

Análisis multivariado aplicado al banco de datos de violencia política en Colombia para el año
2020

Juan Sebastián Barreto Lozano
Frank Sebastian Puentes Vasquez

UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS
BOGOTÁ, D. C. 2022

Análisis multivariado aplicado al banco de datos de violencia política en Colombia para el año
2020

Presentado por:

Juan Sebastián Barreto Lozano

C.C.1070973216

Código: 2017240003

Frank Sebastian Puentes Vasquez

C.C. 1073523618

Código: 2017240059

Trabajo de grado presentado como requisito parcial para optar por el título de Licenciado en
Matemáticas

Director:

Felipe Jorge Fernández Hernández

Magíster en Estadística

UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS
BOGOTÁ, D. C. 2022

Agradecimientos

A mi madre quien a lo largo de su vida a hecho sacrificios para garantizar un mejor futuro sus hijos, todo mi amor y cariño para ella. A todos los profesores de la universidad que hicieron posible mi formación como docente y particularmente al profesor Felipe por prestarnos sus conocimientos y su tiempo para acercarnos a la estadística y a los problemas que a menudo no tienen suficiente atención.

-Juan Sebastián.

A mis padres, por su apoyo incondicional y su amor puro, no hallo manera de pagarles por ello. También al profesor Felipe por su guía durante este estudio, gracias por compartirnos sus conocimientos y ser ayuda en el camino.

-Frank Puentes

RESUMEN

Durante el largo y complejo conflicto armado en Colombia, del cual se tienen registros oficiales o de las ONG, se continúan reconociendo miles de casos de violaciones a los derechos humanos cada año. Esto sin contar con subregistro derivado del miedo a denunciar de las víctimas o de las ventajas del anonimato para los victimarios. Sin embargo, los datos que se logran recolectar son de gran importancia y merecen atención desde la academia. Esta monografía propende por entregar herramientas para interpretar el fenómeno de las violaciones a los derechos humanos, desde una comprensión del contexto a través de la ayuda de un análisis estadístico multivariado. Además, para intuir algunas conclusiones y no quedarse solo en una perspectiva frecuentista, se intenta fortalecer la teoría que hay detrás de la utilización de estos métodos estadísticos, que en últimas facilitaron la visualización de los datos considerados y las conclusiones derivadas de estos.

Palabras clave: derechos humanos, violaciones a los derechos humanos, presuntos responsables, revista Noche y Niebla, análisis de componentes principales, análisis de correspondencias simple, álgebra lineal.

Tabla de contenido

INTRODUCCIÓN	8
PRESENTACIÓN DEL PROBLEMA	8
2.1 JUSTIFICACIÓN	8
2.2 OBJETIVOS	11
2.3 ASPECTOS METODOLÓGICOS	12
ACERCA DE LOS DERECHOS HUMANOS (DDHH)	16
3.1 ALGUNOS DE SUS ANTECEDENTES	16
3.2 REVISTA NOCHE Y NIEBLA	21
3.3 VIOLACIONES A LOS DDHH Y PRESUNTOS RESPONSABLES	23
3.3.1 Violencia Política como Violación de Derechos Humanos (DH)	23
3.3.2 Violencia Político-Social (VPS)	26
3.3.3 Infracciones Graves al Derecho Internacional Humanitario Consuetudinario (DIHC)	28
3.3.4 Acciones béticas o Infracciones graves al Derecho Internacional Humanitario (DIH)	34
3.4 DATOS DE BASE REVISTA NOCHE Y NIEBLA	35
3.4.1 Categorías dadas por la revista	35
3.4.2 Consideraciones acerca de los presuntos responsables de la Tabla 3	38
3.4.3 Decisiones finales acerca de la tabla de contingencia	46
EL ÁLGEBRA LINEAL COMO FUNDAMENTO MATEMÁTICO DEL ANÁLISIS ESTADÍSTICO MULTIVARIADO.....	48
4.1 COMBINACIONES LINEALES, SISTEMA GENERADOR Y BASES DE VECTORES	50
4.2 TRANFORMACIONES LINEALES Y MATRICES	52
4.3 MULTIPLICACIÓN DE MATRICES COMO COMPOSICIÓN	54
4.4 EL DETERMINANTE	56
4.5 MATRICES INVERSAS, ESPACIOS DE LAS COLUMNAS Y NÚCLEO	58
4.6 CAMBIO DE BASE	62
4.7 VALORES Y VECTORES PROPIOS	67
ELEMENTOS DEL ANÁLISIS ESTADÍSTICO MULTIVARIADO	74
5.1 ¿QUÉ ES UN ANÁLISIS MULTIVARIADO?	74
5.2 METODOLOGÍAS Y ENFOQUES DEL ANÁLISIS MULTIVARIADO	76
5.3 TABLA DE CONTINGENCIA	77
5.4 MODELO DE INDEPENDENCIA	79
5.5 TÉCNICAS MULTIVARIABLES	80
5.6 CONSIDERACIONES DEL ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)	80
5.6.1 El ACP según algunos autores	85
RESULTADOS	97
6.1 ANÁLISIS DE CORRESPONDENCIAS SIMPLE (AC)	97
CONCLUSIONES Y REFLEXIONES	110
REFERENCIAS.....	113

Lista de tablas

TABLA 1.	FRAGMENTO DE UNA TABLA DE LA REVISTA NOCHE Y NIEBLA	36
TABLA 2.	FRAGMENTO DE DATOS EN EXCEL	37
TABLA 3.	TABLA DE CONTINGENCIA, PRIMERA VERSIÓN	38
TABLA 4.	MÁS DE UN PR EN MUERTOS DH.....	39
TABLA 5.	MÁS DE UN PR EN DESAPARICIÓN DH.....	40
TABLA 6.	MÁS DE UN PR EN LESIONADOS DH	41
TABLA 7.	MÁS DE UN PR EN LESIONADOS DIHC	42
TABLA 8.	MÁS DE UN PR EN DETENCIÓN ARBITRARÍA DH.....	43
TABLA 9.	MÁS DE UN PR EN AMENAZAS DH.....	44
TABLA 10.	MÁS DE UN PR EN AMENAZA DIHC	45
TABLA 11.	TABLA DE CONTINGENCIA CON SOLO UN PRESUNTO RESPONSABLE POR CATEGORÍA DE VICTIMIZACIÓN.....	46
TABLA 12.	TABLA DE CONTINGENCIA FINAL.....	48
TABLA 13.	TC DE VIOLACIONES A LOS DDHH Y PRESUNTOS RESPONSABLES CON MARGINALES	98
TABLA 14.	TABLA DE FRECUENCIAS RELATIVAS F	98
TABLA 15.	PERFILES FILA.....	99
TABLA 16.	PERFILES COLUMNA	100
TABLA 17.	COORDENADAS DE LAS CATEGORÍAS EN LA FIG. 37.....	108

Lista de Figuras

FIGURA 1.	VECTORES UNITARIOS EN DOS DIMENSIONES.....	50
FIGURA 2.	VECTOR EN TÉRMINOS DE VECTORES UNITARIOS O VECTORES BASE DEL SISTEMA COORDENADO <i>xy</i>	51
FIGURA 3.	TRANSFORMACIÓN LINEAL EN 2D	53
FIGURA 4.	MULTIPLICACIÓN MATRIZ-VECTOR	54
FIGURA 5.	IZQUIERDA: 90° EN SENTIDO ANTIHORARIO, DERECHA: TRANSFORMACIÓN “SHEAR”.	55
FIGURA 6.	ROTACIÓN 90° EN SENTIDO ANTIHORARIO SEGUIDA DE LA TRANSFORMACIÓN “SHEAR”	55
FIGURA 7.	DETERMINANTE VISTO COMO ÁREA, ANTES Y DESPUÉS DE APLICAR LA TRANSFORMACIÓN LINEAL	56
FIGURA 8.	EJEMPLOS DE ALGUNAS ÁREAS RESULTANTES, LUEGO DE APLICAR UNA TRANSFORMACIÓN LINEAL ESPECÍFICA	57
FIGURA 9.	DETERMINANTE MENOR QUE CERO, INVIERTE ORIENTACIÓN DEL ESPACIO	58
FIGURA 10.	MATRICES PARA SOLUCIONAR SISTEMAS DE ECUACIONES	59

