable de investigar y procesar estos actos contribuyen a una situación en la que la capacidad del sistema judicial para garantizar la protección y la justicia para las víctimas se ve afectada. Asimismo, el aumento del discurso de odio hacia ciertos grupos en actos públicos o a través de internet durante la pandemia, resalta la importancia de implementar estrategias de prevención y educación para contrarrestar esta tendencia y promover la cohesión social (Güerri Ferrández, 2015).

Entre algunas líneas de acción que el Ministerio del Interior ha publicado en el “II Plan de Acción de Lucha contra los Delitos de Odio 2022-2024” para prevenir e impulsar la lucha contra este tipo de actos se encuentran: brindar asistencia y apoyo a las víctimas de delitos de odio, mejorar la coordinación entre las Fuerzas y Cuerpos de Seguridad y otras instituciones, prevenir la comisión de delitos de odio con el desarrollo de herramientas para mejorar la efectividad de las investigaciones, crear grupos de lucha contra este tipo de delitos, etc. (II Plan de Acción de Lucha Contra Los Delitos de Odio, 2022)

Para abordar esta problemática se empleará un enfoque metodológico basado en el análisis STATIS, que permitirá identificar las estructuras de covariación de las variables a lo largo del tiempo y con el análisis de K-Means agrupar las comunidades autónomas en función del perfil de victimización y otros indicadores.

El trabajo se estructura en varias secciones. En primer lugar, se establecerán los objetivos del estudio. A continuación, se describirán los métodos utilizados y los datos

analizados. Posteriormente, se presentarán los resultados del análisis y, finalmente, se discutirán las conclusiones y las implicaciones para futuras investigaciones y políticas públicas.

2. OBJETIVOS

Los objetivos del trabajo son:

- Analizar la evolución de los delitos de odio entre los años 2018 y 2022 según el ámbito en el que se han producido.
- Estudiar las estructuras de covariación de las variables relacionadas con los delitos de odio a lo largo de estos años, y determinar si existen diferencias en la estructura entre los distintos periodos.
- Agrupar las diferentes comunidades autónomas en función de su perfil de victimización, detenciones, hechos conocidos y hechos esclarecidos, para identificar patrones similares o distintos entre ellas.

Finalmente, con estos análisis, se pretende concluir si el COVID-19 ha podido tener un impacto significativo en los delitos de odio en España, evaluando cómo la pandemia ha influido en la incidencia y la naturaleza de estos delitos.

3. MATERIAL

Para la realización de este trabajo se construyó una base de datos, a partir de los datos de acceso público disponibles en la fuente publica del Portal Estadístico de la Criminalidad. Para la selección de variables, se han elegido los delitos de odio y se han extraído los datos sobre las victimizaciones, detenciones e investigados, hechos conocidos y hechos esclarecidos por causa de delitos de odio desglosados por tipo de hecho. Dentro de cada uno de ellos se han seleccionado la calificación total, por Comunidades Autónomas, por ámbito de aporofobia, creencias o prácticas religiosas, delitos de odio contra personas con discapacidad, discriminación generacional, discriminación por razón de enfermedad, discriminación por razón de sexo/género, ideología, orientación sexual e identidad de género y racismo/xenofobia; sexo total y para el periodo de 2018 a 2022.

Una vez recopilados todos los almacenes de datos individuales, se procede a diseñar las variables en el almacén de datos final y a completarlas con los datos correspondientes. Para ello se crea la variable “año” la cual abarcará el periodo comprendido entre 2018 y 2022. La variable “comunidad” que identifica la región autónoma asociada a cada registro. El campo “ámbito” que delinea las áreas específicas de la actividad delictiva seleccionada, seguidas de las variables num_vict (número de victimizaciones) que se refiere al recuento de incidentes denunciados por individuos que afirman ser víctimas o perjudicados por alguna infracción; num_detenc (número de detenciones) cantidad de veces que las autoridades han arrestado a personas, privándolas de su libertad, debido a su presunta participación en actividades delictivas, en este caso relacionadas con los delitos de odio; num_ hec_con (número de hechos conocidos) que se refieren al conjunto de infracciones penales y administrativas que han llegado a la atención de las diversas Fuerzas y Cuerpos de Seguridad, ya sea a través de denuncias presentadas o mediante acciones policiales iniciadas de oficio; y num_hec_esc (número de hechos esclarecidos) considerados aquellos en el que haya detención del autor en el acto, identificación completa del autor, existencia de una confesión corroborada o que la investigación demuestre que no había infracción.

Completadas estas variables, se añadirá otra más relacionada con la demografía de las comunidades autónomas, la variable "denspob" (densidad de población), definida como el número de personas que residen en una unidad de superficie, comúnmente expresada en kilómetros cuadrados. Es esencial incorporar una variable demográfica con el fin de facilitar la comparación entre las comunidades autónomas dado que cada una presenta diferencias significativas en términos de superficie y población. La omisión de esta variable podría conducir a conclusiones estadísticas incorrectas.

Se empleó la herramienta Excel para procesar y organizar la base de datos. Posteriormente, para llevar a cabo el análisis de los datos, se utilizó el programa R Studio.

4. MÉTODO

A continuación, se procederá a brindar una explicación teórica de los métodos empleados en el desarrollo del estudio.

4.1.Datos de tres modos

Los datos tri-modales se refieren a observaciones que se organizan en tres diferentes categorías. Estos modos abarcan un conjunto de unidades de observación u objetos, un conjunto de variables, un conjunto de ocasiones, entre otros.

El arreglo de tres modos consiste en identificar a los individuos objeto de estudio mediante un primer índice, se miden las variables correspondientes con una segunda entrada y un tercer índice que abarca las diversas situaciones (condiciones u ocasiones). Este enfoque permite examinar las diferencias y similitudes entre distintos escenarios, basándose en las configuraciones de los individuos y las relaciones entre los diferentes grupos de variables. En este contexto, el primer modo se relaciona con los objetos, el segundo con las variables y el tercero con las ocasiones. Un ejemplo de las mismas variables en el mismo conjunto de objetos son los datos longitudinales.

Podemos clasificarlos en: datos de conjuntos múltiples y datos de tres vías que explicaremos más detalladamente a continuación (Rodríguez Martínez, 2020):

Los datos tres vías se refieren a un conjunto de datos que abarcan todas las observaciones de cada objecto, en todas las variables y en todas las ocasiones. Estos datos se pueden representar mediante un arreglo tridimensional completo y se dividen en:

- Datos de tres vías tres modos: incluyen tres conjuntos de diferentes de entradas, como individuos, variables y ocasiones.
- Datos de tres vías dos modos: en este caso, dos de los modos comparten el mismo conjunto de entradas. Así como, se emplean matrices de correlación para las mismas variables en diversas muestras, siendo las variables y las muestras los dos modos.
- Datos de tres vías un modo: Cada uno de los modos está relacionado con el mismo conjunto de entradas

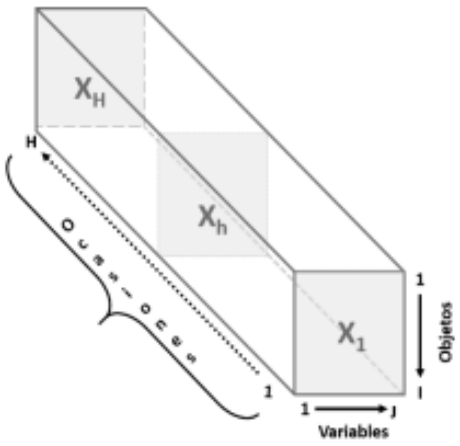


Ilustración 1: Representación de Datos Tridimensionales/ Fuente: (Rodríguez Martínez, 2020)

Se emplea la siguiente notación para representar los conjuntos de datos tridimensionales:

x_{ijh} es un elemento del conjunto con $i = 1, \ldots, I$ (individuos u objetos); $j = 1, \ldots, J$ (variables); y $H = 1, \ldots, H$ (situaciones u ocasiones).

La matriz X_h representa un conjunto de datos tridimensional, donde I denota el número de objectos, J el número de variables, y H el número de ocasiones o condiciones. Cada entrada de esta matriz corresponde a la observación de un objeto en una variable específica durante una ocasión determinada.

Por otro lado, los conjuntos de datos múltiples implican observaciones de diversos conjuntos de objetos, como lugares o individuos, bajo un conjunto de variables común. Generalmente, se supone que estas observaciones se realizan en diferentes momentos o escenarios. No es viable visualizar directamente estos conjuntos de datos en una disposición tridimensional. En lugar de eso, se representan combinando las observaciones de cada conjunto de objetos en una matriz, donde cada fila corresponde a un objeto y cada columna a una variable. Estas matrices se organizan en una super matriz, donde cada matriz individual representa un conjunto de objetos, dispuestas una encima de la otra.

La notación utilizada para describir estos datos incluye I_h, que indica el número de objetos observados en la ocasión H, siendo estas ocasiones variables en cuanto al número de objetos. Además, se emplea I_{ijh} para denotar un elemento específico en esta matriz combinada, donde i representa un objeto, j una variable y h la ocasión o situación. Cada matriz individual X_h contiene las observaciones de los objetos para las variables correspondientes en una ocasión específica. Es crucial señalar que las matrices individuales no tienen necesariamente el mismo tamaño o disposición.

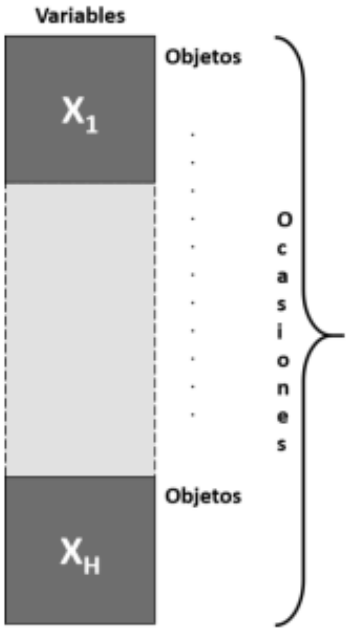


Ilustración 2: Representación de Datos de Conjuntos Múltiples/ Fuente: (Rodríguez Martínez, 2020)

Dentro del ámbito de los datos de conjuntos múltiples, también se incluyen aquellos en los que se registran observaciones de un conjunto de objetos bajo diferentes conjuntos de variables. Este tipo de estructura de datos se caracteriza por su versatilidad y aplicabilidad en diversos campos de estudio.

Para comprender mejor este tipo de datos, es crucial considerar su notación estándar. La variable I representa el número total de objetos observados en el conjunto de datos. Cada objeto individual se denota mediante el índice i, que varía desde 1 hasta I. Por otro lado, la variable J_h hace referencia al número de variables observadas en cada ocasión o situación. Asimismo, se utilizan los subíndices j y h para indicar las variables específicas en cada ocasión.

La representación de estos datos se simplifica mediante una matriz X_h de dimensiones I x J_h. En esta matriz, cada fila corresponde a un objeto individual y cada columna representa una variable específica en una ocasión determinada. Es esencial destacar que las dimensiones de la matriz pueden variar entre diferentes ocasiones, reflejando así la naturaleza dinámica y cambiante de los conjuntos de datos múltiples.

4.2.Análisis de Datos de Tres Modos

En la literatura, existen varios enfoques para el estudio de la integración de matrices. Se pueden clasificar según su origen en la Escuela Anglosajona y la Escuela Francesa, o según el ajuste de las matrices. Algunos métodos ajustan directamente a los datos de tres vías, mientras que otros se ajustan a matrices derivadas de estos datos.

Además, los métodos se pueden clasificar en simétricos y asimétricos. Los simétricos tratan todos los modos (individuos, variables, ocasiones) por igual y se asocian con métodos de componentes, mientras que los asimétricos tratan un modo de forma diferente y se enfocan en el esquema Interestructura-Compromiso-Intraestructura (ICI) (Rodríguez Martínez, 2020).

Dentro de la Escuela Anglosajona, destacan los modelos de las familias Tucker y PARAFAC, estos se enfocan en las variaciones en las estructuras de los elementos o variables idénticos a lo largo de diferentes ocasiones. Estos métodos son simétricos y buscan minimizar la función de pérdida, lo que significa que intentan que los datos reproducidos por el modelo se parezcan lo más posible a los datos originales. De manera general, se emplea el término "three-mode data" para describir las tablas de tres modos (Rodríguez Martínez, 2020).

La Escuela Francesa clasifica los datos de tres modos como Tablas de Tres Índices (Tableaux à Trois Indices). Entre los enfoques más destacados de esta escuela, sobresalen los modelos de la familia STATIS, el Análisis Factorial Múltiple (AFM) y el Análisis Triádico. Estos enfoques difieren en cómo se establece un referencial denominado compromiso, que facilita la representación de sujetos y variables en un mismo subespacio para diferentes situaciones. Además, incluyen el método LONGI, diseñado para el análisis de datos longitudinales (Rodríguez Martínez, 2020).

Por otro lado, existen otros enfoques para analizar datos en tablas de tres modos como las técnicas BIPLOT, de la Escuela Salmantina. Estas técnicas permiten la representación conjunta de individuos y variables. La combinación del BIPLOT con otras técnicas ha generado innovadores métodos de análisis multivariado. Entre estos, destacan el Biplot Conjunto y el Biplot Interactivo, en los cuales el Departamento de Estadística de la Universidad de Salamanca ha colaborado significativamente.
Otra importante contribución de la escuela salmantina al estudio de datos tridimensionales es el ACPR a tres vías. Esta técnica amplía el enfoque del Análisis de Componentes Principales Restringido (ACPR) al ámbito tridimensional, integrando información externa bidimensional proporcionada en matrices. Este enfoque no solo proporciona una visión más completa del ACPR, sino que también permite obtener resultados valiosos, como el Análisis Canónico de Correspondencias (ACC) tridimensional integrando información suplementaria bidimensional (Rodríguez Martínez, 2020).

En términos de ajuste, el Ajuste Directo se refiere al trabajo con las matrices originales, mientras que el Ajuste de Matrices Derivadas utiliza matrices derivadas, como las de productos cruzados.

En nuestro caso, se presentará una explicación más detallada de los métodos asimétricos de la familia STATIS, que es el que utilizaremos para realizar el estudio de nuestros datos más adelante.

4.3.Métodos familia STATIS

La familia de métodos STATIS fue introducida por Escoufier en 1976 y L'Hermier des Plantes en 1976 y más tarde ampliados por Lavit en 1988. Estos métodos representan una extensión del análisis de componentes principales (ACP) y permiten analizar varias tablas de datos simultáneamente. Se aplican a datos cuantitativos recolectados en estos casos (Rodríguez Martínez, 2020):

- Se dispone de H matrices de datos que contienen información recopilada en distintas ocasiones sobre los mismos sujetos. Las variables pueden variar entre cada matriz. Cada estudio se representa como $\boldsymbol{X_h, M_h, D}$, donde h varía de 1 a H. El propósito de este análisis es entender las relaciones relativas entre los sujetos, y esto se alcanza utilizando el método STATIS.