<i>FIGURA 11.</i>	<i>EL VECTOR x ATERRIZANDO EN EL v</i>	59
<i>FIGURA 12.</i>	<i>REPRODUCIENDO EN REVERSA DE LA TRANSFORMACIÓN LINEAL DADA POR UNA MATRIZ A</i>	60
<i>FIGURA 13.</i>	<i>TRANSFORMACIÓN LINEAL QUE PASA DE 2D A 1D Y ENVÍA VECTORES A $(0, 0)$</i>	62
<i>FIGURA 14.</i>	<i>VECTORES ALTERNATIVOS PARA UNA BASE EN 2D</i>	63
<i>FIGURA 15.</i>	<i>VECTOR $(3, 2)$ EN 2D</i>	63
<i>FIGURA 16.</i>	<i>VECTOR $(\frac{5}{3}, \frac{1}{3})$ EN SISTEMA COORDENADO ALTERNATIVO</i>	64
<i>FIGURA 17.</i>	<i>ROTACIÓN DE 90° EN SENTIDO ANTIHORARIO PARA UN ESPACIO GENERADO POR VECTORES ALTERNATIVOS</i>	66
<i>FIGURA 18.</i>	<i>VECTOR SE SALE DE SU “SPAN” LUEGO DE APLICAR LA TRANSFORMACIÓN EN EL SISTEMA COORDENADO EN 2D</i>	67
<i>FIGURA 19.</i>	<i>VECTOR PROPIO DE $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$</i>	68
<i>FIGURA 20.</i>	<i>VECTOR PROPIO COMO EJE DE ROTACIÓN EN 3D</i>	69
<i>FIGURA 21.</i>	<i>VALOR PROPIO DE 1</i>	71
<i>FIGURA 22.</i>	<i>VALOR Y VECTORES PROPIOS DE UNA TRANSFORMACIÓN “SHEAR”</i>	72
<i>FIGURA 23.</i>	<i>CONSIGUIENDO UNA BASE PROPIA</i>	73
<i>FIGURA 24.</i>	<i>OSCURIDAD DEL VINO Y CANTIDAD DE ALCOHOL</i>	82
<i>FIGURA 25.</i>	<i>PROYECTANDO NUBE DE PUNTOS EN UNA RECTA, TRES OPCIONES</i>	82
<i>FIGURA 26.</i>	<i>TABLA DE INDIVIDUOS Y VARIABLES CUANTITATIVAS ACP</i>	86
<i>FIGURA 27.</i>	<i>DOS PERSPECTIVAS PARA ESTUDIAR UNA TABLA DE DATOS: CON I INDIVIDUOS Y K VARIABLES</i>	87
<i>FIGURA 28.</i>	<i>NUBE DE INDIVIDUOS, Y</i>	88
<i>FIGURA 29.</i>	<i>NUBE DE PUNTOS CENTRADA YC</i>	90
<i>FIGURA 30.</i>	<i>PROYECCIÓN DEL VECTOR xi' SOBRE u</i>	92
<i>FIGURA 31.</i>	<i>PRIMER PLANO FACTORIAL DEL ACP ACERCA DE LOS DIEZ CAFÉS</i>	95
<i>FIGURA 32.</i>	<i>FÓRMULAS PARA EL ACP (X, M, N)</i>	97
<i>FIGURA 33.</i>	<i>DIAGRAMA DE BARRAS PERFILES FILA</i>	101
<i>FIGURA 34.</i>	<i>DIAGRAMA DE BARRAS PERFILES COLUMNA</i>	101
<i>FIGURA 35.</i>	<i>PRIMER PLANO FACTORIAL PERFILES FILA</i>	106
<i>FIGURA 36.</i>	<i>PRIMER PLANO FACTORIAL PERFILES COLUMNA</i>	107
<i>FIGURA 37.</i>	<i>REPRESENTACIÓN SIMULTÁNEA DE LOS PLANOS FACTORIALES</i>	108

Introducción

La presente monografía desarrollará un tema de interés de los autores relacionado con la utilización de un análisis estadístico multivariado sobre el contexto de violaciones a los derechos humanos. Los datos para analizar fueron obtenidos de la revista Noche y Niebla en su edición 62 del 2020. El tema se trabaja en seis capítulos: primero, en “*Presentación del problema*” se comentarán los objetivos, justificación y aspectos metodológicos; luego, en “*Acerca de los Derechos Humanos (DDHH)*”, se ubican secciones en las cuales hay antecedentes de este contexto, conceptualización de diferentes categorías relacionadas (utilizadas por la revista), comentarios y decisiones respecto a los datos de base para aplicar el análisis. Los siguientes capítulos, están relacionados con el sustento matemático y la aplicación del análisis seleccionado: en “*El álgebra lineal como fundamento matemático del análisis estadístico multivariado*” y “*Elementos del análisis estadístico multivariado*” se presentan un compendio de una considerable cantidad de conceptos necesarios para comprender este tipo de análisis. Finalmente, en “*Resultados*”, se realiza la explicación del Análisis de Correspondencias simple y paralelamente se analiza la tabla de contingencia surgida de los datos del contexto. Esta herramienta generará interpretaciones que también se comentarán. El documento termina con algunas conclusiones y reflexiones acerca de todo el proceso, desde nuestra perspectiva como profesores de matemáticas.

Presentación del problema

2.1 Justificación

Colombia es un país en el que es posible identificar violaciones de los derechos humanos (en adelante DDHH), sin temor a equivocarse, cada semana. Según varias de las referencias consultadas para este documento, existe una sensación de normalidad en el país ante las

injusticias que se cometan contra la población (ya sea por actores del estado o ilegales) y que fácilmente quedan en la impunidad. Además, la búsqueda de información que al respecto se ha consolidado, ha permitido identificar bancos de datos relacionados con este flagelo. Tales bancos recolectan, muy diligentemente, denuncias, relatos, posibles responsables, entre otros; en aras de darles una voz a aquellas personas que no tendrán posibilidad de obtener justicia mediante la institucionalidad. De hecho, uno de los motivos por los cuales esta labor de los bancos de datos es relevante, tiene que ver con la lentitud con la cual avanzan las investigaciones y los procesos judiciales en Colombia.

Si bien hay disponible mucha información sobre denuncias relacionadas con violaciones a los DDHH, la presentación de dicha información no suele, en lo concerniente a su análisis estadístico, ir más allá de recuentos de frecuencias y/o la publicación de gráficas e indicadores porcentuales básicos. Conviene, sin embargo, advertir que los datos a los que hemos tenido acceso se presentan de manera descriptiva a través de relatos. Luego son clasificados en distintas categorías entre las cuales se identifica qué derecho humano fue vulnerado (y sus respectivas subcategorías), presuntos responsables, ubicaciones, etc. En concreto, existe la posibilidad de tomar estas frecuencias absolutas de las diferentes clasificaciones mencionadas y organizarlas sistemáticamente. En este trabajo de grado se propone una organización de los datos la cual favorece la realización de un análisis estadístico multivariado, concretamente usando las categorías de las violaciones a los DDHH y los presuntos responsables.

Gracias a las herramientas que proveen los análisis de datos multivariados y en nuestro caso específico el análisis de correspondencias simple para las dos categorías anteriormente mencionadas, dan acceso a la posibilidad de, por ejemplo, encontrar características ocultas en estos grandes grupos de datos. Habría que decir también que, aplicar este tipo de herramientas a

los datos considerados, ayuda a producir conclusiones más detalladas, en comparación a las que se derivan de analizar la información recolectada desde puntos de vista descriptivos elementales. Además, estas conclusiones podrían ayudar a darles una observación más clara de este tipo de datos a los actores claves de la sociedad en esta materia, como, por ejemplo, los defensores de DDHH.

Según la revista Noche y Niebla (Cinep/PPP, 2021), uno de los factores más relevantes respecto a las violaciones de los DDHH en Colombia, tiene que ver con la dificultad que supone establecer la frontera entre las luchas legítimas civiles y luchas armadas. El hecho de mantener una falta de claridad entre estas dos dimensiones, por ejemplo, termina siendo un mecanismo que conviene para mantener el estado de las cosas con mejores condiciones para unos pocos, mientras la mayoría del país vive precariamente y con miedo.

Una investigación que se adentre en datos relacionados con las violaciones de los DDHH podría tener validez para ver más claramente el tipo de frontera descrita, reivindicando la lucha social y siendo de gran trascendencia para personas interesadas. Aunque los resultados provengan de la aplicación de análisis estadísticos multivariados, que conllevan usar matemáticas avanzadas, con estos se generan conclusiones o descripciones, que pueden ser presentadas en el lenguaje del contexto.

Para Contreras et al (2019), el contexto es uno de los elementos de la cultura estadística, el cual, por ejemplo, no necesita de una formación matemática avanzada, pero ayuda a obtener una comprensión más profunda de los datos que se puede encontrar una persona del común, evitando en cierto grado la desinformación. Sin embargo, no se excluye la posibilidad de una manipulación intencionada o no de los datos.

Por lo tanto, y atendiendo a los argumentos expresados anteriormente, para este trabajo se plantean los siguientes objetivos.

2.2Objetivos

Como objetivo general, se tiene la intención de caracterizar las violaciones de DDHH ocurridas en el segundo semestre del 2020. A partir de un análisis estadístico multivariado, se pretende relacionar información acerca de los presuntos responsables de una determinada agresión, junto con las diferentes categorías para las violaciones de derechos fundamentales. Estas últimas, son posibles de caracterizar gracias a la revista Noche y Niebla¹, especializada en el tema (en la sección 3.4, se amplía la información acerca de esta fuente).

Este objetivo general, se pretende alcanzar a través de los siguientes objetivos específicos:

- Establecer un marco de referencia para el análisis de violaciones de DDHH.
- Seleccionar y aplicar un método de análisis estadístico multivariado que contribuya a describir las violaciones a DDHH (Análisis de correspondencias simple y/o Análisis de componentes principales).
- Sistematizar y analizar los resultados obtenidos del análisis estadístico multivariado.
- Fundamentar, de manera intuitiva, los aspectos básicos matemáticos del álgebra lineal que requieren los análisis multivariados.

¹ La revista Noche y Niebla es una publicación que cuenta con ISSN 0123-3637, ver: <https://nocheyniebla.org/> y su información se basa en la red Bandatos que cuenta con 43211 casos reportados a la fecha de escritura de este documento. Ver: https://www.noche niebla.org/?page_id=1372

2.3 Aspectos metodológicos

El enfoque que se tuvo en este trabajo de grado sigue los lineamientos de una investigación cuantitativa. Según Sampieri et al (2014), está presenta algunas de las siguientes características: planteamientos acotados, medición de fenómenos, utilización de estadística y pruebas de hipótesis. Para estos autores las fases de una investigación cuantitativa son las siguientes:

Fase 1	Fase 2	Fase 3	Fase 4	Fase 5	Fase 6	Fase 7	Fase 8	Fase 9	Fase 10
Concepción de la idea de investigación	Plantear el problema de investigación	Desarrollar la perspectiva teórica	Definir el alcance de la investigación	Establecer las hipótesis, definiendo las variables	Concebir o elegir un diseño apropiado	Seleccionar una muestra adecuada	Recolectar datos	Analizar los datos	Elaborar el reporte de resultados

Para los propósitos de este trabajo de grado se realizó una adaptación de tales fases. La justificación de esto atiende a que no se considera ineludible trabajar las fases 5 a 8: estas son etapas del proceso que dadas las fuentes de información que se tenían a la mano, no era necesario contemplarlas. En efecto, el establecimiento de hipótesis, la definición de variables, la elección de un diseño experimental, la toma de una muestra y la recolección de datos; no era inevitable tenerlas en cuenta, dado que, por el carácter de datos seleccionados, un trabajo al respecto no tenía lugar. En lo que sigue y para no confundir con las fases de los mencionados autores, utilizaremos la palabra “etapa” en lugar de “fase”, para aludir al proceso seguido en este trabajo.