- Se disponen de H conjuntos de individuos, que pueden variar en cada grupo, midiendo las mismas variables en todos ellos. Estos estudios se representan como $\boldsymbol{X_h, M, D_h}$ para h desde 1 hasta H. El propósito principal es investigar las interrelaciones entre las variables, un objetivo que se logra eficazmente utilizando la técnica STATIS dual.

Cuando se desea examinar matrices que cruzan las mismas variables con los mismos sujetos, se recomienda utilizar ambos métodos. A pesar de sus similitudes, estos métodos parten de diferentes enfoques. El método STATIS comienza con la matriz de similitud entre individuos $\boldsymbol{XX^T}$ y enfatiza las posiciones de los individuos, a diferencia del método STATIS DUAL que se basa en la matriz que captura la estructura de covariación y enfatiza las relaciones entre variables, $\boldsymbol{X^T X}$ (Rodríguez Martínez, 2020). A continuación, se presenta un esquema con dichas diferencias entre ambos métodos:

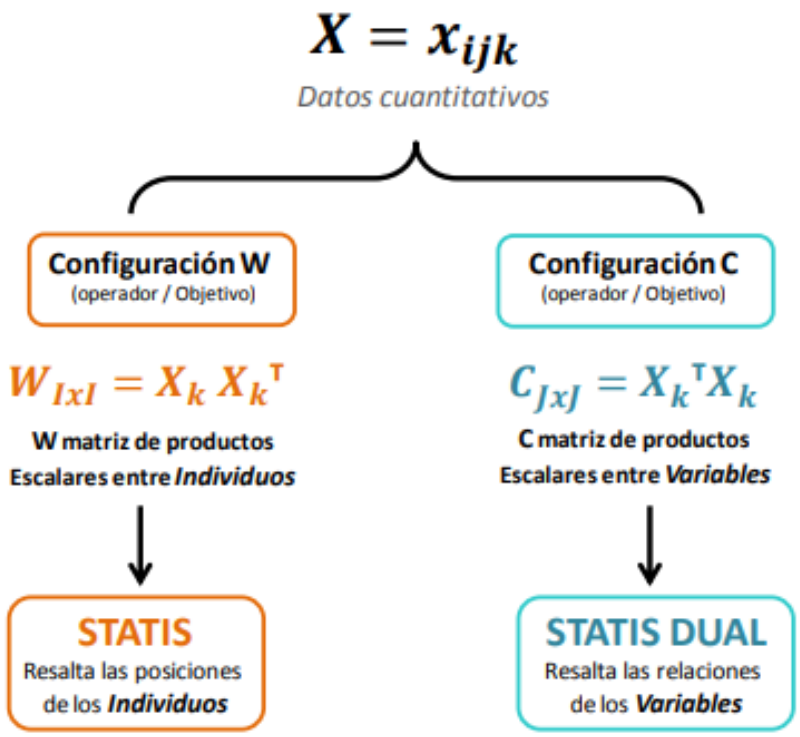


Ilustración 3: Estructura del STATIS y STATIS DUAL / Fuente: (Galindo, 2022)

Para establecer la disposición de las matrices de datos $\boldsymbol{W_k}$ y $\boldsymbol{C_k}$, es crucial establecer un índice de correlación entre estas matrices y calcular las distancias correspondientes. Esto se logra mediante la definición del producto escalar entre los pares de matrices, el cual debe inducir una norma y, consecuentemente, una distancia. Este producto escalar específico es conocido como el producto escalar de Hilbert-Schmidt (HS) (Ballesteros Espinoza, 2022):

$$\langle \boldsymbol{C_k} | \boldsymbol{C_{k'}} \rangle_{\boldsymbol{HS}} = \boldsymbol{tr}(\boldsymbol{C_k C_{k'}})$$

Usando el producto escalar, puede determinar la norma y la distancia Hilbert-Schmidt entre los conjuntos:

$$\| \boldsymbol{C_k} \|^2_{\boldsymbol{HS}} = \langle \boldsymbol{C_k} | \boldsymbol{C_{k'}} \rangle_{\boldsymbol{HS}}$$

$$d_{HS}\langle C_k | C_{k'}\rangle = |C_k - C_{k'}|$$

Además, es fundamental establecer una correlación vectorial basada en el producto escalar de Hilbert-Schmidt, técnica introducida por Escoufier en 1973:

$$r_{kk'} = \frac{S_{kk'}}{\sqrt{S_{kk}}\sqrt{S_{k'k'}}}$$

En el cual,
$S_{kk'}$ representa el producto escalar entre los conjuntos de configuraciones
S_{kk} representa la norma de la configuración S

Cuando la correlación entre matrices alcanza uno, indica que son idénticas en términos de estructura y disposición, lo que implica una fuerte similitud entre las estructuras subyacentes. Cuanto más próximo a uno sea este valor, mayor será la similitud entre las estructuras factoriales (Ballesteros Espinoza, 2022).

El proceso tanto para STATIS como para STATIS dual se desarrolla en cuatro fases (Galindo, 2022):

1. Determinación de la inter-estructura: Se genera una matriz de correlaciones vectoriales entre las diferentes matrices de datos.
2. Representación de la inter-estructura en un espacio reducido: Se realiza una descomposición espectral de la matriz de correlaciones vectoriales y se proyecta en un subespacio de menor dimensión.
3. Análisis de la inter-estructura: Se interpreta el diagrama factorial que resulta de la proyección en el subespacio reducido.
4. Cálculo de la estructura consenso: Se obtiene la matriz consenso como un promedio ponderado de las matrices originales.
5. Estudio de la intra-estructura: Se analizan las trayectorias y relaciones dentro de la estructura consenso para entender mejor las dinámicas internas de los datos.

4.3.1. STATIS

Originario de la escuela francesa, este enfoque se presenta como otra opción para el análisis de conjuntos de datos cuantitativos organizados en k-tablas. Estas tablas pueden consistir en semejantes sujetos observados en diferentes momentos con las mismas variables, los semejantes individuos evaluados en distintos momentos con variables distintas, o distintos grupos de sujetos con variables similares en estudios longitudinales, es decir, medidos a lo largo del tiempo(Ballesteros Espinoza, 2022) .

Este enfoque examina la configuración de cada conjunto de datos y utiliza esa información para calcular un grupo de pesos óptimos. Estos pesos se emplean para crear una representación común de las observaciones, conocida como compromiso. Los pesos se eligen de manera que el compromiso sea lo más representativo posible de todas las tablas. Luego, se realiza un Análisis de Componentes Principales (ACP) del compromiso para determinar la posición de las observaciones en un espacio compartido. Cada tabla de datos tiene sus propias posiciones representadas como puntos en este espacio, y los datos pueden visualizarse como puntos en un espacio de múltiples

dimensiones (Rodríguez Martínez, 2020). A continuación, se representa el esquema general:

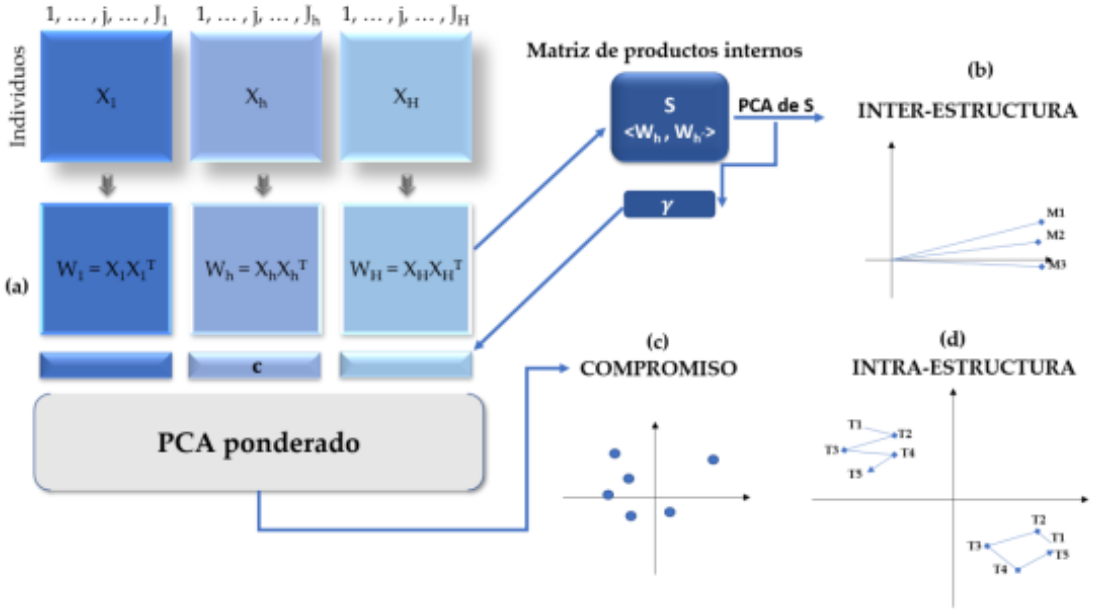


Ilustración 4: Visión general del método STATIS/ Fuente: (Rodríguez Martínez, 2020)

4.2.1.1. Notación

Dentro de los métodos STATIS, las observaciones se organizan en H conjuntos distintos, cada uno nombrado como "escenario". Un escenario se representa con una matriz $I \, x \, J_h$, que es una matriz rectangular referida como Y_h. En esta matriz, I representa el conjunto de observaciones y J_h es el conjunto de variables observadas en los sujetos para el estudio o escenario h-ésimo (Rodríguez Martínez, 2020).

Cada una de estas matrices usualmente pasa por un proceso de pretratamiento, que puede incluir centrado, normalización u otros procedimientos. La matriz resultante de este pretratamiento se denota como X_h. Las H matrices de datos X_h, cada una de dimensión $I \, x \, J_h$, se concatenan para formar una matriz completa X con tamaño I x J:

$$\boldsymbol{X} = [\boldsymbol{X1}| \cdots |\boldsymbol{Xh}| \cdots |\boldsymbol{XH}],$$

que es la utilizada para el análisis.

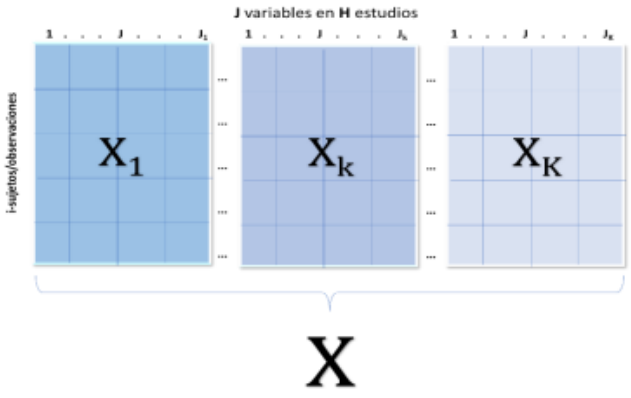


Ilustración 5: Matriz X resultante de la concatenación de las H matrices X_h. /Fuente:

Posteriormente, las matrices X_h se convierten en matrices de productos cruzados de tamaño I x I, representadas de la siguiente manera:

$$W_h = X_h X_h^T$$

Esta matriz W_h de un estudio indica cómo se relacionan entre sí las observaciones dentro de ese escenario.

Cuando presentamos la matriz de datos X en forma de bloques, podemos expresar su producto cruzado de la siguiente forma:

$$XX^T = [\mathrm{X_1}| \cdots |\mathrm{X_k}| \cdots |\mathrm{X_h}] \text{ x } [\mathrm{X_1}| \cdots |\mathrm{X_k}| \cdots |\mathrm{X_h}]^{\mathrm{T}} = \sum_h \mathrm{X_h}\, X_h^T = \sum_h \mathrm{W_h}$$

Como se ha comentado previamente, cada estudio se caracteriza por un conjunto de datos ($\mathrm{X_h}$), una métrica en el espacio de los sujetos ($\mathrm{M_h}$) y una métrica en el espacio de las variables ($\mathrm{D_h}$). Usualmente, estas métricas son representadas por la matriz identidad. Esta configuración se denota como $\mathrm{W_h}$ y describe cómo se relacionan entre sí las observaciones dentro de cada estudio (Rodríguez Martínez, 2020).

4.2.1.2. Análisis de la Inter-Estructura

En el método STATIS, el primer paso implica examinar la similitud entre las estructuras de todas las tablas de datos, lo que se conoce como estudio de la inter-estructura (Rodríguez Martínez, 2020).

Para evaluar esta similitud, primero se transforman las matrices de datos en matrices de productos cruzados. La similitud entre dos matrices, por ejemplo, X_h y X_h', se representa como $Sh,\mathrm{h'}$.

Esta similitud se calcula mediante el producto interno de Hilbert-Schmidt entre las matrices de productos cruzados W_h y W_h' que se expresa como,

$$\langle W_h | W_h' \rangle_{HS}$$

y cuyos coeficientes Sh,h son obtenidos de la siguiente manera:

$$Sh,\mathrm{h'} = \langle W_h | W_h' \rangle_{HS} = tr[X_h X_h^T, xX_{h'} X_h^T,] = \mathrm{tr}[W_h W_h'] =$$

$$= vec(W_h)^T x\ vec(W_{h'}) = \sum_i^I \sum_j^I W_{i,j,h}\, W_{i,j,h'}$$

Este producto interno se puede interpretar como el producto escalar entre matrices semi-definidas positivas $W_h W_h'$, lo que significa que siempre es mayor o igual a cero. Además, se emplea para establecer la norma de una matriz, la cual se define de la siguiente manera:

$$||W_h||^2 = \langle W_h, W_h \rangle$$

Es fundamental tener en cuenta que cuando las matrices W_h y W_h' de forma que el sumatorio de los cuadrados de sus componentes sea uno, el producto interno resultante entre estas matrices representa el coseno del ángulo que las separa. Este valor, conocido como coeficiente RV, fue creado para medir la similitud entre matrices simétricas cuadradas, particularmente en el contexto de matrices semi-definidas positivas.