Etapa 1. Concepción de la idea de investigación

Esta etapa, por supuesto está incluida en el trabajo de grado. Surgió de la necesidad de comprender a profundidad un tema complejo como el de las violaciones a los DDHH, pero que está latente en la realidad colombiana. Esta idea de investigación propende por conceptualizar, según Sampieri et al (2014), esto es: “detectar un fenómeno o problema de investigación que requiere indagarse en profundidad o aportarse mayor conocimiento o evidencia para conocerlo, definirlo, describirlo y/o comprenderlo” (p.25).

También es relevante mencionar que la idea se originó en el momento convulsionado del paro del 2021. Por esas fechas y a pesar de la virtualidad en la que se desarrollaban casi todas las actividades de la universidad, se decidió apoyar el paro. Particularmente en la Licenciatura en Matemáticas hubo una agenda propuesta por estudiantes y docentes para tratar temas relacionados con la difícil situación del país en lugar de las clases. En uno de esos espacios, de parte de algunos docentes, se hizo un acercamiento inicial a una respuesta desde la academia, utilizando bases de datos de violaciones a los DDHH y la aplicación de análisis estadísticos sobre estos.

Etapa 2. Plantear el problema de investigación

En esta etapa es necesario plantear objetivos, preguntas, justificación y viabilidad. Los objetivos y la justificación, que ya se mencionaron anteriormente. Ahora, para las preguntas de investigación, se considerarán las siguientes:

- ¿Se puede detectar algún patrón entre los datos de las violaciones de los DDHH con los presuntos responsables?
- ¿Al realizar la descripción de los datos se podrán hacer comparaciones entre grupos al margen de la ley y la institucionalidad?
- ¿Puede ser un análisis estadístico multivariado una herramienta útil para este tipo de contexto?

Respecto a la viabilidad de la investigación, consideramos que el trabajo de grado la tiene. Aunque no se recolectan datos para analizarlos con la estadística, si se usa esta última en un grupo de datos confiable, de fuentes reputadas que llevan años trabajando en el tema. Además, de acuerdo con los tipos de variables utilizados en el banco de datos consultado, se

utilizaron métodos estadísticos que se sugieren en la literatura para realizar un estudio de esta naturaleza.

Etapa 3. Desarrollar la perspectiva teórica

La perspectiva teórica que se desarrolla a lo largo del documento se divide en dos partes. Primero, una relacionada con el ámbito del contexto de los DDHH y las diferentes categorías, definiciones y conceptos involucrados en los temas relacionados con las violaciones a los DDHH. Dichos temas están descritos en la sección: “*Acerca de los derechos humanos*”. La otra parte conceptual del documento está enfocada en sustentar el método estadístico multivariado utilizado en el análisis de los datos. Para este cometido, hay primero una sección relacionada con los conceptos básicos del álgebra lineal, necesarios para comprender el sustento matemático de los análisis estadísticos multivariados, desde una perspectiva intuitiva (ver “*El álgebra lineal como fundamento matemático del análisis multivariado*”); luego, en la sección “*Elementos del análisis multivariado*”, se amplía la información respecto a estos. Finalmente, en la sección “*Resultados*” se desarrollan las explicaciones del Análisis de Correspondencias simple (análisis utilizado para el contexto de violaciones a los DDHH) a la par que se van describiendo los resultados obtenidos.

Etapa 4. Definir el alcance de la investigación

En esta etapa se define el alcance que puede ser exploratorio, descriptivo, correlacional o explicativo. Estos alcances deben verse como una causalidad y no como un tipo de investigación, puesto que no es una clasificación y se pueden tener elementos que sean de más de un alcance (Sampieri et al, 2014).

Para este trabajo de grado, los alcances utilizados son el descriptivo y el de correlación.

El primero es la base para las investigaciones correlacionales, tiene como objetivo “medir o recoger información de manera independiente o conjunta sobre los conceptos o las variables a las que se refieren” (Sampieri et al, 2014, p. 92). Este alcance no indica cómo se relacionan las variables; a diferencia del correlacional, que tiene como objetivo identificar la relación que existe entre dos o más variables/categorías en algún contexto en particular.

Cabe distinguir que el alcance de la investigación no pretende ser inferencial ni enunciar generalizaciones; más bien, en primera medida, el propósito es descriptivo² respecto a las variables involucradas en el estudio. Posteriormente en la fase de selección y agrupación de las variables que conducen a la tabla de contingencia resultante, el proceso deriva en un Análisis de Correspondencias simple (AC) que tiene el alcance de un estudio correlacional. En otras palabras, dicho análisis estadístico multivariado entregará posibles relaciones entre las categorías de las columnas y las filas de la tabla de contingencia. Esto último se amplía en la sección “*Resultados*”.

Etapa 5. Analizar los datos

En esta etapa se consideran los siguientes elementos (basados en Sampieri, et al, 2014, p. 272): primero los softwares Excel y R como herramientas apropiadas para analizar los datos; segundo la exploración de los datos gracias a las salidas generadas por estos programas: en la sección “*Resultados*” se podrá ver cómo es posible sacar algunas intuiciones generadas a partir de histogramas. Por otro lado, un elemento de importancia es el uso de paquetes de R, los cuales

² Una de las ganancias que tiene el *análisis estadístico multivariante descriptivo* es el de congregar métodos estadísticos que se orientan a observar y procesar simultáneamente diferentes variables estadísticas para lograr información relevante.

permiten generar salidas en gráficas para presentar los resultados de aplicar un Análisis de Correspondencias simple a los datos considerados. Este último elemento, habilita finalmente poder explorar cómo se agrupan o dispersan los datos, gracias a que el análisis aplicado da acceso a una visualización con más significado que solo la tabla de contingencia.

Etapa 6. Elaborar el reporte de resultados

En esta etapa es importante definir el motivo por el cual se originó el estudio; el tipo de receptor que tomará las decisiones con base en los resultados; y el contexto en el cual será mostrada la investigación. Para este caso, el motivo es realizar una monografía para obtener un grado académico (licenciatura), el contexto será académico y los receptores serán los jurados en primera instancia, para luego ser profesores o estudiantes de la universidad.

Los apartados que contiene el trabajo de grado para la etapa de reportes de resultados, al ser de contexto académico son: la portada, el resumen, el índice, el cuerpo del documento y las referencias. En cuanto al cuerpo del documento se incluye: la introducción, el marco teórico, los resultados y conclusiones.

Acerca de los Derechos Humanos (DDHH)

3.1 Algunos de sus antecedentes

El interés en el estudio estadístico de algún tipo de datos, situados en un contexto, requiere la recolección de información que permita comprender con mayor profundidad el significado de los datos. Detrás de las cifras o de disponer números en tablas, están las historias que permiten darles significado y que ayudarían a interpretar con mayor profundidad los análisis aplicados.

Específicamente respecto a los DDHH, es común escuchar diferentes voces acerca del tema. En el caso colombiano predomina la información relacionada con atropellos, tanto del estado como de grupos al margen de la ley, contra la población. Estos han sido visibilizados ya sea gracias a alguna ONG, medios independientes, entes internacionales, etcétera. El tema está en el orden del día, más aún en Colombia: un país históricamente atravesado por el flagelo de la violencia.

Pensado detenidamente en la información que le llegaría a un individuo promedio, por ejemplo, en el contexto colombiano acerca de los DDHH, no es arriesgado concluir que su conocimiento del tema estará limitado a: noticias alarmantes, cifras que revelan una constante acumulación de casos en los cuales se violentan los derechos fundamentales de las personas y la impunidad relacionada con estos hechos. Entonces, para superar un poco este nivel superficial de comprensión, valdría la pena tener en cuenta de qué manera se llegaron a concebir los DDHH en cuanto a su naturaleza y algo de su historia.

La naturaleza de los DDHH está en el concepto de la dignidad de las personas, Carpizo (2011), ofrece una definición clara de esto en el siguiente aparte:

La dignidad de la persona es el rasgo distintivo de los seres humanos respecto de los seres vivos, la que constituye a la persona como un fin en sí mismo, impidiendo que sea considerada un instrumento o medio para otro fin, además de dotarlo de capacidad de autodeterminación y de realización del libre desarrollo de la personalidad (p. 6).

En esencia la dignidad de una persona tiene que ver con su libertad para decidir³.

Describir los marcos en los cuales esta libertad tiene cabida, dependerá en gran parte de la garantía de los derechos que especialmente el estado, por encima de las acciones que hagan las personas, pueda ofrecerles a sus habitantes. Lo anterior está en concordancia con el reconocimiento de las personas como seres con racionalidad y por ello, la dignidad humana hay que protegerla y defenderla.

Una de las características de los DDHH descritas por Carpizo (2011) tiene que ver con la historicidad, por eso es de relevancia mencionar las cuatro etapas esbozadas por el autor respecto a la aparición de los DDHH. Estas no han ocurrido de manera simultánea y están en constante evolución. La primera se inicia con la era moderna y la presencia de la burguesía, creándose diversas declaraciones como las americanas y francesas del siglo XVIII, con las que se precisaron derechos civiles y políticos de carácter individualista y liberal. La segunda ocurre antes, durante y después de la Primera Guerra Mundial, con la consagración de los derechos sociales y económicos. La tercera se origina poco antes, durante y especialmente después de la Segunda Guerra Mundial, como consecuencia de los horrores cometidos durante ese conflicto, y que impulsa la universalización e internacionalización de los derechos humanos, etapa en la cual nos encontramos y que aún dista mucho de haberse perfeccionado, aunque los avances son enormes. La cuarta se empalma en los últimos años con la anterior: la precisión de los derechos de solidaridad o de la tercera generación (Carpizo, 2011, p.18).