Los productos internos entre todas las matrices se recopilan en una matriz de productos internos, denotada como S_{HXH}. Además, esta matriz puede ser obtenida mediante la matriz Z_{HxI^2} expresada de la siguiente manera,

$$Z = [vec(W_1)|\dots|vec(W_h)|\dots|vec(W_H)]^T$$

A partir de la matriz Z, es posible determinar S de la siguiente manera,

$$S = ZZ^T$$

Por lo tanto, S es una matriz semi-definida positiva, lo que implica que todos sus valores propios son no negativos y sus vectores propios son reales y ortogonales. Esto permite que S pueda ser descompuesta en valores propios de la siguiente manera:

$$S = V\tau V^T$$

donde,

$$V^T V = I$$

La descomposición en valores propios de esta matriz S ofrece el Análisis de Componentes Principales (ACP) de la estructura de similitud entre las tablas, que se puede visualizar como puntos en un gráfico de ACP.

Además de ofrecer visualmente la similitud, esta descomposición proporciona pesos óptimos para combinar las tablas en un "compromiso". La separación de los autovalores propios de la matriz S siempre produce un primer vector propio cuyos elementos tienen el mismo signo, usualmente considerados positivos. Esta característica se explica por el teorema de Perron-Frobenius, el cual establece que las matrices semi-definidas positivas, con todos sus elementos positivos, poseen un primer vector propio cuyos componentes tienen el signo idéntico. Estos pesos se pueden utilizar para ponderar las tablas y dar más importancia a aquellas que representan mejor el conjunto de datos. Los pesos se pueden almacenar en un vector o en una matriz diagonal, dependiendo de la preferencia del análisis.

4.2.1.3. Análisis del compromiso

El análisis de la inter-estructura revela cuán similares son las diferentes tablas. Cuando se encuentran semejanzas significativas entre ellas, resulta útil crear una matriz de compromiso o consenso C para representar los estudios. Esta matriz tiene que sintetizar de manera efectiva la información de todas las configuraciones (González Bartolomé, 2015).

Los pesos obtenidos se emplean con el objetivo de realizar la factorización generalizada de los valores singulares de la matriz X, siguiendo las limitaciones impuestas por la matriz de masas M. Esta matriz M se construye a partir de las masas m_i asignadas a cada observación en X, formando así un vector de masas m. Todas las masas en m son positivas o nulas y su suma total es igual a uno (Rodríguez Martínez, 2020):

$$M=\text{diag}(m).$$

La descomposición generalizada de valores singulares se representa como:

$$X = G\Theta P^T,$$

donde *G* incluye los vectores singulares izquierdos y *P* los vectores singulares derechos. La ecuación $G^T MG = P^T CP =$ I indica que esta descomposición genera puntajes de factores para las observaciones y cargas factoriales para las variables. En términos de análisis de componentes principales (ACP), la ecuación puede simplificarse a

$$\mathrm{X} = \mathrm{FP}^{\mathrm{T}},$$

donde $F = G\Theta$ contiene los puntajes factoriales.

Dado que la matriz X se compone de H tablas, cada una con J_h variables, la matriz P puede dividirse de forma similar a X. Así, la matriz P se expresa como

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_h \\ \vdots \\ P_H \end{bmatrix} = [P_1^T | \cdots | P_h^T | \cdots | P_H^T]^T$$

donde cada P_h es una matriz de dimensiones $J_h \times RJ_h \times R$, siendo R el rango de X. De esta manera, la ecuación $X = G\Theta P^T$ puede reescribirse para reflejar esta estructura en bloques, de la siguiente forma:

$$\begin{aligned} X &= (X_1 | \ldots | X_h | \ldots | X_H) = G\Theta P^T \\ &= G\Theta[(P_1^T | \cdots | P_h^T | \cdots | P_H^T]^T]^T \\ &= G\Theta(P_1^T | \cdots | P_h^T | \cdots | P_H^T) \\ &= (G\Theta P_1^T \cdots G\Theta P_h^T \ldots G\Theta P_H^T) \end{aligned}$$

El puntaje factorial de X se determina como $F = G\Theta$, representando el compromiso óptimo para el conjunto de las H matrices. Estos puntajes permiten graficar las observaciones, y la varianza de dichos puntajes se calcula utilizando las masas contenidas en la matriz M.

En resumen, la ecuación $F=XCP$ indica que los puntajes factoriales se derivan de X al combinar las ecuaciones de descomposición. Los puntajes factoriales parciales correspondientes a una tabla específica se obtienen al proyectar esa tabla en sus vectores

singulares P_h. La matriz de puntajes factoriales de compromiso se calcula como la media ponderada de los puntajes de factores parciales.

Finalmente, los biplots resultan útiles para análizar la estructura estadística de cada bloque, aunque las posiciones relativas de los puntajes factoriales y las cargas no pueden interpretarse de manera directa. Una representación alternativa traza las correlaciones entre las variables originales de X y las puntuaciones factoriales en un gráfico bidimensional con un círculo de correlación. Mientras más cerca se encuentren las variables del círculo, mejor serán explicadas por los componentes empleados para generar la gráfica.

Para calcular la configuración de compromiso, se utiliza un promedio ponderado de las tablas originales, asignando a cada tabla un peso que corresponde a su contribución al primer vector propio de la inter-estructura (González Bartolomé, 2015). De esta manera,

$$W = \sum_{k=1}^{K} \propto_k W_k$$

los coeficientes $(\propto_{1,\dots,} \propto_k)$ que generan la combinación lineal más correlacionada con cada ocasión individual se encuentran al ajustar los coeficientes del primer vector eigenvector de C. Destaca que los objetos de compromiso comparten la misma naturaleza y pueden ser integrados en el mismo contexto que los objetos individuales.

4.2.1.4. Análisis de la Intra-Estructura

La investigación sobre la intra-estructura se enfoca en examinar la composición interna de cada matriz y cómo se vincula con el compromiso. Hay diversas formas de representar este análisis, como los sujetos dentro de cada matriz, las variables y sus trayectorias. En este estudio, dado su interés y relevancia, pondremos especial atención en las trayectorias de los I individuos a lo largo de las T matrices originales (González Bartolomé, 2015).

Cada individuo proyectado en todas las matrices (tiempos) resulta en una serie de puntos llamada trayectoria. Al situar estos IxT puntos en el espacio de los componentes principales de la matriz de compromiso C, es posible seguir cómo evoluciona cada individuo con el paso del tiempo. Además, el lugar que ocupa cada individuo en la matriz de compromiso será el centro de masa de su trayectoria (González Bartolomé, 2015).

4.2.1.5. Nuevas contribuciones al Método STATIS

Los avances recientes en los enfoques de la familia STATIS se pueden categorizar de la siguiente manera según Vicente-Galindo (2013): en función de los datos iniciales utilizados, de la asignación de pesos al construir la matriz de consenso, de la consideración de información externa, y de la disponibilidad de pares de tablas en diferentes contextos o momentos temporales. Esto implica que hay tres posibilidades en cuanto a los datos iniciales (Rodríguez Martínez, 2020):

- En situaciones donde todas las tablas de datos contienen información sobre las mismas variables observadas en los mismos individuos, pero en contextos distintos, se recurre al X-STATIS o Análisis Parcial Triádico (PTA). Dicho método opera tanto con las matrices de datos como con operadores.

Por otro lado, al trabajar con las mismas variables y sujetos, utilizando una matriz de operadores como punto de partida, se emplean técnicas como COVSTATIS y DISTA.

- En situaciones donde los individuos están estructurados en grupos, se utiliza la técnica CANOSTATIS.
  Por otro lado, en una versión de STATIS, los pesos asignados para crear la matriz consenso difieren de la idea inicial de L'Hermier. Este enfoque se conoce como Power-STATIS, donde el peso asignado a cada matriz está vinculado a las componentes del primer vector propio de la matriz de correlaciones vectoriales.

Además, hay otra clasificación relacionada con la incorporación de información externa, la cual se puede presentar de dos maneras (Rodríguez Martínez, 2020):

- Una de ellas es al incorporar las H matrices utilizadas en el STATIS, junto con una matriz adicional que contiene información externa sobre los individuos. En estos casos, proponen el t+1 STATIS. Posteriormente, se introduce el STATIS-4, que es una extensión del t+1 STATIS.
- Por otro lado, cuando los individuos están estructurados en grupos, se utiliza CANOSTATIS.

En otra situación, se presenta el caso en el que se cuentan con pares de tablas en H momentos temporales distintos. Entre las técnicas relacionadas con este escenario, encontramos el STATICO, que combina COINERCIA y STATIS, así como el COSTATIS, que primero emplea STATIS y después COINERCIA.

4.4. Software para STATIS

Actualmente, hay numerosos programas informáticos diseñados para llevar a cabo el análisis de múltiples tablas. No obstante, en este trabajo, nos centraremos en el paquete MultBiplotR, elaborado en la Universidad de Salamanca, así como el paquete ade4 disponibles en el lenguaje RStudio para implementar el método STATIS.

El paquete Ade-4 ofrece una variedad de métodos para el análisis de datos multivariados. Entre estos métodos se incluyen análisis de una sola tabla, como el análisis de componentes principales, así como métodos de acoplamiento de dos tablas, como el análisis de correspondencia canónica, análisis de redundancia y análisis de coinercia. Además, Ade-4 proporciona herramientas para el análisis de tablas de tres vías, conocidas como tablas K, utilizando métodos desarrollados principalmente en la escuela francesa, tales como STATIS, X-STATIS, y STATICO, entre otros. Desde 2002, Ade-4 está disponible como un paquete de software R compatible con sistemas operativos Windows, Mac OS, Linux y Unix, facilitando el análisis multivariado a través de diversas herramientas gráficas y analíticas (Rodríguez Martínez, 2020).

Por otro lado, el paquete MultBiplotR permite realizar análisis multivariados clásicos a través de Biplots de PCA con características avanzadas, como transformaciones de datos no convencionales y escalas ajustables para las variables. Además, ofrece diversas ayudas visuales, tales como variaciones en tamaño y color de los puntos basadas en sus cualidades de representación y capacidad predictiva. También incluye procedimientos de Alternating Least Squares (ALS) o Criss-Cross para calcular aproximaciones de

rango reducido, manejando datos incompletos y aplicando pesos diferenciales a cada elemento de la matriz de datos, e incluso versiones robustas de estos procedimientos (Vicente-Villardón et al., 2023).

Este paquete forma parte de un proyecto más amplio denominado MULTBIPLOT, que abarca múltiples técnicas de biplot. MultBiplotR es una adaptación en R del paquete original MULBIPLOT desarrollado en MATLAB. Actualmente, se está desarrollando una interfaz gráfica de usuario (GUI) para facilitar su uso (Vicente-Villardón et al., 2023).

4.5. Biplot

Un BIPLOT es una herramienta visual utilizada para representar datos multivariados. Similar a cómo un gráfico de dispersión muestra la relación entre dos variables, un BIPLOT permite visualizar la interacción de tres o más variables simultáneamente.

El BIPLOT reduce la dimensionalidad de un conjunto de datos multivariantes, usualmente representándolos en dos dimensiones. Esta representación gráfica incluye tanto los puntos de la muestra como las variables medidas. Las variables se muestran como vectores, que apuntan en las direcciones que mejor reflejan las variaciones específicas de cada una.

Nos podemos encontrar diferentes tipos de Biplots cada uno con unas determinadas características(Nieto Librero, 2022):

- GH-Biplot: En el GH-Biplot, los productos escalares de las columnas de una matriz coinciden con los productos escalares de los marcadores columna. La distancia de Mahalanobis entre las filas de la matriz se estima utilizando la distancia euclidiana entre los marcadores fila. Esto asegura una buena representación de las variables en el biplot.

- JK-Biplot: El JK-Biplot se distingue por asegurar que los productos escalares de las filas de una matriz coincidan con los productos escalares de los marcadores fila. Además, la distancia entre los marcadores fila refleja la distancia entre las filas de la matriz original. Los marcadores filan también coinciden con las coordenadas de las filas en un Análisis de Componentes Principales (ACP). Por otro lado, las coordenadas para las columnas se obtienen mediante las proyecciones de los ejes originales en el espacio de las Componentes Principales (CPs). Estas características garantizan una representación adecuada de las filas en el biplot.

- El HJ-Biplot se caracteriza porque los marcadores correspondientes a las filas se alinean con las coordenadas de los individuos en el espacio de representación. Asimismo, los marcadores de las columnas coinciden con las coordenadas de las variables. Este enfoque garantiza una óptima calidad de representación tanto para las filas como para las columnas en el biplot.

Desde la perspectiva del usuario, los biplots resultan valiosos debido a que su interpretación se fundamenta en conceptos geométricos básicos, familiares para quienes poseen una formación matemática (Vicente Villardón, n.d.) :

- En un biplot, la similitud entre los individuos se refleja inversamente en la distancia que los separa en la representación gráfica: cuanto más cercanos estén dos puntos, mayor será su similitud.

- En los JK-Biplot y HJ-Biplot, las magnitudes y direcciones de los vectores que representan las variables se comprenden en relación con la variabilidad y covariabilidad, respectivamente.
- La asociación entre individuos y variables se analiza a través del producto escalar, es decir, considerando las proyecciones de los puntos individuales sobre los vectores de las variables.
- En el HJ-Biplot, la búsqueda de variables diferenciadoras está estrechamente ligada a la relación con los ejes factoriales. Este enfoque permite identificar qué variables tienen una mayor influencia en la distribución de los individuos en el espacio representado por los ejes factoriales del biplot.

La medida que evalúa las relaciones entre los ejes del Biplot y las variables observadas se llama contribución relativa del factor al elemento. Esta medida indica la proporción de variabilidad de cada variable explicada por cada factor, y puede entenderse como el coeficiente de determinación en regresión. Específicamente, si los datos están centrados, representa el coeficiente de determinación en la regresión de cada variable sobre su respectivo eje. Las contribuciones relativas son valiosas para identificar qué variables tienen una mayor asociación con cada eje, lo que permite comprender mejor qué variables influyen más en la posición de los individuos en cada eje del biplot.

4.5.1. STATIS-BIPLOT

Es sabido que realizar un análisis STATIS equivale a efectuar un Análisis de Componentes Principales de la matriz :

$$X_S = [\sqrt{\alpha_1}\, X_1, \ldots, \sqrt{\alpha_K}\, X_K]$$

Ahora, consideremos que la Descomposición en Valores Singulares (SVD):

$$X_S = U\Delta^{1/2}P^T$$

Aquí, U contiene los vectores propios de $X_S X_S^T$, P incluye los vectores propios de $X_S^T X_S$ y $\Delta^{1/2}$ representa las raíces cuadradas de los valores propios no nulo de ambas matrices, que son idénticos.

La SVD, que está estrechamente vinculada al PCA, permite generar un Biplot de XS, que se expresa como:

$$X_S = \Delta P^T$$

usando como coordenadas para los individuos las primeras columnas de $A = U\Delta^{1/2}$, que constituye el compromiso STATIS.