Durante la tercera etapa, la mayoría de los países del mundo logró juntarse en la Organización de las Naciones Unidas (ONU) para colaborar en la protección de los DDHH y

³ Hablamos de libertad para decidir en el sentido mencionado por el profesor F. López Quintás. Ver: [Libertad para decidir. - YouTube](#)

garantizar su universalización e internacionalización. De manera muy resumida, los países miembros (entre ellos Colombia como uno de los países fundadores) se comprometieron a intentar no repetir lo que sucedió en las guerras mundiales. Uno de los primeros documentos emitidos por la ONU y que representó un reconocimiento de los derechos de las personas, pero sin hacerlos explícitos, fue la Carta de San Francisco. Vargas et al. (2020) resaltan de este tratado, que marca inicio de

los procesos de internacionalización de los DDHH, de humanización del derecho internacional y de desarrollo progresivo de normas que tienen como característica fundamental resguardar a los individuos de la arbitrariedad y fijar precisos límites a la soberanía y potestad discrecional de los Estados (pág. 245).

Hoy en día, luego de 75 años de su adopción, es posible encontrar un robusto conjunto de “normas jurídicas no solo del orden universal, sino también desde los contextos regionales, que además de consagrar derechos, han avanzado hacia la creación de mecanismos de protección de diversa índole” (Vargas et al. 2020, p.245). En Carpizo (2011), es posible evidenciar diferentes ejemplos de cómo las constituciones de los países, incluida la de Colombia, concebida y adoptada en 1991, fueron incluyendo progresivamente el concepto de dignidad humana en sus escritos. Esto permitió darle un carácter jurídico a este concepto, al ser incorporado en los instrumentos internacionales para la garantía de los DDHH, así como a nivel local en cada país.

Desde la conformación de la ONU cada país miembro delega a un representante que tiene voz y voto en la asamblea general. Allí pueden existir consensos respecto a lo discutido y esto significa que cada gobierno deberá:

1) conducir sus esfuerzos para mantener la paz, la seguridad, la tolerancia y la convivencia pacífica; 2) abstenerse de usar la fuerza y solamente emplearla en beneficio del bien común; y 3) promover el progreso del orden económico y social para todas las naciones (Vargas et al. 2020, p.245).

Estas obligaciones internacionales, se incorporan en el ordenamiento interno de cada país en una figura denominada “bloque de constitucionalidad” (Vargas et al. 2020). En uno de los últimos eslabones, para el caso colombiano, estos instrumentos internacionales empiezan a cobrar sentido a través de las sentencias sucesivas de la Corte Constitucional.

Según Vargas et al. (2020), Colombia ha ratificado más de medio centenar de tratados relacionados con el ámbito global o regional, los cuales incluyen la vigilancia de diferentes órganos facultados para monitorear el cumplimiento de dichos tratados. Sin embargo, estos autores concluyen que “si bien los esfuerzos realizados por el Estado colombiano para suscribir y ratificar tratados internacionales en materia de protección de los DDHH han sido importantes, aún resultan insuficientes” (p.257).

Esta carencia, por ejemplo, se presenta con más fuerza en la historia reciente del país, un evento relevante es el acuerdo de paz con las FARC. Infortunadamente y a pesar de las muchas cosas positivas de acabar un conflicto de más de 50 años, los instrumentos surgidos de las negociaciones dejaron de lado cosas como garantizar la dignidad humana de los excombatientes, que se acogieron a este acuerdo. Estos, desde la firma del acuerdo, han sido estigmatizados, asesinados, excluidos, etc. Por otro lado, la falta de compromiso del gobierno de Iván Duque con el acuerdo de paz no permitió que el estado llegará a las zonas abandonadas por las FARC, permitiendo la proliferación de bandas criminales que tomaron estos territorios y que son ahora las responsables de muchas violaciones a los DDHH.

Estás fallas en la correcta implementación de los acuerdos relacionados con la garantía de la dignidad humana derivó según Vargas et al. (2020), en que “las víctimas surgieron como un actor protagónico en el ‘nuevo clima de justicia’ de las sociedades contemporáneas y sus experiencias de sufrimiento se convirtieron en un imperativo moral” (p. 259). La Comisión de la Verdad, por ejemplo, como uno de los organismos extrajudiciales surgidos del acuerdo de paz con las FARC, se encargó de acumular relatos de las víctimas para poder construir sobre esta concepción: ayudar a visibilizar a quienes realmente se vieron afectados por el conflicto. El informe final de esta comisión termina siendo algo que apoya la dignidad de las víctimas y permite no dejar en el olvido sus historias de vida.

Una de las fuentes que registra diferentes violaciones a los DDHH que han ocurrido en Colombia es la revista Noche y Niebla, de la cual se hablará en los apartados siguientes.

3.2 Revista Noche y Niebla

Como se describe en el más reciente Marco conceptual de la revista Noche y Niebla (Cinep/ppp, 2016), desde 1979 varios sectores democráticos daban cuenta del deterioro de los derechos fundamentales, y en vista de que nadie estaba procesando esa información y no se tenía una visión conjunta de lo que ocurría en Colombia, se concibe la idea de sistematizar los datos. Un motivo principal para esta labor resalta el Cinep/ppp (2016), tiene que ver con los altos niveles de violencia en Colombia, que se han presentado desde hace años.

Por esta razón, luego de ser creada la Comisión Intercongregacional de Justicia y Paz en 1987, se diseñó junto con el Centro de Investigación y Educación Popular (CINEP) un Banco de Datos de Derechos Humanos. Esto con la intención de responder a la necesidad del campo de la información y de la denuncia. Así, en 1988 comenzó a editarse el boletín trimestral Justicia y Paz, como un recurso para divulgar y denunciar los atentados más graves contra la vida

cometidos en Colombia. Este boletín se apoyaba en una lectura sociológica de los hechos, porque recogía y sistematizaba toda la información dispersa en medios de comunicación. Además, se recolectaban denuncias directas de organizaciones y entornos sociales de las víctimas. De esta forma se generaba un primer aviso humanitario, que podría ayudar en el seguimiento e investigación de las denuncias.

Como en el país muchas de estas denuncias quedaban en la impunidad, dentro del mismo Banco de Datos surgió el interés de hacer una evaluación más teórica de los hechos que ocurrieron y utilizar los documentos jurídicos existentes. No fue hasta 1996 que se logró elaborar un marco conceptual más técnico que empezó una nueva etapa con una publicación llamada *Noche y Niebla*. Este Marco Conceptual integró el deseo de una mayor rigurosidad jurídica y, por otra parte, el deseo de no desperdiciar información limitada que podría llegar a revelar indicios de móviles políticos en los actos de violencia, ocultos en las sombras.

Para lograr este Marco Conceptual se utilizó el instrumental del “Derecho Internacional de los Derechos Humanos y del Derecho Internacional Humanitario, como normatividad internacional adoptada por la Organización de Naciones Unidas” (Cinep/ppp, 2016). Además, abrieron un buzón de “Violencia Política y Social” donde se consignan los casos que no tienen evidencias que señalen a alguna autoría determinada, pero si con algunas pruebas que permitan detectar, además de presuntos responsables, unos posibles móviles políticos.

Hasta el momento la revista *Noche y Niebla* ha realizado 65 publicaciones, compilando relatos y datos de violaciones a los DDHH con dos ediciones al año. La edición de interés para este trabajo de grado es la número 62 (Cinep/ppp, 2021), de la cual es importante mencionar que su enfoque es el *anonimato como factor dinamizador del genocidio impune*. Muchas de las narraciones que conforman los datos, tienen que ver con actos de violencia realizados por

personajes no identificados y encapuchados, especialmente apuntando a estigmatizar a líderes y lideresas que trabajan para mejorar las condiciones de vida de las comunidades.

El anonimato termina revelando móviles políticos, que intentan conservar el estado de precariedad en el cual vive gran parte de la población colombiana, a costa de cuidar los intereses de la élite imperante. Según el análisis realizado en la presentación de la revista, este problema ha alcanzado su clímax en los últimos años y especialmente con el gobierno de Iván Duque. También concluyen que el imaginario ahora es que “existe UN MISMO LIBRETO que se activa intermitentemente para eliminar grupos humanos que, en opinión de la clase dirigente, NO CABEN EN ESTA NACIÓN” (Cinep/ppp, 2021).

3.3 Violaciones a los DDHH y presuntos responsables

En el marco conceptual de la revista Noche y Niebla se presenta un glosario que brinda definiciones precisas para cada uno de los términos o categorías utilizadas, en los relatos de los actos de violencia registrados en la revista. Estas permiten clasificar las diferentes violaciones a los Derechos Humanos.

En lo que sigue daremos cuenta de las principales: *Violencia Política como Violación de Derechos Humanos (DH)*, *Violencia Político-Social (VPS)*, *Infracciones Graves al Derecho Internacional Humanitario Consuetudinario (DIHC)* y *Acciones bélicas o Infracciones graves al Derecho Internacional Humanitario (DIH)*.

3.3.1 Violencia Política como Violación de Derechos Humanos (DH)

Para la Defensoría del Pueblo, la violación de DDHH debe entenderse como: “conducta positiva o negativa mediante la cual un agente directo o indirecto del Estado vulnera, en cualquier persona y en cualquier tiempo, uno de los derechos enunciados y reconocidos por los

instrumentos que conforman el Derecho Internacional de los DDHH” (Cinep/PPP, 2016, pág. 15).

De esta forma, se resalta la importancia de dos elementos para que un acto de violencia sea considerado como violación a los DDHH: el autor y la materia. Si el derecho vulnerado es alguno de los consagrados en los pactos internacionales de DDHH y si el autor está relacionado directa o indirectamente con el estado, entonces se considerará como una violación de DDHH.

La autoría estatal puede darse de varias maneras, se mencionarán dos: por un lado, cuando el acto de violencia es realizado por un agente del estado (que ejerce una función política), y por otro cuando el acto de violencia es realizado por particulares que son apoyados o tolerados por el Estado.