Los individuos tienen las mismas coordenadas principales que las coordenadas de consenso obtenidas anteriormente. Esto se debe a que:

$$X_S X_S^T = \sum_k \propto_k X_k X_k^T = \sum_k \propto_k W_k = W$$

Para obtener un Biplot de $X = [X_1, \ldots, X_K]$, se debe considerar que $X_S = XD_{\propto}^{1/2}$ donde $D_{\propto}$ es una matriz diagonal por bloques con la forma $diag\ (D_{\propto 1}, \ldots, D_{\propto k})$, y cada bloque $D_{\propto k}$ es igual a $\propto_k I_{Jk}$.

Un Biplot de la matriz de datos observada $X = AB^T$ se puede generar fácilmente utilizando $B = D_{\propto}^{-1/2}$V. De este modo, se obtiene un Biplot en el que las coordenadas de las filas son las de consenso obtenidas mediante el método STATIS

La matriz B puede ser segmentada en tantos bloques como ocasiones haya:

$$B = [B_1, \ldots, B_K]$$

Esto permite crear un Biplot separado para cada ocasión $X_k = AB_k^T$ utilizando el método STATIS. En lugar de maximizar la variabilidad total de cada ocasión por separado, se enfoca en la estructura común a todas las ocasiones. Este biplot de predicción tiene una calidad de ajuste inferior comparado con el biplot separado, pero esta también puede ser evaluada.

Cada matriz X_k se puede aproximar de la siguiente manera:

$$X_k = \widetilde{X_k} + E = AB_k^T + E$$

Donde $\widetilde{X_k}$ representa los valores ajustados. Las dimensiones de las matrices de coordenadas se han omitido para mayor simplicidad.

La métrica de variabilidad de la ocasión en el nuevo biplot se obtiene al comparar la traza de la matriz de valores ajustados con la traza de la matriz original:

$$p_k^2 = \frac{traza(\widetilde{X_k}\widetilde{X_k^T})}{traza(X_k X_k^T)}$$

Para evaluar la calidad de la representación de cada columna en la solución, se calcula el cociente entre la norma al cuadrado de la columna correspondiente de los valores ajustados y la norma al cuadrado de la misma columna en los datos originales:

$$p_k^2(j) = \frac{||\widetilde{x_k(j)}||^2}{||X_k(j)||^2}$$

Este indicador también puede interpretarse como el cuadrado del coseno del ángulo entre el vector que representa la variable en el espacio J-dimensional completo y su proyección en el subespacio de dimensiones reducidas.

4.6.Análisis de Correlaciones

Dadas dos variables aleatorias X e Y con varianza estrictamente positiva, el coeficiente de correlación entre estas dos variables se define mediante la fórmula presentada (Kenney & Keeping, 1951),

$$P_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}Var(Y)}$$

La matriz de correlación se define como una matriz cuadrada y simétrica donde los elementos de la diagonal son unos, y los otros elementos representan los coeficientes de correlación entre las variables. Esta matriz se denota de la siguiente manera:

$$R = \begin{pmatrix} 1 & p_{12} & \cdots & p_{1k} \\ p_{21} & 1 & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \ldots & 1 \end{pmatrix},$$

Si las variables son estandarizadas como se indica en la ecuación:

$$X_i^* = \frac{X_i - \mu_i}{\sigma_i},$$

para los valores de i=1, 2,…,k se puede verificar que la matriz de covarianza S coincide con la matriz de correlaciones R de las variables sin estandarizar. Dado que $Var\,(X_i^*)$=1 y se cumple lo establecido en:

$$Cov\left(X_i^*, X_j^*\right) = E\left(X_i^* X_j^*\right) - E(X_i^*)E\left(X_j^*\right)$$

$$= E\left(\left(\frac{X_i - \mu_i}{\sigma_i}\right)\left(\frac{X_i - \mu_j}{\sigma_j}\right)\right)$$

$$= \frac{1}{\sigma_i \sigma_j} E\left(X_i X_j - X_i \mu_j - X_j \mu_i + \mu_i \mu_j\right)$$

$$= \frac{1}{\sigma_i \sigma_j}\left(E\left(X_i X_j\right) - E\left(X_i \mu_j\right) - E\left(X_j \mu_i\right) + E\left(\mu_i \mu_j\right)\right)$$

$$= \frac{1}{\sigma_i \sigma_j}\left(E\left(X_i X_j\right) - \mu_i \mu_j - \mu_j \mu_i + \mu_i \mu_j\right)$$

$$= \frac{1}{\sigma_i \sigma_j}\left(E(X_i X_j) - \mu_i \mu_j\right) = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = p_{ij}$$

por lo tanto, se puede expresar como sigue:

$$S^* = \begin{pmatrix} 1 & \sigma_{12}^* & \dots & \sigma_{1k}^* \\ \sigma_{21}^* & 1 & \dots & \sigma_{2k}^* \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1}^* & \sigma_{k2}^* & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & p_{12} & \dots & p_{1k} \\ p_{21} & 1 & \dots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \dots & 1 \end{pmatrix} = R$$

Esto lleva a una conclusión directa de que la matriz de correlación es positiva-semidefinida. Esta afirmación se deriva de la consistencia entre la matriz de covarianza y S^*, lo cual implica que R es también positiva-semidefinida.

Esta matriz ofrece una visión clara de cómo se relacionan todas las combinaciones de valores de una tabla. En cada celda se muestra el coeficiente de correlación entre cada par de variables. Este coeficiente puede variar entre -1 y +1: indicando una correlación negativa perfecta si el valor es de -1, indicando una correlación positiva perfecta si es +1, y sin correlación si el valor es de 0 (Ortega, n.d.).

4.7.Análisis de Clúster

El análisis de Clusters, también conocido como Análisis de Conglomerados, es un método de Análisis de datos Exploratorio que resuelve problemas de clasificación. Su propósito es agrupar objetos (variables, personas, animales…) en conjuntos (conglomerados o clusters) de tal manera que la similitud entre los miembros del mismo clúster sea mayor que la relación entre los miembros de diferentes clusters. Cada conjunto se describe como la clase a la que pertenecen sus miembros (Vicente Villardón, 2007).

El análisis de conglomerados es una técnica que puede revelar asociaciones y relaciones en los datos que pueden no ser evidentes inicialmente, pero que pueden resultar valiosas una vez que se han identificado. Los resultados del análisis pueden ayudar a definir formalmente un esquema de clasificación, como una taxonomía de un conjunto de objetos, una propuesta de modelo estadístico para describir una población, la asignación de nuevos individuos a categorías con fines de diagnóstico e identificación…(Vicente Villardón, 2007).

Existen dos tipos fundamentales de métodos de clasificación: jerárquicos y no jerárquicos. Principalmente, estos dos tipos se distinguen en que en los métodos jerárquicos la clasificación obtenida presenta un número creciente de clases anidadas, mientras que en los no jerárquicos no.

4.7.1. K-means

El método de k-medias se emplea para agrupar o predecir datos a través de un proceso de aprendizaje no supervisado. El usuario determina el número de conglomerados, denotado como $\boldsymbol{k}$. Cada una de las agrupaciones cuenta con un vector medio, llamado centroide. Para determinar el clúster más cercano a un objeto, se utiliza la distancia euclidiana, la cual se lleva a cabo de manera iterativa, asignando cada objecto al clúster con el centroide más cercano (Vintimilla et al., 2017).

El método tiene tres etapas(Gonzalo, 2019):

a. Inicialización: después de seleccionar $\boldsymbol{k}$ (el número de grupos), se fijan los centroides en el espacio de los datos. Esto puede lograrse asignando aleatoriamente $\boldsymbol{k}$ puntos como centroides.
b. Asignación de las observaciones a los centroides: utilizando una media de distancia específica se asigna cada observación al centroide más próximo.
c. Actualización de los centroides: calculando la media de la ubicación de las observaciones en ese grupo actualizamos la posición de los centroides.

Los pasos 2 y 3 vuelven a ejecutarse todas las veces que sea necesario hasta que los centroides permanezcan constantes o estén por debajo de la distancia umbral establecida (Gonzalo, 2019).

Las n observaciones se pueden expresar como un vector en un espacio de $\boldsymbol{d}$ dimensiones, donde $\boldsymbol{d}$ es igual al número de variables o características que representan cada observación. Este algoritmo para formar $\boldsymbol{k}$ grupos, de modo que minimiza la suma de distancias entre las observaciones y sus centroides respectivos en cada grupo $\boldsymbol{S} = \{\boldsymbol{S_1}, \boldsymbol{S_2}, \ldots, \boldsymbol{S_K}\}$.

$$\boldsymbol{min\ E\ (\mu_i) = min} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \sum_{\boldsymbol{x_j \epsilon S_i}} \boldsymbol{\left\| x_j - \mu_i \right\|^2}$$

El conjunto de datos $\boldsymbol{S}$ está compuesto por las observaciones $\boldsymbol{x_j}$, representadas como vectores, donde cada elemento del vector denota una característica o atributo. Estos datos se dividen en $\boldsymbol{k}$ grupos o clústeres, cada uno con su respectivo centroide $\boldsymbol{\mu_i}$.

En cada iteración de la actualización de los centroides se aplica la condición de que la función $\boldsymbol{E(\mu_i)}$ alcance un extremo. En el caso de la función cuadrática, esta condición se expresa como:

$$\boldsymbol{\frac{\partial E}{\partial \mu_i} = 0 \quad \mu_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \epsilon S_i^{(t)}} x_j}$$

Luego, se determina cada nuevo centroide como la media de los elementos en cada grupo. Si se quiere emplear el algoritmo, es esencial establecer el número de grupos ($\boldsymbol{k}$), la métrica de distancia a utilizar y cómo se inicializan los centroides. En términos generales, el algoritmo tiende a converger hacia un mínimo local en lugar de un mínimo global.

Selección de la métrica de distancia en k-means

Las diferentes métricas de distancia en el algoritmo k-means ofrecen diversas formas de medir la separación entre los puntos en un espacio multidimensional. Cada métrica tiene sus propias características y aplicaciones, por lo que seleccionar la métrica adecuada es crucial en problemas de agrupamiento, ya que alterar la medida de similitud entre elementos afectará al cálculo de los clústeres (Gonzalo, 2019).

A continuación, se presentan las distancias más usadas:

a. Distancia euclídea

$$\boldsymbol{d_{euc}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}$$

b. Distancia Manhattan

$$\boldsymbol{d_{man}(x,y) = \sum_{i=1}^{n}|(x_i - y_i)|}$$

c. Distancia de correlación de Pearson

$$\boldsymbol{d_{cor}(x,y) = 1 - \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}}$$

d. Distancia de correlación de coseno

$$\boldsymbol{d_{eisen}(x,y) = 1 - \frac{|\sum_{i=1}^{n} x_i y_i|}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}}$$

Por lo general, se emplea la distancia euclídea como medida por defecto. No obstante, según las características de los datos y las interrogantes que se busquen responder, puede ser más adecuado optar por otras métricas de distancia (Gonzalo, 2019).

Elección de k, el número de grupos a buscar

- **Método del codo**

Se busca llevar a cabo una prueba que contraste la mejora al seleccionar $\boldsymbol{k+1}$ grupos en comparación con k, evaluando la distancia intra-cluster obtenida al variar el número de grupos. Para ello, se ejecuta el algoritmo de k-medias con distintos valores de k, se calcula la distancia intra-cluster en cada caso y se representa ambos conjuntos de datos en un gráfico. En dicho gráfico se seleccionará el punto donde la curva parece doblarse en un codo (Perucha Jurjo, 2022).

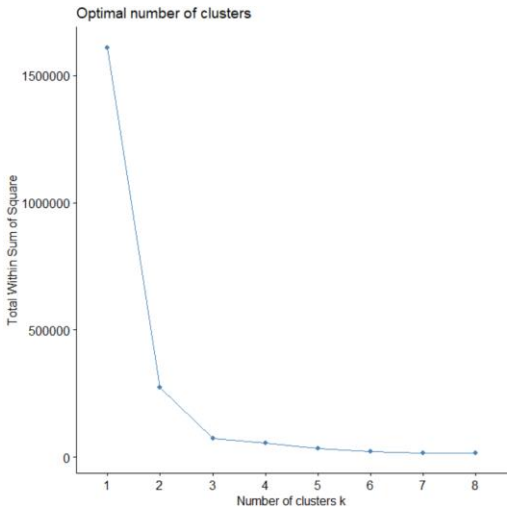


Ilustración 6: Método del Codo /Fuente: Elaboración propia con RStudio

- **Método de la Silueta**

Dicho método asigna a cada punto del conjunto de datos una medida de similitud entre él y el resto de los individuos dentro del mismo clúster, en contraste con aquellos clústeres diferentes. Cada punto recibe un valor dentro del rango de -1 a 1, conocido como índice de silueta. Este índice refleja la adecuación del punto al clúster al que se asigna y su predisposición a no pertenecer a otros grupos, es decir, un valor alto indica una mejor correspondencia con el clúster al que debe pertenecer y una menor probabilidad de estar en otro grupo. El índice de silueta promedio de todos los puntos se utilizará para indicar la calidad del agrupamiento (Perucha Jurjo, 2022).

Para un conjunto de datos $\{\boldsymbol{x_1}, \dots, \boldsymbol{x_n}\}$, al considerar un punto $\boldsymbol{x_i}$, si definimos $\boldsymbol{a_i}$ como la media de las distancias entre $\boldsymbol{x_i}$ y los otros puntos dentro del mismo clúster, y $\boldsymbol{b_i}$ como la media de las distancias entre $\boldsymbol{x_i}$ y los puntos de otros clústeres, el índice de silueta de $\boldsymbol{x_i}$ se determina de la siguiente manera, con k representando el número de clústeres:

$$\boldsymbol{s}(\boldsymbol{x_i}, \mathbf{k}) = \frac{\boldsymbol{b_i} - \boldsymbol{a_i}}{\boldsymbol{máx}\{\boldsymbol{b_i}, \boldsymbol{a_i}\}}$$

Así, un valor cercano a 0 sugiere que la observación está próxima a un clúster vecino, un índice cercano a 1 indica que el punto está altamente integrado en su clúster y distante de otros grupos, mientras que un valor próximo a -1 señala que los clústeres no han sido asignados adecuadamente.
Al calcular el promedio de los índices de silueta de todos los puntos, obtenemos el índice de silueta para k grupos.

$$\boldsymbol{S}(\boldsymbol{k}) = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{s}(\boldsymbol{x_i}, \mathbf{k}) = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \frac{\boldsymbol{b_i} - \boldsymbol{a_i}}{\boldsymbol{máx}\{\boldsymbol{b_i}, \boldsymbol{a_i}\}}$$

Por lo tanto, la estrategia para seleccionar k será la siguiente:

1. Se ejecuta el algoritmo para diversos valores de k
2. Se calcula un índice de silueta $\boldsymbol{S}(\boldsymbol{k})$ para cada agrupación en k grupos.
3. Se elige como número óptimo de clústeres aquel que el índice de silueta sea el máximo

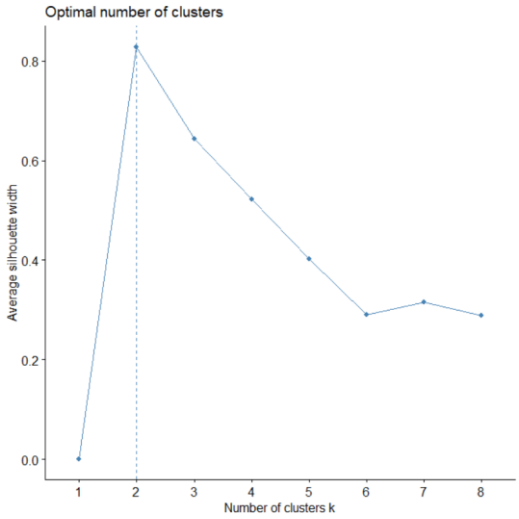


Ilustración 7: Método de Siluette /Fuente: Elaboración propia con RStudio

5. RESULTADOS

5.1.Análisis descriptivo

Una vez obtenida la base de datos final para nuestro estudio, procederemos a llevar a cabo un análisis descriptivo de las variables de interés por ámbito y por año.

Los boxplots realizados (Ilustración 8, Anexos: Ilustración 9,10 y 11) revelan una amplia gama de tendencias y patrones en la evolución de las variables relacionadas con los delitos de odio en diferentes ámbitos sociales.

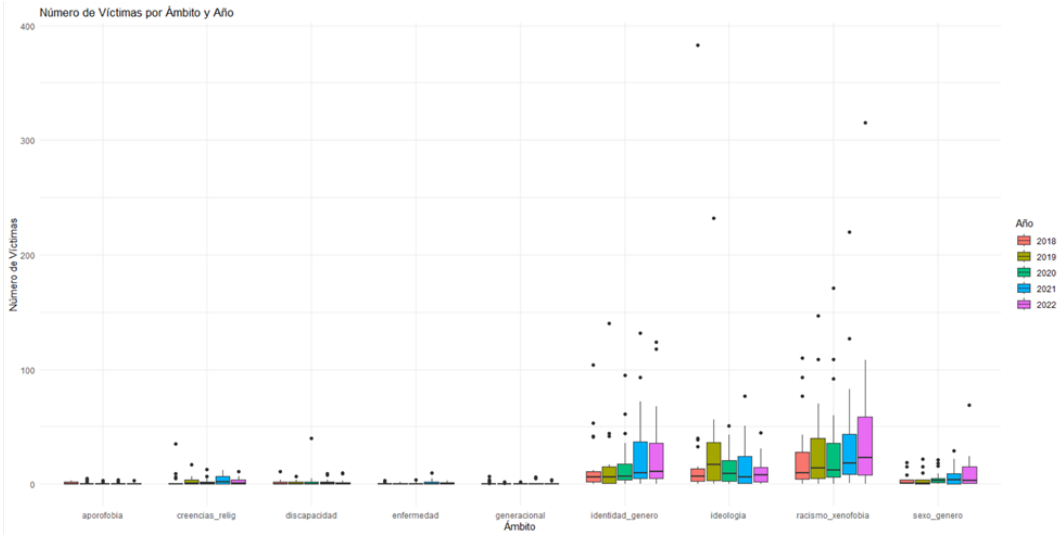


Ilustración 8: Boxplot Número de Victimizaciones por Ámbito y Año /Fuente: Elaboración propia con Rstudio

En general, hay una tendencia a la baja en la mayoría de las medidas de aporofobia, con valores medianos de cero que indican que muchos años no se registraron incidentes. Sin embargo, hay fluctuaciones en los valores máximos y las medias anuales. Por ejemplo, el número de detenciones (Anexos: Ilustración 9) alcanzó un máximo de 6 en algunos años, con una media de 1.11 en 2018, disminuyendo en

años posteriores. Los hechos conocidos y esclarecidos (Anexos: Ilustración 10 y 11) muestran patrones similares con un leve incremento hacia el final del periodo. Esto sugiere una prevalencia baja pero variable de incidentes de aporofobia.

En cuanto a las creencias religiosas, los datos muestran variabilidad considerable en valores máximos y medias anuales. Aunque la mediana es cero en la mayoría de los años, indicando pocos incidentes, los valores máximos señalan casos esporádicos pero severos, como 27 detenciones en 2022. Los hechos conocidos y esclarecidos también varían, con medias entre 1.21 y 3.84 para hechos conocidos, y entre 1.21 y 2.05 para hechos esclarecidos. El número de víctimas sigue un patrón similar, sugiriendo que los incidentes, aunque no frecuentes, pueden ser graves

En el caso de la discapacidad, el número de detenciones relacionadas con discapacidad presenta fluctuaciones notables, con medias anuales entre 0.789 y 1.68, y máximos que varían de 6 a 18. Los hechos conocidos y esclarecidos también muestran variaciones significativas, con medias desde 1.05 hasta 3.11 para hechos conocidos y desde 0.895 hasta 1.37 para hechos esclarecidos. El número de víctimas refleja incidentes graves y esporádicos, con medias que van desde 1.37 en 2018 hasta un máximo de 3.16 en 2020, indicando una severidad variable en los incidentes.

Para los delitos de odio contra personas con enfermedad, los datos de detenciones y hechos conocidos muestran un bajo número de incidentes anuales, con medias que oscilan entre 0 y 1.11, y máximos que llegan hasta 5 en 2021. Los hechos en esclarecidos y el número de víctimas también son bajos, con medias generalmente por debajo de 1, destacando solo algunos picos en 2021.

En el ámbito generacional, se observa una mayor variabilidad, especialmente en hechos conocidos con un máximo de 23 en 2021 y una media anual que llega hasta 1.89. Las detenciones y hechos en esclarecidos muestran menores incidentes con medias generalmente menores a 1, y un número de víctimas con una media de hasta 0.842 en 2021.

Por otro lado, para el ámbito de la identidad de género, se observa un aumento significativo en el número de detenciones y hechos conocidos a lo largo de los años. En 2022, las detenciones alcanzaron un máximo de 68, con una media anual de 11.7. Los hechos conocidos también muestran un incremento, llegando a 101 en 2021 y una media anual de 25.1. Los hechos esclarecidos y el número de víctimas reflejan tendencias similares, con picos de 66 y 124 respectivamente en 2022, y medias que van aumentando consistentemente cada año.

Para la ideología, los datos muestran altos números en 2018 y 2019, especialmente en hechos conocidos y el número de víctimas. En 2018, se registraron 350 hechos conocidos con una media anual de 31.4, y 383 víctimas con una media de 30.6. Estos números descienden en los años posteriores, con una caída notable en 2022 a 46 hechos conocidos y 45 víctimas, y medias anuales también en declive.

En lo que respecta al racismo y xenofobia, hay un patrón de aumento constante en todos los indicadores. Las detenciones aumentaron de 50 en 2018 a 174 en 2022, con medias anuales que pasan de 10.2 a 19. Los hechos conocidos y esclarecidos también muestran incrementos significativos, con máximos de 205 y 130 respectivamente en 2022. El número de víctimas alcanza su pico en 2022 con 315, reflejando una tendencia al alza en estos incidentes.

En cuanto al sexo y género muestra fluctuaciones más moderadas en comparación con los otros ámbitos. Las detenciones y los hechos conocidos tienen valores máximos más bajos, con 26 detenciones en 2018 y 64 hechos conocidos en 2022. Las medias anuales de estos indicadores también son relativamente bajas. Los hechos esclarecidos y el número de víctimas siguen patrones similares, con máximos de 21 y 69 respectivamente en 2022, y medias que se mantienen bajas a lo largo de los años.

5.2. Análisis de Correlaciones

Realizaremos un análisis de correlaciones entre las distintas variables de estudio, num_vict, num_deten, num_hec_con y num_hec_esc, que nos mostrará la fuerza y la dirección de la relación lineal entre cada par de variables a través de la matriz de correlaciones.

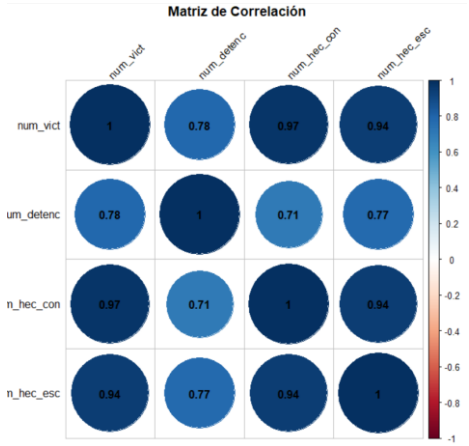


Ilustración 12: Matriz de Correlaciones de las variables de estudio /Fuente: Elaboración propia con Rstudio

Los resultados de la matriz de correlaciones (Ilustración 12) revelan relaciones significativas entre las variables de estudio. Se observa una correlación positiva fuerte entre el número de víctimas y el número de detenciones (0.78), lo que sugiere una relación estrecha entre el crimen y las acciones policiales. Además, se observa una asociación muy marcada entre el número de víctimas y los hechos conocidos por las autoridades, así como con los hechos esclarecidos (0.97 y 0.94 respectivamente), indicando una estrecha conexión entre la victimización y la actuación policial. Por otro lado, se identifica una correlación moderada entre el número de detenciones y los hechos conocidos, así como con los hechos esclarecidos 0.71 y 0.77, respectivamente, lo que implica una relación significativa aunque menos fuerte. Finalmente, entre los hechos conocidos y esclarecidos por las autoridades se encuentra una correlación muy alta de aproximadamente 0.94, subrayando una fuerte asociación entre ambos aspectos del proceso de investigación y resolución del delito.

5.3. Análisis STATIS

Una vez realizados los análisis descriptivos básicos procederemos a la realización del análisis STATIS. En nuestro caso trabajaremos con una tabla tres vías por ello, agruparemos nuestra base de datos de partida por año y por comunidades y

transformaremos los datos de nuevo en una lista, donde cada elemento de la lista corresponde a la matriz de datos para cada ocasión.

En primer lugar, al realizar el análisis STATIS sobre los datos de delitos de odio, se obtuvo la matriz de coeficientes de correlación vectorial (RV). Esta matriz refleja la similitud en la estructura de correlaciones entre las variables a lo largo de los diferentes años analizados.

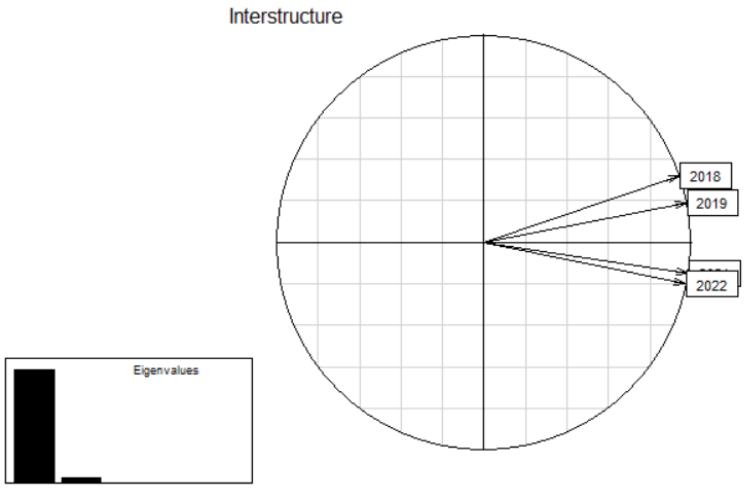


Ilustración 13 : Inter-estructura /Fuente: Elaboración propia con Rstudio

Como podemos observar (Ilustración 13), los valores de los coeficientes fuera de la diagonal principal son bastante altos, oscilando entre 0.85 y 1. Esto sugiere una correlación significativa entre las estructuras de los conjuntos de datos de diferentes años.

Por ejemplo, la estructura de correlaciones en 2018 está altamente correlacionada con la de 2019, con un coeficiente de 0.9787. Esto indica que, a pesar del paso de los años, las características subyacentes que relacionan los delitos de odio en 2018 y 2019 son muy similares. Del mismo modo, los años 2020 y 2021 muestran una correlación notablemente alta de 0.9972, reflejando una gran similitud en la estructura de los delitos de odio en estos dos años. Además, la correlación entre 2021 y 2022 también es extremadamente alta, alcanzando 0.9970, lo que sugiere una continuidad casi perfecta en las relaciones entre las variables durante estos años. La correlación entre 2020 y 2022 es de 0.9972, lo que subraya aún más la consistencia en los patrones de delitos de odio en estos años consecutivos.

Sin embargo, aunque la mayoría de los pares de años tienen correlaciones elevadas, se observan algunas variaciones que indican ligeros cambios en las relaciones entre las variables. Por ejemplo, la correlación entre 2018 y 2020 es ligeramente menor (0.8881) comparada con la de 2018 y 2019, lo que podría sugerir algún cambio en las características de los delitos de odio entre esos años. Los años 2018 y 2022 tienen una correlación más baja de 0.8506, y la correlación entre 2018 y 2021 es de 0.87. Estas variaciones pueden sugerir cambios sutiles en la estructura de las relaciones entre las variables en ciertos años, posiblemente debido a cambios sociales, legislativos....

En general, la alta consistencia en las correlaciones en la mayoría de los años indica una estabilidad en las características subyacentes de los delitos de odio en España durante el 2018 y 2022. No obstante, las variaciones detectadas en algunos años podrían reflejar factores externos como cambios en la política, variaciones en la denuncia de delitos, o efectos específicos de eventos como la pandemia de COVID-19. Es notable que la alta correlación entre los años más recientes (2020, 2021 y 2022) puede estar influenciada por el impacto del COVID-19, que podría haber homogenizado ciertos patrones de delitos de odio debido a las restricciones y cambios en la dinámica social.

Por consiguiente, obtenemos la tabla con los Cos^2, estos son indicadores de la calidad de la representación de los años en el compromiso global obtenido a través del análisis STATIS. Un valor alto de Cos^2 sugiere que el compromiso representa bien los datos de ese año específico. En general podemos observar que la mayoría de los valores de Cos^2 son muy altos, lo que indica una excelente representación de los datos de cada año en el compromiso global y que, además el análisis STATIS ha capturado de manera efectiva la estructura subyacente de los datos de delitos de odio en estos años.