Algunos de los móviles que pueden llegar a originar violaciones a los DDHH son: la persecución política (PP), la cual ocurre en actividades -que pueden ser encubiertas o no- para mantener el orden público o en la represión de ideologías que critiquen al Estado; el abuso o exceso de autoridad (AA), que tiene que ver con el uso de la fuerza desmedida con el fin de cumplir las funciones de agentes del Estado; y la manifestación de intolerancia social (IS), la cual se presenta cuando el acto de violencia está dirigido a eliminar personas que son consideradas por sus victimarios como problemáticas para la sociedad.

Estos móviles se dividen en tres modalidades específicas de violación de los Derechos Humanos: violación del derecho a la vida (VDV), violación del derecho a la integridad personal (VIP) y violación del derecho a la libertad personal (VDLP), estas modalidades tienen actos de violencia específicos, los cuales la revista Noche y Niebla clasifica por medio de códigos a partir

de letras y números. Cada uno de estos códigos es utilizado para identificar el crimen, el móvil y la modalidad.

La modalidad específica de VDV se compone de las siguientes acciones:

- Ejecución extrajudicial, con los códigos A10 (PP), A20 (AA), A30 (IS)
- Atentado, con los códigos A16 (PP), A26 (AA), A37 (IS)
- Amenaza individual, con los códigos A15 (PP), A25 (AA), A35 (IS)
- Amenaza colectiva, con los códigos A18 (PP), A28 (AA), A38 (IS)

La modalidad VIP comprende las siguientes acciones de violencia:

- Tortura, con los códigos A12 (PP), A22 (AA), A36 (IS)
- Lesión física, con los códigos A13 (PP), A23 (AA), A33 (IS)
- Colectivo lesionado, con los códigos A17 (PP), A231 (AA), A331 (IS)
- Violencia sexual, con los códigos A19 (PP), A29 (AA), A39 (IS), esta se subcategoriza en:
 - Violación, con los códigos A191 (PP), A291 (AA), A391 (IS)
 - Embarazo forzado, con los códigos A192 (PP), A292 (AA), A392 (IS)
 - Prostitución forzada, con los códigos A193 (PP), A293 (AA), A393 (IS)
 - Esterilización forzada, con los códigos A194 (PP), A294 (AA), A394 (IS)
 - Aborto forzado, con los códigos A197 (PP), A297 (AA), A397 (IS)
 - Esclavitud sexual, con los códigos A195 (PP), A295 (AA), A395 (IS)
 - Abuso sexual, con los códigos A196 (PP), A296 (AA), A396 (IS)

Para la modalidad VDLP se consideran las siguientes acciones de violencia:

- Desaparición forzada e involuntaria, con los códigos A11 (PP), A21 (AA), A302 (IS)
- Detención arbitraria, con los códigos A14 (PP), A24 (AA), A301 (IS)
- Judicialización arbitraria, con los códigos A141 (PP), A241 (AA), A341 (IS)
- Deportación, con el código A101(PP)
- Desplazamiento forzado colectivo, con los códigos A102 (PP), A27 (AA), A34 (IS)
- Confinamiento como represalia o castigo colectivo, con el código A104 (PP)

3.3.2 Violencia Político-Social (VPS)

Este tipo de violencia es aquella que es ejercida por organizaciones o grupos motivados por el poder político o la intolerancia frente a ideologías, razas, religiones o sectores sociales. También, se considera como Violencia Política Social a las prácticas de actores armados no estatales que no pueden ser clasificados como violaciones a los Derechos Humanos.

En estas acciones no se identifica como responsable a un autor estatal o para-estatal; aquí se identifica la diferencia con el DH, debido a que no se puede clasificar como violación a los DDHH. En mayor medida los perpetradores están amparados por una clandestinidad que los hace difíciles de identificar, en otros casos se identifican como autores particulares sin vínculos con el estado.

Es importante aclarar que en esta categoría de violencia no se incluyen a las acciones de grupos insurgentes, exceptuando los casos de secuestro y de prácticas de intolerancia social. Las acciones violentas de los grupos insurgentes son tipificadas como infracción al Derecho Internacional Humanitario Consuetudinario.

Los móviles de la violencia político-social se dividen en dos: por persecución política (PP) y por intolerancia social (IS). Respecto a los móviles políticos “pueden percibirse en el análisis de las razones que impulsaron a los perpetradores del crimen y dichas razones pueden revelarse a través de las características de la víctima” (Cinep/PPP, 2016, p.27). En cuanto a los móviles de intolerancia se observan cuando la víctima hace parte de “sectores sociales considerados por sus victimarios como disfuncionales o problemáticos para la sociedad” (Cinep/PPP, 2016, p.27).

Acorde con estos móviles las acciones de VPS se pueden clasificar en tres diferentes modalidades específicas de violencia: contra la vida, contra la integridad personal y contra la libertad personal. Como en el DH, cada uno de estos actos de violencia tiene un código, que designa el banco de datos en su marco teórico utilizado para identificar el crimen, el móvil y la modalidad.

Para la modalidad específica de violencia contra la vida se tienen las siguientes acciones de violencia, con su respectivo código:

- Asesinato, con los códigos B40 (PP), B50 (IS)
- Atentado, con los códigos B46 (PP), B57 (IS)
- Amenaza individual, con los códigos B45 (PP), B55 (IS)
- Amenaza colectiva, con los códigos B49 (PP), B59 (IS)

Para la modalidad específico de violencia contra la integridad personal se tienen los siguientes actos de violencia, con su respectivo código:

- Tortura, con los códigos B47 (PP), B56 (IS)
- Lesión física, con los códigos B43 (PP), B53 (IS)

- Colectivo lesionado, con los códigos B402 (PP), B502 (IS)
- Violencia sexual por persecución política, con los códigos B420 (PP), B520 (IS)

Las siguientes violaciones a los derechos humanos también se encuentra en DH, pero aquí cambia el móvil y el autor:

- Violación, con los códigos B421 (PP), B521 (IS)
- Embarazo forzado, con los códigos B422 (PP), B522 (IS)
- Prostitución forzada, con los códigos B423 (PP), B523 (IS)
- Esterilización forzada, con los códigos B424 (PP), B524 (IS)
- Aborto forzado, con los códigos B427 (PP), B527 (IS)
- Esclavitud sexual, con los códigos B425 (PP), B525 (IS)
- Abuso sexual, con los códigos B426 (PP), B526 (IS)

Para el móvil específico de violencia contra la integridad personal se tienen los siguientes actos de violencia, con su respectivo código que lo identifica:

- Secuestro perpetrado por organizaciones insurgentes, con el siguiente código B41 (PP)
- Rapto, con los códigos B48 (PP), B58 (IS)
- Desplazamiento colectivo, con los códigos B401 (PP), B501 (IS)

3.3.3 Infracciones Graves al Derecho Internacional Humanitario Consuetudinario (DIHC)

La violencia que se genera debido a los conflictos armados, que puede ser entre dos estados o entre fuerzas armadas en el interior del estado, debe ser clasificada de otra manera, diferente a las vistas anteriormente. Esto se debe a que un conflicto armado puede privar de la

vida o la libertad a los contrincantes, generando afectaciones a la integridad de los partícipes del estado y destrucción en bienes públicos.

La edición de la revista estudiada aclara que el Derecho Internacional Humanitario no entra a calificar la legitimidad de la guerra o los métodos del conflicto en relación con su efectividad. Lo único que se proponen los DIH es evitar el sufrimiento y daños innecesarios, aquellas acciones violentas que no tienen como objetivo una ventaja militar sobre el adversario.

Los DIH están basados en acuerdos o protocolos surgidos de los convenios de la Haya de 1899; los cuatro convenios de Ginebra de 1949; y dos protocolos adicionados en 1977. Estas declaraciones internacionales son las que restringen algunas acciones cometidas en la guerra. Sin embargo, pese a que toda esta normatividad fue asumida por una gran cantidad de Estados, las infracciones a estas normas continuaron siendo numerosas.

Por lo tanto, colectivamente se generó una conciencia internacional sobre los fallos que hubo en la formulación de los DIH. Esto principalmente en los protocolos, que no fueron ratificados en varios estados. Existían dos protocolos, el primero con una normatividad más detallada, aplicaba para conflictos internacionales y el segundo con una normatividad precaria, aplicaba a conflictos no internacionales. Esto fue contraproducente, dado que la mayoría de los conflictos en la actualidad son de carácter no internacional.

Debido a los hechos descritos, se realizó una conferencia internacional en pro de una mejor protección para las víctimas de la guerra en 1993. En esta se convocó a un grupo de expertos para identificar las normas aplicables en la guerra que tuvieran un carácter de derecho consuetudinario.

De acuerdo con la edición 62 de la revista Noche y Niebla “el Derecho Consuetudinario es un derecho reconocido por la costumbre universal, con sustento en la más amplia gama de legislaciones, sentencias de tribunales internacionales, opiniones de los más prestigiosos juristas” (Cinep/ppp, 2021, p.8). Esto quiere decir que se impone por su fuerza jurídica y son aceptadas universalmente como derecho. Esta publicación de normas fue realizada en 2005 contando con un catálogo de 161 normas.

Es cierto que esa formulación de normas para los DIHC, al elaborarse sobre conflictos entre fuerzas equilibradas, no es útil para registrar infracciones donde haya un modelo de Guerra de Guerrillas o Guerra Sucia. En la mayoría de los casos los conflictos suelen ser entre un grupo estatal y un movimiento de oposición surgido de las capas sociales más excluidas o perseguidas; por ende, estos conflictos no suelen apegarse a las características de una guerra regular.

Desde los años 60 las estrategias contrainsurgentes del estado colombiano han sido documentadas y revelan que están diseñadas principalmente contra la población civil, que debido a “ideologías o arraigos territoriales pudiera albergar niveles de simpatía hacia propuestas de cambio del sistema económico-político” (Cinep/ppp, 2021, p. 9). Al involucrar en el combate armado a la población civil se dificulta la identificación de las personas protegidas o de los no combatientes según los parámetros del DIHC y del DIH.