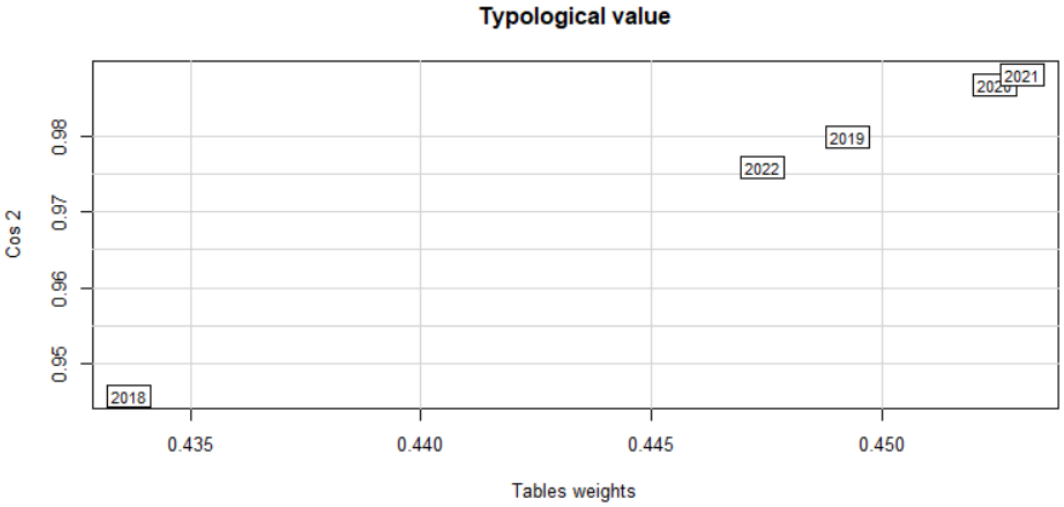


Ilustración 14: Representación de los pesos de los años con cos^2 /Fuente: Elaboración propia con Rstudio

Entre los años con cos^2 extremadamente altos nos encontramos a los años 2020 y 2021 con valores muy cercanos a 1, sugiriendo una estabilidad y coherencia significativas en los patrones de delitos de odio en estos periodos. Por otro lado, con un valor más inferior comparado con el resto de los años se encuentra el 2018. Aunque sigue siendo alto, con un 0.9394, no está tan bien representado en el compromiso global como los otros años.

Los pesos de las 5 tablas obtenidos en el análisis STATIS indican la contribución de cada año al compromiso global (Ilustración 14). Estos pesos reflejan la importancia relativa de cada año en la estructura total de los datos. En nuestro caso, los pesos asignados a cada año son relativamente similares, lo que sugiere que todos los años contribuyen de manera significativa a la estructura global de los datos. Sin embargo, hay ciertas diferencias poco significativas que pueden ser atribuibles a cambios específicos en los datos de delitos de odio. Los años 2020 y 2021 tienen una influencia ligeramente mayor, posiblemente a los efectos de la pandemia Covid-19. El año 2018, con el peso más bajo, podría haber tenido variaciones específicas

que lo hicieron menos representativo de la estructura general en comparación con otros años.

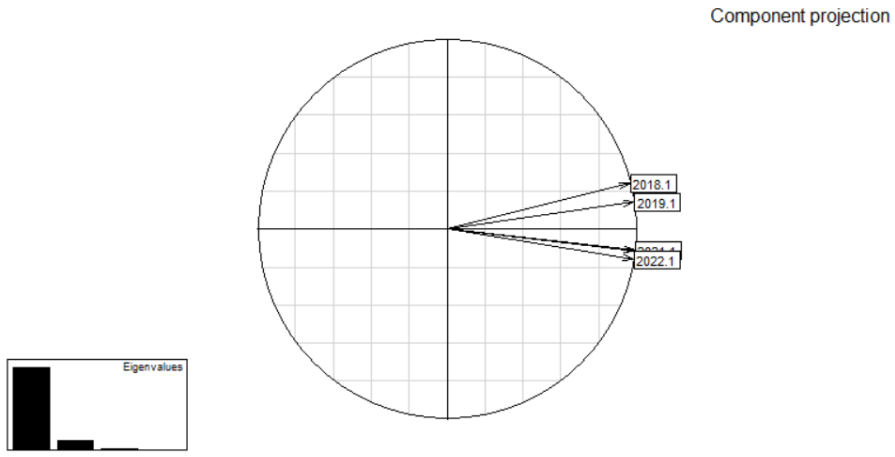


Ilustración 15: Proyección de los componentes /Fuente: Elaboración propia con Rstudio

Los valores propios obtenidos para la inter-estructura, indican la proporción de la variabilidad total en los datos que es capturada por cada uno de los componentes principales en la inter-estructura. La mayor parte de la varianza es explicada por los primeros dos componentes principales (99.74%), mientras que los componentes restantes contribuyen con porcentajes mucho menores. Esto hace indicar que las principales tendencias y patrones en los datos de delitos de odio pueden ser capturados efectivamente en un espacio bidimensional.

Por otro lado, los valores propios para el compromiso reflejan cuánta de la variabilidad en los datos puede ser capturada por cada componente en el compromiso. En nuestro caso, la combinación de estos dos componentes principales explica el 97.82% de la varianza total, lo que permite una representación precisa y efectiva.

Una vez estudiada la estructura de los datos según los años del estudio procederemos a estudiar la posición de las comunidades con respecto a los demás, estudiar los indicadores más importantes que diferencian a cada comunidad o grupos de comunidades y además, estudiar las modificaciones de las distintas comunidades a lo largo de los años.

A partir de los análisis realizados, se ha identificado una clara separación de los años en dos grupos distintos. El primer grupo comprende los años 2018 y 2019, mientras que el segundo grupo incluye los años 2020, 2021 y 2022. Esta agrupación se fundamenta en la inter-estructura y los diversos análisis llevados a cabo, que muestran diferencias significativas entre los dos conjuntos de años.

Para profundizar en estas diferencias, se separarán los años en dos grupos, diferenciando los periodos antes y durante la pandemia de Covid-19, lo que permitirá una comparación detallada permitiendo una comparación detallada de los mismos. A continuación, se presenta un Biplot que ilustra las comunidades autónomas compromiso y las proyecciones de las variables relevantes para cada uno de los grupos de estudio según los años, proporcionando una visualización clara de

cómo varían las características entre los dos periodos temporales y como se relacionan con las comunidades autónomas.

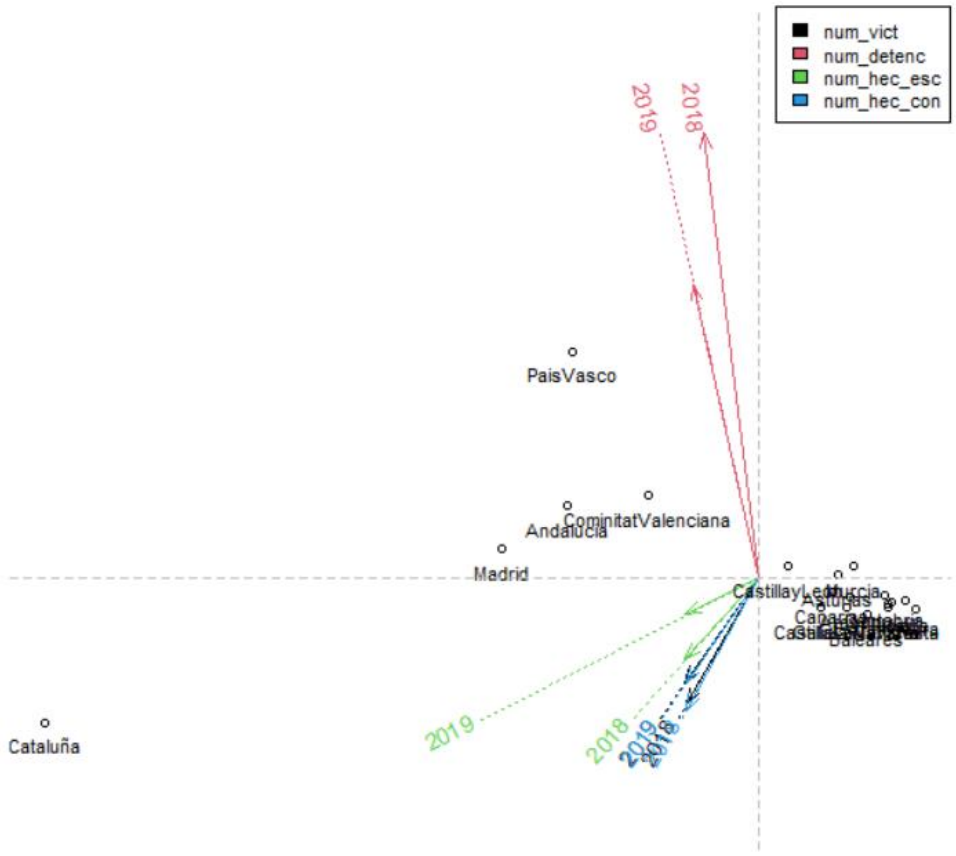


Ilustración 16: Biplot según las Comunidades Autónomas compromiso y las proyecciones de las variables antes del Covid-19/ Fuente: Elaboración propia con Rstudio

El análisis STATIS (Ilustración 16) muestra que el primer componente principal explica un 88.67% de la varianza, mientras que el segundo componente principal explica un 9.98%. Esto significa que el primer componente captura la mayor parte de la variabilidad en los datos, y los dos primeros componentes juntos explican un 98.65% de la varianza total. Este alto porcentaje de varianza explicada por los dos primeros componentes indica que el modelo es muy efectivo para representar la estructura de los datos.

Las contribuciones relativas de las comunidades autónomas al primer componente (Dim 1) son bastante altas en la mayoría de los casos. Destacan particularmente Cantabria (97.63%), Melilla (96.41%), y Cataluña (95.94%), entre otras, lo que indica que estas comunidades tienen una fuerte influencia en la estructura del primer componente. En el segundo componente (Dim 2), la Comunidad Valenciana (34.00%) y el País Vasco (58.59%) tienen contribuciones significativas, sugiriendo que estas regiones presentan variaciones específicas que son capturadas por el segundo componente.

Las calidades de representación acumuladas para las filas muestran que la mayoría de las comunidades autónomas están bien representadas por los dos primeros componentes, con muchas de ellas alcanzando valores cercanos al 100%. Las excepciones son Castilla y León (68.87%) y Galicia (78.76%), que tienen representaciones más bajas.

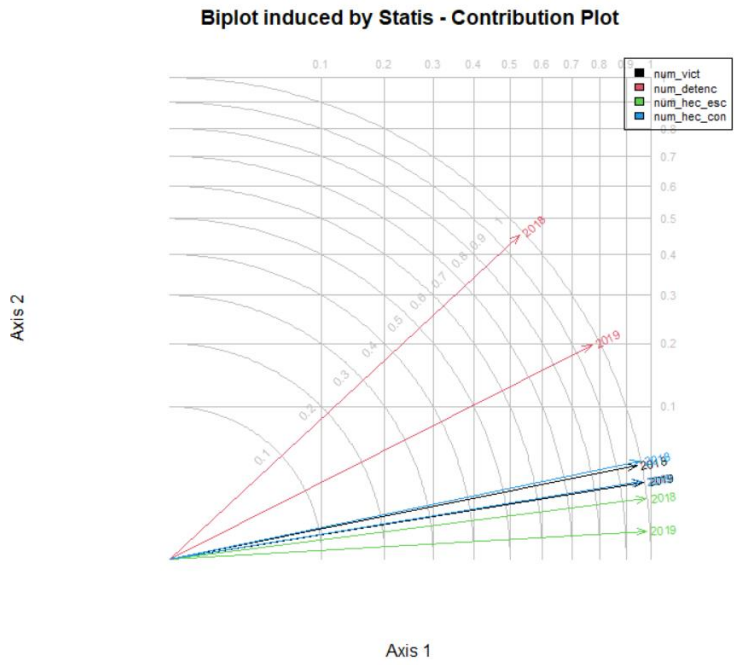


Ilustración 17: Gráfico de Contribuciones antes del Covid-19/Fuente: Elaboración propia con RStudio

En términos de las variables, las contribuciones relativas (Ilustración 17) al primer componente son muy altas para todas las variables en ambos años, con valores superiores al 90%. Esto sugiere que estas variables tienen un impacto significativo en la estructura de la primera componente. En la segunda componente, las variables "num_detenc-2018" (45.31%) y "num_detenc-2019" (19.73%) son las que más contribuyen, indicando que las detenciones presentan variaciones específicas que son capturadas por este componente.

Las calidades de representación acumuladas para las columnas también son altas, con la mayoría de las variables alcanzando valores cercanos al 100%. Esto confirma que las dos primeras dimensiones son adecuadas para representar las variaciones en las variables medidas en los años 2018 y 2019.

La posición relativa de las Comunidades Autónomas con respecto a las variables estudiadas puede interpretarse observando sus ubicaciones en el plano. Cuanto más cerca estén los puntos que representan a las comunidades autónomas, más parecidas serán sus puntuaciones en los índices analizados. De este modo, aquellas Comunidades autónomas que se encuentran en el primer y cuarto cuadrante de la figura, son las que tanto su número de victimizaciones, número de detenciones, número de hechos conocidos y número de hechos esclarecidos son más bajos para los distintos años de estudio, lo que sugiere que estas comunidades tienen menos problemas relacionados con los delitos de odio. Destacan algunas Comunidades Autónomas como la Rioja, Castilla y León, Asturias o Melilla.

Las Comunidades Autónomas que se ubican en dirección del eje 2, muy cerca de éste o que se encuentran en el cuadrante 2 del plano, se caracterizan por valores intermedios en todas las variables, destacan el País Vasco con mayor número de detenciones tanto en 2018 como en 2019. Las comunidades que se encuentran en este cuadrante tienen una mayor incidencia de delitos y una mayor respuesta en términos de detenciones y esclarecimientos de casos comparado con las Comunidades de los anteriores cuadrantes.

Las que se encuentran en el cuarto cuadrante, en este caso, solo Cataluña se caracterizan por tener un elevado número de victimizaciones, hechos conocidos y hechos esclarecidos tanto en 2018 como en 2019.

Para el segundo grupo, formado por los años 2020,2021 y 2022 obtenemos los siguientes resultados:

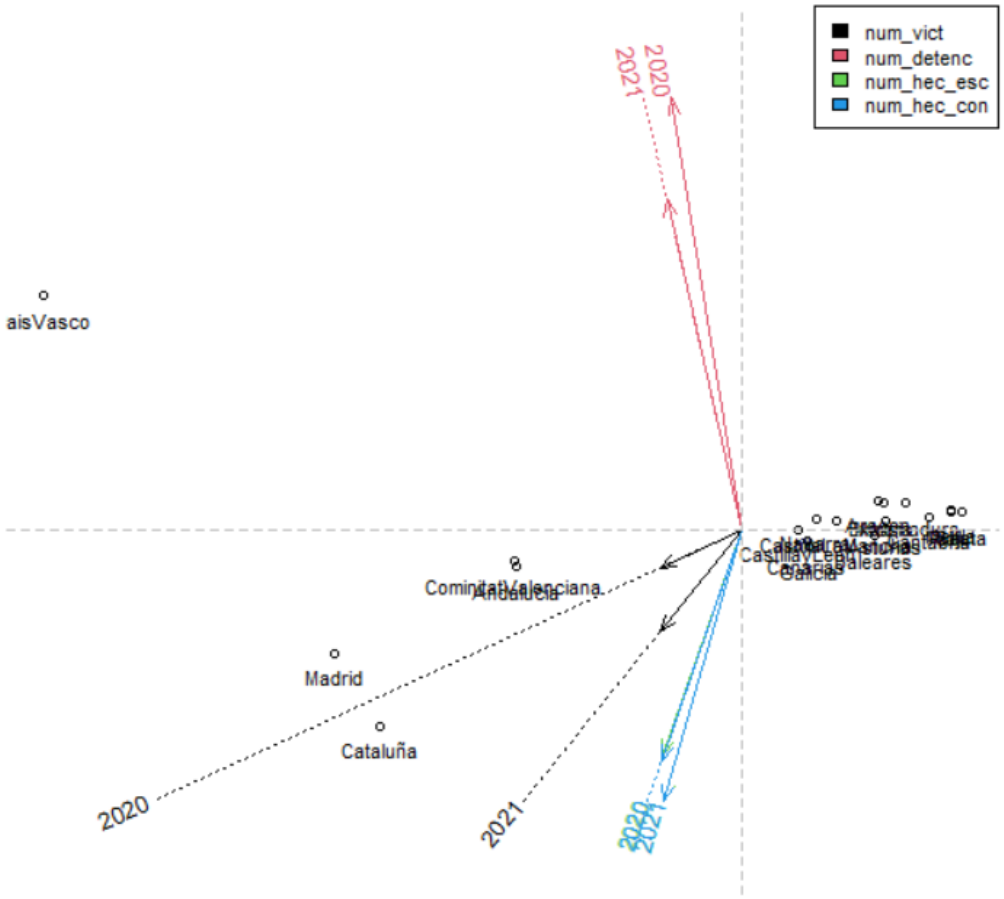


Ilustración 18: Biplot según las Comunidades Autónomas compromiso y las proyecciones de las variables después del Covid-19/ Fuente: Elaboración propia con Rstudio

El análisis STATIS (Ilustración 18) muestra que el primer componente principal explica un 88.92% de la varianza, mientras que el segundo componente principal explica un 8.63%. En conjunto, los dos primeros componentes explican un 97.55% de la varianza total. Este alto porcentaje de varianza explicada sugiere que el modelo es muy efectivo para representar la estructura de los datos.

Las contribuciones relativas de las comunidades autónomas al primer componente (Dim 1) son bastante altas en la mayoría de los casos. Destacan particularmente Cantabria (98.71%), Ceuta (98.95%), y Rioja (98.95%), entre otras, lo que indica que estas comunidades tienen una fuerte influencia en la estructura del primer componente. En el segundo componente (Dim 2), Cataluña (22.11%) y el País Vasco (10.19%) tienen contribuciones significativas, sugiriendo que estas regiones presentan variaciones específicas que son capturadas por el segundo componente.

Las calidades de representación acumuladas para las filas muestran que la mayoría de las comunidades autónomas están bien representadas por los dos primeros

componentes, con muchas de ellas alcanzando valores cercanos al 100%. Las excepciones son Castilla y León (71.15%) y Navarra (66.31%), que tienen representaciones más bajas.

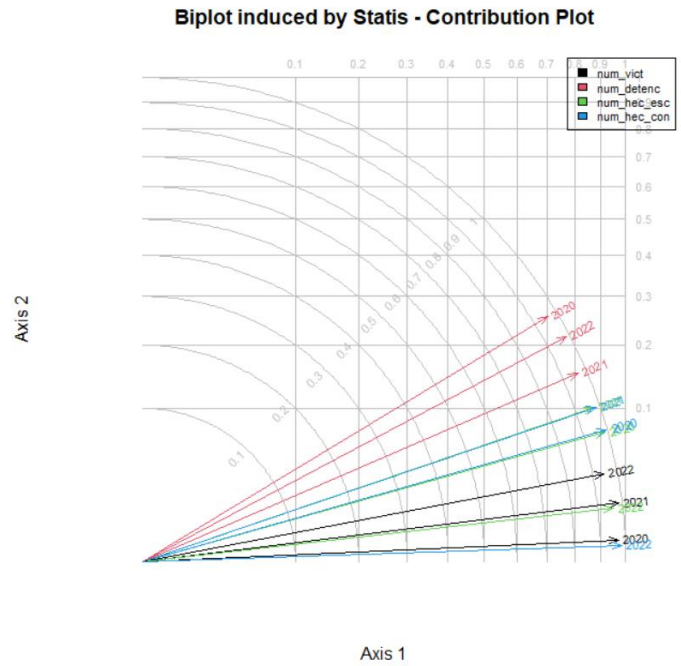


Ilustración 19: Gráfico de Contribuciones después del Covid-19 /Fuente: Elaboración propia con RStudio

En términos de las variables (Ilustración 19), las contribuciones relativas al primer componente son muy altas para todas las variables en los años 2020, 2021 y 2022, con valores superiores al 70%. Esto sugiere que estas variables tienen un impacto significativo en la estructura del primer componente. En el segundo componente, las variables "num_detenc-2020" (25.72%) y "num_detenc-2022" (21.55%) son las que más contribuyen, indicando que las detenciones presentan variaciones específicas que son capturadas por este componente.

Las calidades de representación acumuladas para las columnas también son altas, con la mayoría de las variables alcanzando valores cercanos al 100%. Esto confirma que las dos primeras dimensiones son adecuadas para representar las variaciones en las variables medidas en los años 2020, 2021 y 2022.

El análisis STATIS para los años 2020, 2021 y 2022, durante la pandemia de COVID-19, revela que las estructuras de datos son altamente similares entre estos tres años. El primer componente principal captura la mayoría de la variabilidad, lo que sugiere que las diferencias entre las comunidades autónomas en estos años están bien representadas por este componente. Las contribuciones significativas de las comunidades autónomas como Cantabria, Ceuta y Rioja al primer componente, y de Cataluña y el País Vasco al segundo componente, indican que estas regiones tienen patrones distintivos que afectan la estructura de los datos. Las variables relacionadas con el número de víctimas y hechos delictivos son las que más influyen en la estructura de los datos, lo que subraya su importancia en el análisis de la criminalidad en estos años.

En los siguientes gráficos, se presentan las trayectorias para los individuos en el subespacio del compromiso para ambos grupos:

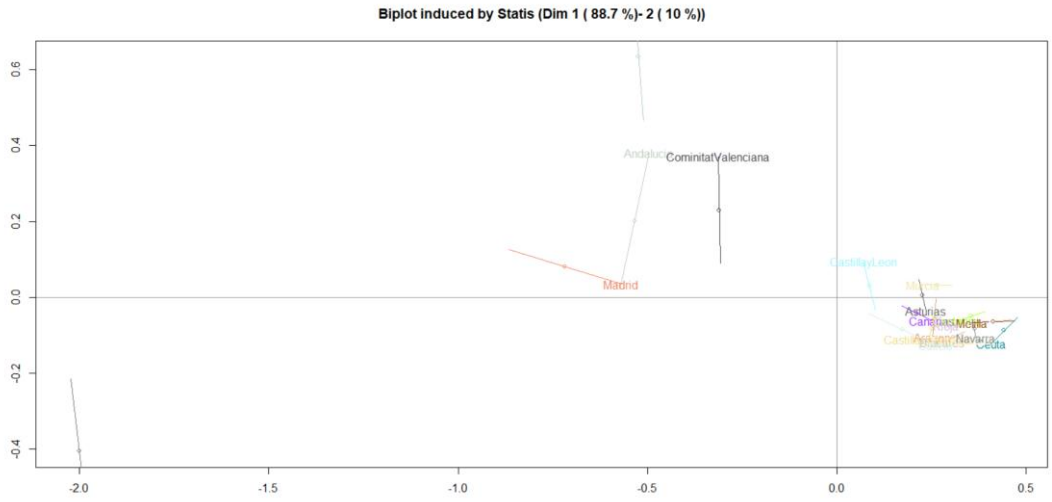


Ilustración 20: Trayectorias para los individuos antes del Covid-19/ Fuente: Elaboración propia con RStudio

Para los individuos del grupo 1 (años 2018 y 2019), se observan trayectorias excéntricas en algunas comunidades autónomas, lo que indica una mayor variabilidad y cambios significativos en la incidencia de delitos de odio durante estos años. Entre estas comunidades destacan: Comunidad Valenciana, Madrid y Andalucía. Por otro lado, otras comunidades autónomas presentan trayectorias más envolventes y regulares, lo que sugiere una estabilidad relativa en la incidencia de delitos de odio a lo largo del tiempo. Entre estas comunidades destacan: Castilla-La Mancha o Canarias.

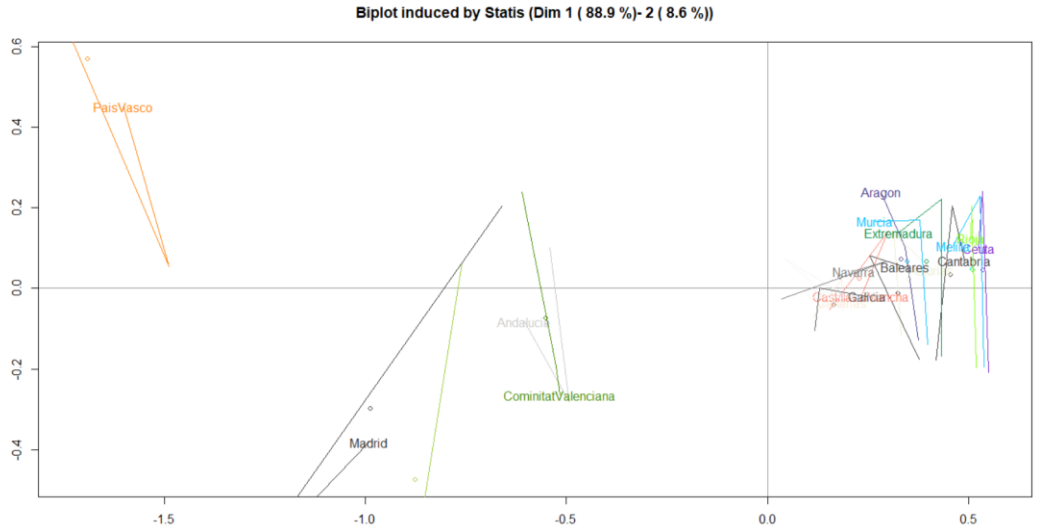


Ilustración 21: Trayectorias para los individuos después del Covid-19/ Fuente: Elaboración propia con RStudio

Para los individuos del grupo 2 (años 2020, 2021 y 2022), se observan trayectorias excéntricas también en las Comunidades de País Vasco, Madrid y Comunidad Valenciana, pero en este segundo periodo más pronunciadas que en el periodo anterior para todas ellas. El resto de Comunidades Autónomas también presentan trayectorias excéntricas pero en menor medida que en el resto.

5.4. Análisis de Clúster

Una vez realizado el análisis STATIS y su representación con distintos Bitplot, realizaremos un análisis K-Means para agrupar las Comunidades Autónomas en función de las distintas variables (num_vict, num_detenc, num_hec_esc, num_hec_con) que miden los delitos de odio. Este enfoque nos permitirá comprender mejor cómo se relacionan las comunidades autónomas y nos ayudará a identificar posibles áreas de intervención para la prevención y mitigación de los delitos de odio.

Antes de realizar el clustering k-means, es crucial seleccionar el número óptimo de clústeres, k, utilizando el método del Codo o el método de Siluette. Estos métodos nos ayudarán a determinar el número más adecuado de clústeres que mejor representen la estructura subyacente de los datos.

Como se puede observar, tanto para el gráfico del Método del Codo del primer grupo (Anexos: Ilustración 22) como para el gráfico del Metodo del Codo para el segundo grupo (Anexos: Ilustración 23), sugieren que el número óptimo de clústeres, k, sea igual a 3 para ambos conjuntos de datos. Ya seleccionado el número de clusters óptimo, aplicamos el algoritmo k-means con k=3 a nuestros datos.

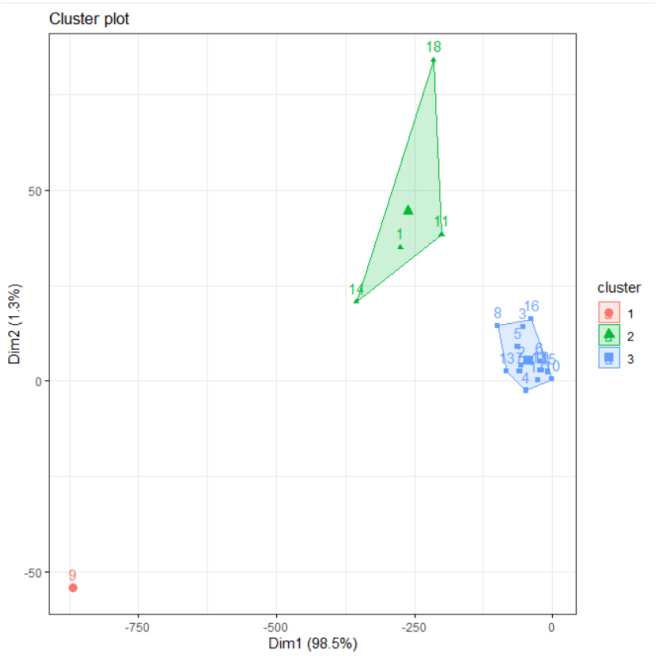


Ilustración 22: Clúster K-means por Comunidades Autónomas antes del Covid-19 /Fuente: Elaboración propia con RStudio

Los resultados del modelo K-means (Ilustración 22) para el grupo 1 (periodo 2018 y 2019) revelan la existencia de tres clusters con diferentes tamaños: el primer cluster formado por 1 individuo, el segundo formado por 4 individuos y el tercero con 14 individuos, como se evidenciaba en el Biplot anteriormente realizado.