Otra consideración de importancia, desde los parámetros del Derecho Humanitario, es la dificultad para la lectura del conflicto colombiano debido al desarrollo progresivo del paramilitarismo. Esto se puede evidenciar en los análisis estadísticos puesto que “en las últimas décadas, las curvas de responsabilidad de la fuerza pública en violaciones graves a los derechos humanos y en crímenes de guerra solo descienden en la medida en que ascienden las de responsabilidad de los paramilitares” (Cinep/ppp, 2021, p. 9).

Las infracciones a los DIHC en la revista Noche y Niebla se clasifican en cuatro clases: por atacar objetivos ilícitos de guerra, por el empleo de métodos ilícitos de guerra, y por el trato afrentoso al ser humano. Cada una de estas infracciones tienen un código que lo identifica y son asignados por el banco de datos.

Para las infracciones por atacar objetivos ilícitos de guerra (objetivos no militares), se tienen las siguientes acciones:

- El ataque indiscriminado, con el código D90
- Atacar o impedir misión médica o sanitaria, con el código D707
- Atacar o impedir misión religiosa, con el código D708
- Atacar o impedir misión humanitaria, con el código D709
- Atacar o impedir misiones de paz, con el código D710
- Atacar o impedir misión informativa, con el código D711
- Ataque a zonas humanitarias, con el código D712
- Ataque a bienes culturales, con el código D85
- Ataque a obras o instalaciones que contienen fuerzas peligrosas, con el código D801
- Ataque al medio ambiente natural, con el código D84

Para las infracciones por el empleo de métodos ilícitos de guerra se tienen las siguientes acciones de violencia:

- Asumir el método de “Guerra sin cuartel”, con el código D905
- Recurrir al pillaje, con el código D95
- Utilizar el hambre como método de guerra, con el código D86

- Utilizar la perfidia, con el código D91
- Impedir conversaciones de paz a través de mediadores, con el código D713

Para las infracciones por el empleo de medios ilícitos de guerra se tienen los siguientes actos realizados en la guerra:

- Empleo de armas absolutamente prohibidas, con el código D92. Algunas de ellas son:
 - Armas envenenadas
 - Armas biológicas
 - Armas químicas
 - Herbicidas como método de guerra
 - Balas que se expande o que explotan
 - Armas nucleares
- Empleo ilícito de armas de uso restringido, con el código D93. Algunas de ellas son:
 - Minas terrestres
 - Armas incendiarias

Para las infracciones por el trato afrentoso al ser humano se tienen las siguientes acciones:

- Homicidio internacional de persona protegida, con el código D701
- Muerte causada por empleo de métodos y medios ilícitos de guerra, con el código D97
- Muerte causada por ataque a bienes civiles, con el código D87
- Muerte de civil en acción bélica, con el código D703
- Lesiones intencionales a la integridad personal de persona protegida, con el código D702

- Lesiones a la integridad personal de persona protegida por empleo de métodos o medios ilícitos de guerra, con el código D98
- Lesiones a la integridad personal de persona protegida como consecuencia de ataques a bienes de carácter civil, con el código D88
- Lesiones a la integridad personal de civiles como consecuencia de una acción bélica, con el código D704
- Colectivo lesionado como consecuencia de otras infracciones al DIHC, con el código D705
- Tortura y tratos crueles e inhumanos, atentados contra la dignidad personal, tratos humillantes y degradantes y castigos corporales, como instrumentos de guerra, con el código D72
- Violencia sexual como instrumento de guerra, con el código D77. Hay diversas formas de violencia sexual que también cuentan con su propio código:
 - Violación (código D771)
 - Esclavitud sexual (código D775)
 - Prostitución forzada (código D773)
 - Embarazo forzado (código D772)
 - Esterilización forzada (código D774)
 - Aborto forzado (código D777)
 - Aborto sexual (código D776)
- Esclavitud y trabajos forzados como instrumentos de guerra, con el código D714
- Toma de rehenes como instrumento de guerra, con el código D74
- Uso de escudos humanos como instrumento de guerra, con los códigos D78, D904)

- Desaparición forzada como instrumento de guerra, con el código D76
- Judicialización arbitraria como instrumento de guerra, con el código D15
- Desplazamiento forzado como instrumento de guerra, con el código D903
- Reclutamiento de menores y su utilización en la guerra, con el código D75

3.3.4 Acciones béticas o Infracciones graves al Derecho Internacional Humanitario (DIH)

El Banco de Datos de Noche y Niebla utiliza el nombre de Acciones Béticas para clasificar los “actos ejecutados por los actores del conflicto armado de carácter no internacional que se da en el país, y que por acomodarse a las normas del ‘ius in Bello’ (Derecho en la Guerra) son acciones legítimas de guerra” (Cinep/ppp, 2021, p. 10).

Pero no son pocos los problemas que surgen a la hora de interpretar y hacer una clasificación de las acciones béticas. Por ejemplo, hay que tener cuidado de no atribuir a los dos polos armados, una infracción que tiene como sujeto activo de la acción a un solo polo (Cinep/ppp, 2016).

Por lo tanto, para el caso del Banco de Datos de la revista, se deben contar tantas acciones béticas que no resulten contrarias a los criterios que establece el derecho penal. En las acciones béticas hay modalidades específicas:

- Combate, con el código C62
- Emboscada, con el código C63
- Uso de minas-minado de un campo, con el código C64
- Bombardeo-ametrallamiento, con el código C65
- Bloqueo de vías, con el código C66

- Ataque a objetivo militar, con el código C67
- Sabotaje, con el código C68

3.4 Datos de base revista Noche y Niebla

Con base en los datos reportados en la edición 62 de la revista Noche y Niebla, a continuación, se describirán: el tratamiento de la información y las decisiones acerca de la sistematización de estos datos. Conviene distinguir que esto colaboró a generar una matriz de datos acorde con los lineamientos que se requieren para hacer un análisis multivariado.

Primero, se debe tener en cuenta que en la sección 3.3, se trajeron cuatro grandes categorías de violaciones a los DDHH y se usaron siglas para cada una, por ejemplo, Violencia Político-Social como VPS. En esta sección delante de cada palabra que se refiere al acto delictivo, se pone la sigla correspondiente según lo convenido en la sección 3.3. Por ejemplo, Muertos DH, hace referencia a casos en los cuales se clarifican muertes cometidas en las condiciones descritas en la sección 3.3.1 para Violaciones a los derechos humanos (DH).

3.4.1 Categorías dadas por la revista

Al final de la edición número 62 de la revista (y de cualquier entrega), figuran tablas que representan el compendio de cada uno de los casos de violaciones a los DDHH. Cada tabla tiene la fecha, ubicación, nombre de la víctima, categoría de la victimización y presunto responsable. Las últimas dos categorías son las de interés, y se clasifican usando letras y números, como se puede ver en la parte superior de la Tabla 1, que representa un fragmento de una tabla más grande:

Tabla 1. Fragmento de una tabla de la revista Noche y Niebla

Categorías de victimización																																					
A (MUERTOS-DH): 30 20 30 ; B (MUERTOS-DH): 87 97 701 203 ; C (MUERTOS-VPS): 40 50 ; D (DESAPARICIÓN-DH): 11 21 302 ; E (DESAPARICIÓN-DH): 76 ; F (TORTURA-DH): 12 22 36 ; G (TORTURA-DH): 72 ; H (TORTURA-VPS): 47 56 ; I (LESIONADOS-DH): 13 21 35 ; J (LESIONADOS-VPS): 43 53 ; L (DETENCIÓN ARBITRARIA-DH): 14 24 301 ; M (AMENAZA-DH): 15 25 35 ; N (AMENAZA-VPS): 45 55 ; O (ATENTADO-DH): 12 13 14 ; P (ATENTADO-VPS): 46 57 ; Q (DUDICLACION ARBITRARIA-DH): 141 241 341 ; R (DUDICLACION ARBITRARIA-DH): 715 ; S (VIOLENCIA SEXUAL-DH): 19 29 39 ; T (VIOLENCIA SEXUAL-DH): 74 ; U (VIOLENCIA SEXUAL-VPS): 40 50 ; V (DEPORTACIÓN-DH): 101 ; W (SECUESTRO-VPS): 41 ; X (RAPTO-VPS): 48 58 ; Y (TOMA DE REHENES-DH): 74 ; Z (ESCLAVITUD Y TRABAJOS FORZADOS-DH): 734 ; AA (NEGACIÓN DE DERECHOS A PRISIONEROS DE GUERRA-DH): 756 ; AB (NEGACIÓN DE ATENCIÓN A PERSONAS VULNERABLES-DH): 717 ; AC (PROFANACIÓN Y OCULTAMIENTO DE CADABRES-DH): 73 ; AD (RECLUTAMIENTO DE MENORES-DH): 79 ; AE (ENCUENTRO INDIVIDUAL-DH): 78																																					
Presuntos responsables																																					
1: ESTADO COLOMBIANO; 2: FUERZA PÚBLICA; 3: FUERZAS MILITARES; 4: EJERCITO; 5: ARMADA; 6: FUERZA AÉREA; 7: POLICIA; 8: GUÍA; 9: SUB; 10: DEBH; 11: INPEC; 12: DAS; 13: CTI; 14: PAMI; 15: AUC; 16: ACCU; 17: AUTODEFENSAS DE RAMÓN ISAZA; 18: AUTODEFENSAS DE LOS LLANOS ORIENTALES; 19: AUTODEFENSAS DE SANTANDER Y SUR DEL CESAR; 20: AUTODEFENSAS DE CASANARE; 21: AUTODEFENSAS CAMPESINAS DEL MAGDALENA MEDICO; 22: AGENTE EXTRANJERO; 23: GUERRILLA; 24: MILICIAS; 25: FARC-EP; 26: ELN; 27: MOVIMIENTO JAIME BATEMAN CAYON; 28: ERP; 29: EPL; 30: ERG; 31: GRUPOS DE INTOLERANCIA; 32: SIN INFORMACIÓN; 33: OTROS; 34: COMBATIENTES; 35: FISCALIA; 36: POLO INSURGENTE; 37: AUROMAR S.A.; 38: AMERISUR EXPLORACIÓN COLOMBIA; 39: EMPRESAS; 40: PABEL RESOURCES																																					
Fecha	Ubicación	Víctima	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	IP			
08/ago/2020	NARIÑO - CUMBAL	WILLIAM ARENEJO GUADIR TARAPIQUES																																		35	
13/ago/2020	NARIÑO - IPiales	WILLIAM ALEXANDER DOMINGUEZ																																			7
23/jul/2020	CAUCA - SUCRE	WILLIAM DE JESÚS DIAS MORALES																																			35
25/ago/2020	MURTE DE SANTANDER - ABREGO	WILMAR CARREÑO																																			35
04/sep/2020	PUTUMAYO - PUERTO ASÍS	WILMAR MACRONERO																																			1-34

Fuente: Revista Noche y Niebla

Todas las tablas disponibles en esta edición fueron transferidas a un archivo de Excel, en el cual fue posible dejar todas las columnas con un filtro. Para facilitar la lectura de los datos, se llenó con 0 las casillas vacías de manera que el sistema interpretara que la víctima no corresponde a esa categoría de victimización. Los 1 se conservaron del documento original, indicando la presencia de la violación de DDHH. Por ejemplo, se puede ver en la Tabla 2, la aplicación de un filtro a la columna D, con 8 casos, la cual corresponde a la categoría “Desaparición DH”:

Tabla 2. Fragmento de datos en Excel

Fuente: elaboración propia

Las columnas relevantes para abordar el estudio de la información contenida con base en un análisis multivariado fueron las categorías de victimización (columnas A hasta AD en la Tabla 2) y presunto responsable (última columna de la Tabla 2: “PRTodos”). Realizando filtrados consecutivos como el que se ilustra en la Tabla 2, llegamos a la siguiente tabla de contingencia (ver Tabla 3) que relaciona las variables mencionadas. Como se pudo notar en los ejemplos de la revista y del archivo de Excel, hay casos los cuales tienen más de un presunto responsable, es decir, más de un número en la columna “PRTodos”. A esto se le prestó atención en la Tabla 3, representándolo en las últimas tres columnas (de izquierda a derecha). Primero, la columna “Total fila” es la suma de cada fila, es decir el total de presuntos responsables por cada categoría de victimización, la cual tiene en cuenta los casos de víctimas que reportan más de un presunto responsable. Segundo, la columna “sum real” que es la suma de la cantidad real de víctimas que hay por categoría de victimización, ya que las víctimas con más de un presunto responsable inflan de manera artificial la frecuencia descrita en “Total fila”. Para corregir este problema, se hace un filtrado de cada columna del archivo de Excel anterior, para obtener solo las filas con frecuencia 1. Finalmente, la columna “Dif”, es la que representa la cantidad de

coincidencias por categoría de victimización, en las cuales la responsabilidad del suceso se le atribuye a más de un presunto responsable. Esta columna de diferencias obliga a realizar una revisión más detallada de los casos de violaciones a los DDHH, como se explicará en la siguiente sección.

Tabla 3. Tabla de contingencia, primera versión

	PR1	PR2	PR3	PR4	PR7	PR9	PR11	PR14	PR25	PR28	PR33	PR35	PR37	PR38	Total fila	sum real	Dif.	
A Muertos-DH	0	0	0	14	41	0	0	59	0	0	0	0	0	0	114	102	12	
B Muertos-DIH	0	0	0	0	0	0	0	39	10	0	0	0	10	0	59	58	1	
C Muertos-VPS	0	0	0	0	0	0	0	0	0	0	0	12	224	0	236	236	0	
D DESAPARICION-DH	1	0	1	4	0	0	0	6	0	0	0	0	0	0	12	8	4	
E DESAPARICION-DIHC	0	0	0	0	0	0	0	12	0	0	0	0	0	0	12	12	0	
F TORTURA-DH	0	0	0	12	17	0	0	14	0	0	0	0	0	0	48	43	2	
G TORTURA-DIHC	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	
H TORTURA-VPS	0	0	0	0	0	0	0	0	0	0	1	10	0	0	11	11	0	
I LESIONADOS-DH	0	6	0	13	74	1	0	6	0	0	0	0	0	0	180	95	5	
J LESIONADOS-DIHC	0	0	0	0	0	0	0	16	1	0	0	0	22	0	39	35	4	
K LESIONADOS-VPS	0	0	0	0	0	0	0	0	0	0	0	2	48	0	46	46	0	
L DILEXION ARBITRARIA-DH	2	1	0	19	27	0	0	2	0	0	0	0	0	4	75	70	5	
M AMENAZA-DH	40	1	0	54	26	0	1	251	0	0	0	0	0	1	374	328	46	
N AMENAZA-DIHC	5	0	0	9	8	0	0	0	70	63	7	1	0	0	154	148	5	
Ñ AMENAZA-VPS	0	0	0	0	0	0	0	0	0	0	0	12	31	0	103	103	0	
O ATENTADO-DH	0	0	0	2	1	0	0	1	0	0	0	0	0	0	4	4	0	
P ATENTADO-VPS	0	0	0	0	0	0	0	0	0	0	0	0	12	0	12	12	0	
S VIOLENCIA SEXUAL-DH	0	0	0	1	7	0	0	1	0	0	0	0	0	0	9	7	2	
U VIOLENCIA SEXUAL-VPS	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	
W SECUESTRO-VPS	0	0	0	0	0	0	0	0	7	0	3	0	0	0	10	10	0	
X RAPTO-VPS	0	0	0	0	0	0	0	0	0	0	1	14	0	0	15	15	0	
AD RECLUTAMIENTO DE MENORES-DIHC	0	0	0	0	0	0	0	4	0	0	0	0	1	0	5	4	1	
Total columna																1436	1350	86

Fuente: elaboración propia

3.4.2 Consideraciones acerca de los presuntos responsables de la Tabla 3

Nótese que la Tabla 3 no es realmente una tabla de contingencia, debido a las repeticiones que hay de presuntos responsables. Si se deja en ese estado, no sería de utilidad para llevar a cabo un Análisis de Correspondencias simple. Por este motivo, fue necesario realizar lecturas de los diferentes relatos en los cuales había coincidencias (que se pueden identificar en la columna “Dif” de la Tabla 3 en donde hay valores mayores que cero). Esto con el propósito de decidir, en concordancia con la descripción de cada hecho, quién de los presuntos responsables tendría mayor responsabilidad en el acto vulnerador de los DDHH y así dejar solo uno. A continuación, describimos las decisiones para cada una de las categorías de victimización que presentan más de un presunto responsable en la Tabla 3.

Muertos DH. Para esta violación, en la Tabla 4 se pueden identificar 3 presuntos responsables en las últimas dos columnas: Ejército (PR4) Policía (PR7) y Paramilitares (PR14). Ya sean estos tres o solo dos, están en los 8 registros. La elección del victimario en las primeras cinco filas fue para PR4. Con base en los relatos fue posible identificar a PR4 como el mayor responsable, excluyendo a los cómplices PR7 y PR14 que no son mencionados directamente en las narraciones de las víctimas. En las últimas tres filas, por los relatos que los responsabilizan, ahora se elige directamente a PR14; en donde PR7 y PR4, se excluyeron en estos tres casos porque participaban, por ejemplo, con la omisión o la falta de atención a las víctimas, pero no cometían directamente la acción de violencia.

Tabla 4. Más de un PR en Muertos DH

Fecha	Ubicación	Víctima	A	PRTodos
25-novi-20	PUTUMAYO - PUERTO ASÍS	FERNANDO MEJIA ANGARITA	1	4; 7; 14
13-agosto-20	CAUCA - CDRINTO	JOHAN ERNESTO RIVERA	1	4; 7
27-oct-20	AMAZONAS - LETICIA	N N	1	4; 7
27-oct-20	AMAZONAS - LETICIA	N N	1	4; 7
27-oct-20	AMAZONAS - LETICIA	N N	1	4; 7
20-oct-20	CHOCÓ - CARMEN DEL DARIÉN	OSWALDO RODRÍGUEZ	1	4; 7; 14
13-oct-20	PUTUMAYO - PUERTO ASÍS	PEDRO ABEL CAÑAS	1	4; 7; 14
12-oct-20	PUTUMAYO - PUERTO ASÍS	WILFRIDO JIMÉNEZ	1	4; 7; 14

Fuente: elaboración propia

Muertos DIH. En el caso de esta violación solo hay una incongruencia y aparece AGROMAR (PR37) por primera vez como presunto responsable; PR37 no se tiene en cuenta, porque en el relato se identifica a las FARC-EP (PR25) como los victimarios.

Muertos VPS. Aunque no había más de un presunto responsable en los datos que se atribuyen a esta violación, se identificó una cantidad elevada de registros (ver Tabla 3). Una

explicación para esto, luego de la lectura de los múltiples relatos, tiene que ver con que la presencia de la institucionalidad es exigua. En las narraciones que responsabilizan a la Fiscalía (PR35) y otros (PR33), puede pensarse en términos de VPS, porque las organizaciones al margen de la ley delinquen a sus anchas en los territorios y esto provoca ausencia de denuncias, procesos judiciales, impunidad para los criminales, etc. De esta manera no se logra identificar quién de los presuntos responsables comete el asesinato y se asigna la responsabilidad a la Fiscalía, por su omisión en estos casos.

Desaparición DH. En este tipo de violación, en el único caso de interés, en la Tabla 5, no se atribuye responsabilidad al Estado Colombiano (PR1) ni a las Fuerzas Militares (PR3), porque la desaparición es cometida por los Paramilitares (PR14), según el relato.

Tabla 5. Más de un PR en Desaparición DH

Fecha	Ubicación	Víctima	D	PRTodo
6-sep-20	ANTIOQUIA - MURINDÓ	LUIS HERNANDO RODRIGUEZ DELGADO	1	3; 1; 14

Fuente: elaboración propia

Tortura DH. Como en el caso anterior, para esta violación solo hay un registro a revisar, considerando como presuntos responsables a los Paramilitares (PR14), el Ejercito (PR4) y Policía (PR7). En la narración se identifica la responsabilidad directa de PR14; PR4 y PR7 son cómplices por omisión.