El cluster 1, compuesto únicamente por Cataluña, se caracteriza por una alta incidencia de delitos de odio. Cataluña presenta un promedio de 603 víctimas, 85 detenidos, 280 hechos esclarecidos y 555 hechos conocidos. Estos números significativamente elevados en comparación con otras regiones reflejan la complejidad y los desafíos específicos que enfrenta esta comunidad.

Este segundo cluster incluye a Andalucía, Comunidad Valenciana, Madrid y País Vasco, regiones que muestran una incidencia moderada de delitos de odio con un promedio de 162 víctimas, 82.125 detenidos, 98.625 hechos esclarecidos y 167.125 hechos conocidos. Estas regiones urbanas, con importantes centros metropolitanos, presentan desafíos moderados en términos de delitos de odio, lo que sugiere la necesidad de políticas preventivas efectivas y programas de concienciación para promover la convivencia pacífica y la cohesión social. La diversidad dentro de este grupo refleja diferentes enfoques y contextos locales que deben considerarse al diseñar intervenciones específicas.

El tercer cluster agrupa a una variedad de regiones, incluyendo Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla-La Mancha, Castilla y León, Ceuta, Extremadura, Galicia, Melilla, Murcia, Navarra y La Rioja. Estas regiones muestran una baja incidencia de delitos de odio, con un promedio de 25.107 víctimas, 10.821 detenidos, 17.964 hechos esclarecidos y 29.071 hechos conocidos. La baja incidencia en estas áreas puede atribuirse a políticas locales efectivas y menores densidades poblacionales.

El análisis k-means muestra una mayor variabilidad interna en el Cluster 2 podría reflejar la diversidad de las regiones incluidas y las diferencias en las políticas locales y contextos sociales. La proporción de la suma de cuadrados entre los clusters respecto al total es del 96.1%, lo que indica que los clusters explican la mayor parte de la variabilidad en los datos. Esto sugiere que los grupos identificados son bastante representativos de las diferencias regionales en la incidencia de delitos de odio.

Para el segundo grupo compuesto por los periodos 2020, 2021 y 2022, es decir, después del Covid-19 obtenemos lo siguiente (Ilustración 23):

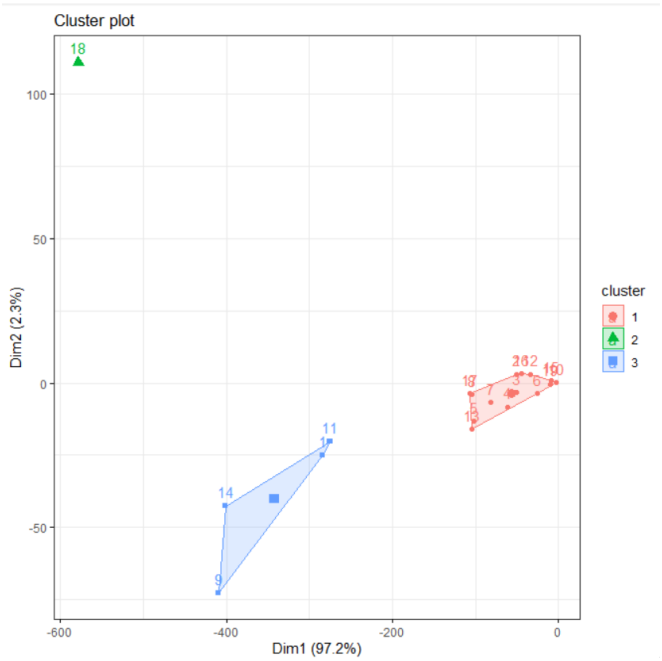


Ilustración 23: Clúster K-means por Comunidades Autónomas después del Covid-19 /Fuente: Elaboración propia con RStudio

El primer cluster incluye 14 comunidades autónomas: Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla-La Mancha, Castilla y León, Ceuta, Extremadura,

Galicia, Melilla, Murcia, Navarra y La Rioja. Estas regiones presentan una baja incidencia de delitos de odio, con un promedio de 34.619 víctimas, 14.524 detenidos, 22.738 hechos esclarecidos y 35.286 hechos conocidos, tiene una dispersión moderada, mostrando cierta heterogeneidad entre sus miembros, pero siguen estando significativamente más bajos en incidencia comparados con otros clusters. Las cifras relativamente bajas en estas áreas podrían ser indicativas de una menor actividad delictiva relacionada con el odio o de una mayor eficacia en las medidas preventivas y de resolución de conflictos.

El segundo cluster está compuesto únicamente por el País Vasco, destacándose por su muy alta incidencia de delitos de odio, con promedios de 420.333 víctimas, 245 detenidos, 166 hechos esclarecidos y 288.333 hechos conocidos. Estos altos números reflejan una problemática significativa que requiere intervenciones especializadas. Los factores que podrían contribuir a esta alta incidencia incluyen tensiones sociopolíticas, alta densidad poblacional y tal vez un mayor reporte de estos incidentes. Abordar esta situación requiere un enfoque multidimensional que incluya reforzar la vigilancia, mejorar la respuesta de las fuerzas de seguridad y desarrollar programas educativos y de integración comunitaria que promuevan la convivencia pacífica y el respeto mutuo.

El tercer cluster incluye a las comunidades de Andalucía, Cataluña, Comunidad Valenciana y Madrid, las cuales presentan una incidencia moderada a alta de delitos de odio. Los promedios de este grupo son de 215.333 víctimas, 67.333 detenidos, 142.417 hechos esclarecidos y 218.5 hechos conocidos. presenta la mayor dispersión interna, indicando una diversidad considerable en las incidencias de delitos de odio dentro de este grupo. Estas regiones abarcan áreas urbanas con altas densidades de población y diversidad cultural, factores que pueden contribuir tanto a la incidencia como a la visibilidad de los delitos de odio.

Los resultados del análisis K-Means revelan patrones distintivos en la incidencia de delitos de odio en las comunidades autónomas de España, tanto antes como después del COVID-19. Antes de la pandemia, las regiones se agruparon en clusters que reflejaban diferencias en la incidencia de delitos de odio, con algunas regiones mostrando una alta incidencia, otras una moderada y otras una baja.
.

6. CONCLUSIONES

El análisis exhaustivo de los delitos de odio en diferentes ámbitos sociales y comunidades autónomas a lo largo de los años estudiados proporciona varias conclusiones clave sobre las tendencias, patrones y factores asociados a estos incidentes. A partir de los resultados obtenidos en el análisis descriptivo, el análisis STATIS y el análisis de clúster, se pueden extraer las siguientes conclusiones:

1. El análisis descriptivo muestra una tendencia a la baja en los delitos de odio relacionados con la aporofobia lo que sugiere una escasez de incidentes en muchos casos, aunque pueden ocurrir de forma esporádica y grave. Por otro lado, los delitos de odio relacionados con creencias religiosas y discapacidad muestran una variabilidad considerable, con incidentes esporádicos pero graves, particularmente destacados en el caso de las creencias religiosas, donde los incidentes más graves parecen concentrarse en ciertos años. Además, se observa un aumento significativo en los delitos de odio relacionados con la identidad de género y el racismo/xenofobia. Este aumento se refleja

en un incremento considerable en las detenciones, hechos conocidos y el número de víctimas, lo que evidencia una creciente preocupación social por estos tipos de delitos. En ámbitos como la ideología y el sexo/género, aunque existen fluctuaciones, los patrones generales muestran una cierta estabilidad. Los incidentes son menos frecuentes que en otros ámbitos, pero pueden aumentar en respuesta a eventos sociales o políticos específicos.

2. El análisis STATIS revela una alta similitud en las relaciones entre las variables de los delitos de odio a lo largo de los años 2018 y 2022, indicando una continuidad notable en los patrones subyacentes. Aunque la mayoría de los años muestran correlaciones elevadas, se observan algunas variaciones sutiles que sugieren cambios en la estructura de estos delitos, posiblemente influenciados por factores externos como la pandemia de COVID-19. Los años 2020 y 2021 destacan por su estabilidad y coherencia significativas en los patrones de delitos de odio, posiblemente influenciados por la situación sanitaria. La contribución relativa de cada año a la estructura global de los datos es similar, aunque se observa una mayor influencia de los años 2020 y 2021. Además, se identifica una clara separación entre los años 2018-2019 y 2020-2022, indicando diferencias significativas en las estructuras de covariación entre estos conjuntos de años.

3. El análisis de clúster revela diferencias significativas entre las Comunidades Autónomas. Cataluña destaca con una alta incidencia de delitos de odio, mientras que otras regiones como Andalucía, Comunidad Valenciana, Madrid y País Vasco muestran una incidencia más moderada. Por otro lado, las demás regiones presentan una incidencia baja en comparación.

   Tras el impacto del COVID-19, se observaron cambios significativos en la dinámica regional de los delitos de odio en España. Mientras que algunas regiones mantuvieron una baja incidencia, otras experimentaron un aumento notable. Este cambio se evidencia aún más al observar los patrones de incidencia a lo largo de los años, donde entre 2020 y 2022 se registró un aumento general en la incidencia de estos delitos, manteniendo una configuración de clúster similar a años anteriores. Sin embargo, resalta que durante este período, el País Vasco destaca significativamente entre todas las demás regiones en cuanto a la frecuencia y gravedad de estos incidentes.

En general, la pandemia de COVID-19 ha tenido un impacto significativo en la incidencia de delitos de odio en España, influenciada por diversos factores. El aumento de la xenofobia y el racismo, particularmente hacia las comunidades asiáticas y otros grupos minoritarios, fue exacerbado por la crisis sanitaria y económica, que generó un clima de incertidumbre y estrés social. El confinamiento y el aislamiento social también contribuyeron a un incremento de la agresividad y la exposición a discursos de odio en línea. Además, la mayor conciencia y denuncia de estos delitos ha llevado a una visibilidad incrementada de los mismos.

En conclusión, este estudio proporciona una visión integral de la evolución y los factores que influyen en los delitos de odio en España, destacando el impacto de la pandemia de COVID-19 y las diferencias regionales significativas. Estas diferencias resaltan la importancia de adaptar las estrategias de prevención y respuesta de los delitos de odio para abordar los desafíos emergentes, con una colaboración entre comunidades, fuerzas de seguridad, y organizaciones de la sociedad civil siendo fundamental para promover entornos seguros y acogedores para todos los ciudadanos.

7. BIBLIOGRAFÍA

Ballesteros Espinoza, V. I. (2022). *Análisis Multivariante de los estilos de aprendizaje, estilos de pensamiento e inteligencia emocional en estudiantes de nivel medio y superior*. Universidad de Salamanca.

Cámara Arroyo, S. (2017). *El concepto de delitos de odio y su comisión a través del discurso. Especial referencia al conflicto con la libertad de expresión*.

Díaz López, J. A. (2020). *Informe de delimitación conceptual en materia de delitos de odio*.

Galindo, P. (2022). *Análisis Multivariante de Tablas de Tres Vías: Métodos de la Familia STATIS*.

García Domínguez, I. (2020). El tratamiento penal de los delitos de odio en España con la adopción de una perspectiva comparada. *Anuario Iberoamericano de Derecho Internacional Penal*, *8*. https://doi.org/10.12804/revistas.urosario.edu.co/anidip/a.9899

Giménez-Salinas Colomer, E. I., Román Maestre, B., & García Solé, M. (2003). *Sociedad abierta y delitos de odio en la era de la globalización*.

Giraldo Pérez, S. (2022). Hate crimes. Incidence of the COVID pandemic on hate crimes in Spain. *Sociologia y Tecnociencia*, *12*(1), 216–240. https://doi.org/10.24197/st.1.2022.216-240

González Bartolomé, D. (2015). *Análisis de tablas de tres entradas como herramienta en las ciencias atmosféricas. Estudio de las depresiones del Golfo de Génova y del Golfo de Cádiz mediante el STATIS DUAL y el Análisis Parcial Triádico*. Universidad de Salamanca.

Gonzalo, Á. (2019, March 8). *Segmentación utilizando K-means en Python*. Machine Learning Para Todos. https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/

Güerri Ferrández, C. (2015). La especialización de la fiscalía en materia de delitos de odio y discriminación. *InDret*.

II Plan de Acción de Lucha contra los Delitos de Odio. (2022).

Kenney, J. F., & Keeping, E. S. (1951). Mathematics of Statistics. Princeton, NJ: Van Nostrand.

López Ortega, A. I. (2017). Análisis y evolución de los delitos de odio en España (2011-2015). *Revista Extremeña de Cienciass Sociales "ALMENARA."*

Metodología Estadística. (n.d.). www.estadisticasdecriminalidad.es

Nieto Librero, A. B. (2022). *Los Métodos Biplot*.

Ortega, C. (n.d.). *Matriz de correlación: Qué es, cómo funciona y ejemplos*. Retrieved May 26, 2024, from https://www.questionpro.com/blog/es/matriz-de-correlacion/

Pérez Álvarez, F., Janeth Villasante Arroyo, N., Cortés, D., Mariola, L., Campo, H., & Teresa, M. (2016). *Propuestas penales: Nuevos retos y modernas tecnologías*.

Pérez, M., Andrea, R., Framis, G.-S., Méndez, R. C., Ana, L., Martínez, S., & Chiclana De La Fuente, S. (2021). *Informe del Estudio sobre Delitos de Odio: Perfil de las personas condenadas por delitos de odio a prisión y a penas y medidas alternativas a la prisión*.

Perucha Jurjo, C. (2022). *El Método de K Medias*.

Rodríguez Expósito, C. (2016, July 11). *Discurso del odio o Hate Speech*. Crimipedia. https://crimipedia.umh.es/topics/discurso-del-odio-hate-speech/

Rodríguez Martínez, C. C. (2020). *Contribuciones a los Métodos STATIS basados en Técnicas de Aprendizaje no Supervisado*.

Vicente-Galindo, P. (2013). Análisis de tablas de tres vías : recientes desarrollos del STATIS. Trabajo fin de master, 1-65.

Vicente Villardón, J. L. (n.d.). *LOS MÉTODOS BIPLOT*.

Vicente Villardón, J. L. (2007). *Introducción al Análisis de Clúster*.

Vicente-Villardón, J. L., Vicente-Gonzalez, L., & Frutos-Bernal, E. (2023). *Package "MultBiplotR" Type Package Title Multivariate Analysis Using Biplots in R*.

Vintimilla, C., Astudillo-Salinas, F., Severeyn, E., Encalada, L., & Wong, S. (2017). *Agrupamiento de K-medias para estimación de insulino-resistencia en adultos mayores de Cuenca*.