Lesionados DH. Para los cinco registros de esta violación, en la Tabla 6, se observa que para las dos primeras filas (que tienen el mismo lugar y fecha) se deja al Ejército (PR4) como presunto responsable, porque, de acuerdo con los hechos descritos, accionaron las armas. En la

tercera fila, se deja a los Paramilitares (PR14), debido a que la Policía (PR7) fue cómplice de acuerdo con el relato, pero no lesionaron directamente a la víctima. En las dos últimas filas, también con la misma fecha, la Policía (PR7) estaba al frente del operativo según lo narrado. El Ejército (PR4) estaba apoyando, por ende, se dejó a PR7 como los presuntos responsables principales.

Tabla 6. Más de un PR en Lesionados DH

Fecha	Ubicación	Víctima	I	PRTodos
13-agosto-20	CAUCA - CORINTO	JULIO TUMBO	1	4; 7
13-agosto-20	CAUCA - CORINTO	LEONIDAS PERDOMO	1	4; 7
15-noviembre-20	ANTIOQUIA - APARTADÓ	N N	1	7; 14
27-octubre-20	PUTUMAYO - ORITO	N N	1	4; 7
27-octubre-20	PUTUMAYO - ORITO	N N	1	4; 7

Fuente: elaboración propia

Lesionados DIHC. En el caso de esta violación, en la Tabla 7 se pueden ver los cuatro casos de interés. Aparece AGROMAR (PR37) nuevamente, con FARC-EP (PR25) como presuntos responsables. En los relatos se pudo notar que PR37 aparece seguido como cómplice de PR25. Además, en estas lecturas, se identificó que varios casos en los cuales el presunto responsable es PR37, tienen que ver con víctimas de minas antipersonales. Este podría ser un error en la clasificación, debido a que no fue posible encontrar información fidedigna de qué es AGROMAR, pero debido a los relatos fue posible identificar que en realidad la responsabilidad finalmente recaería más sobre las disidencias de PR25.

Tabla 7. Más de un PR en Lesionados DIHC

Fecha	Ubicación	Víctima	J	PRTodos
18-sep-20	CAUCA - CALOTO	ALBERTO DARIO COICUE	1	37; 25
18-sep-20	CAUCA - CALOTO	CHELI SENOVIA	1	37; 25
18-sep-20	CAUCA - CALOTO	JHEIN GENI NOSCUE	1	37; 25
18-sep-20	CAUCA - CALOTO	JOHAN ANDRES MEDINA	1	37; 25

Fuente: elaboración propia

Lesionados VPS. Para esta violación no hay casos con más de un presunto responsable, pero es interesante notar que en los culpables se identifican a Fiscalía (PR35) y otros (PR33), lo cual indica la relación con la VPS en las categorías anteriores.

Detención arbitraria DH. En el caso de este tipo de violación, las filas 1, 2, 5 y 6 de la Tabla 8, en los que se tiene a policía (PR7) y AMERISUR (PR38), se deja PR7 porque de acuerdo con los relatos son quienes realizan la detención arbitraria. Con una consulta rápida se conoció que AMERISUR es una petrolera con presencia en varias zonas del país, pero en los relatos no había explicaciones respecto al involucramiento de esta última en las detenciones. Por otro lado, en las filas 3 y 4 de la Tabla 8, que tiene como presuntos responsables al Estado Colombiano (PR1) y a los Paramilitares (PR14), se deja PR14 porque los relatos revelaron que la culpa de PR1 es indirecta, debido a su falta de presencia en la zona, lo cual permite el actuar de PR14.

Tabla 8. Más de un PR en Detención arbitraria DH

Fecha	Ubicación	Víctima	L	PRTodos
15-dic-20	META - VILLAVICENCIO	ADELSO GALLO	1	7; 38
29-sep-20	CAUCA - POPAYÁN	FABIÁN ANDRÉS SÁNCHEZ CORREA	1	7; 38
8-nov-20	ANTIOQUIA - MUTATÁ	FERNANDO DOGARI BAILARIN	1	1; 14
8-nov-20	ANTIOQUIA - MUTATÁ	MILTON BAILARIN	1	1; 14
16-dic-20	NARIÑO - SAN PABLO	ROBERT DAZA GUEVARA	1	7; 38
15-dic-20	BOLÍVAR - BARRANCO DE LOBA	TEOFILO MANUEL ACUÑA RIBON	1	7; 38

Fuente: elaboración propia

Amenaza DH. En este tipo de violación primero, se reportan 32 casos en los cuales se identifican como presuntos responsables al Estado Colombiano (PR1) y a los Paramilitares (PR14). En estos se dejó a PR14, porque son quienes comenten esta violación a los DDHH. Con base en las narraciones se intuyó que estas amenazas, no se producirían si hubiera presencia de PR1 en los territorios, pero en estos casos solo actúa como cómplice.

En la Tabla 9 se pueden ver algunos de los casos particulares en los cuales, además de responsabilizar a PR1 y PR4, se añade a las Fuerzas Militares (PR3), el Ejército (PR4), la Policía (PR7). Se deja a PR14, porque en la lectura de los relatos se concluyó algo similar a lo descrito en el párrafo anterior: PR3, PR4, PR7 hacen parte de las posibilidades que el estado podría desplegar en los territorios, pero no hacen presencia. Aunque comparten algo de la culpa, quienes perpetran la violación de DDHH son PR14.

Tabla 9. Más de un PR en Amenazas DH

Fecha	Ubicación	Víctima	M.	PR Tdc
6-sep-20	ANTIOQUIA - MURINDÓ	ALDEMAR DOMÍCIO	1	3;1;14
10-nov-20	ANTIOQUIA - MUTATÁ	ARGIMIRO BAILARÍN	1	4;7;14
6-sep-20	ANTIOQUIA - MURINDÓ	CARLOS ALFREDO DELGADO	1	3;1;14
26-nov-20	BOGOTÁ, D.C. - BOGOTÁ, D.C.	HARRY ALEJANDRO GIL BRICEÑO	1	4;13
6-sep-20	ANTIOQUIA - MURINDÓ	LUIS HERNANDO RODRIGUEZ DELGADO	1	3;1;14
19-jul-20	CHOCÓ - CARMEN DEL DARIÉN	LUIS SINI GUI	1	4;1;14
27-oct-20	CHOCÓ - CARMEN DEL DARIÉN	LUIS SINI GUI	1	4;1;14
24-jul-20	ANTIOQUIA - APARTADÓ	N N	1	4;7;1;14
10-oct-20	PUTUMAYO - PUERTO ASÍS	SANDRA LAGOS	1	4;7;14

Fuente: elaboración propia

En la Tabla 9 hay un caso particular con PR4 y el CTI (PR13) como presuntos responsables. El relato describe un falso positivo judicial por medio de amenazas y participación de las dos entidades. Se deja a PR13 porque son la entidad que puede realizar la captura, aunque, según el relato, se trataba de un falso positivo judicial.

Amenaza DIHC. Como se puede ver en la Tabla 10, para este tipo de violación hay cinco casos en los cuales se tiene al Estado Colombiano (PR1) y a los Paramilitares (PR14) como presuntos responsables. En estos casos se deja a los segundos dado que son quienes cometan la acción, según lo narrado. Se intuye también que estas amenazas no se producirían si hubiera presencia de PR1 en los territorios.

Tabla 10. Más de un PR en Amenaza DIHC

Fecha	Ubicación	Víctima	N	PRTodos
4-dic-20	PUTUMAYO - PUERTO ASÍS	ANDRES CANCIMANCE	1	1; 14
4-dic-20	PUTUMAYO - PUERTO ASÍS	SANDRA LAGOS	1	1; 14
4-dic-20	PUTUMAYO - PUERTO ASÍS	WILMAR MADRÓNERO	1	1; 14
4-dic-20	PUTUMAYO - PUERTO ASÍS	YULE ANZUETA	1	1; 14
4-dic-20	PUTUMAYO - PUERTO ASÍS	YURI QUINTERO	1	1; 14

Fuente: elaboración propia

Violencia Sexual DH. Para esta violación, solo hay un caso que reporta como presuntos responsables al Ejército (PR4), a los Paramilitares (PR14) y a la Policía (PR7). El relato relacionado, responsabiliza directamente a un paramilitar, con nombre propio, como el perpetrador del hecho, por ese motivo PR14 es el presunto responsable en este caso.

Reclutamiento de menores DIHC. Para esta violación, también solo se presenta un caso con AGROMAR (PR37) y FARC-EP (PR25) como presuntos responsables. Según el relato, directamente tiene que ver con un hecho cometido por las disidencias de PR25, entonces se deja este último como presunto responsable.

En la Tabla 11 se puede ver el resultado final de solucionar las incongruencias con los presuntos responsables de la columna “Dif”.

Tabla 11. Tabla de contingencia con solo un presunto responsable por categoría de victimización

	PR1	PR2	PR3	PR4	PR7	PR9	PR13	PR14	PR15	PR20	PR31	PR33	PR35	PR37	PR39	Total fila	sum real	Dif
A Muertos-DH	0	0	0	10	33	0	0	59	0	0	0	0	0	0	0	102	102	0
B Muertos-DIH	0	0	0	0	0	0	0	0	39	10	0	0	0	0	0	58	58	0
C Muertos-VPS	0	0	0	0	0	0	0	0	0	0	12	224	0	0	0	296	296	0
D DESAPARICIÓN-DH	0	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	10	0
E DESAPARICIÓN-DIHC	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	12	12	0
F TORTURA-DH	0	0	0	11	16	0	0	16	0	0	0	0	0	0	0	43	43	0
G TORTURA-DIHC	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0
H TORTURA-VPS	0	0	0	0	0	0	0	0	0	0	0	1	10	0	0	11	11	0
I LESIONADOS-DH	0	6	0	11	71	1	0	6	0	0	0	0	0	0	0	95	95	0
J LESIONADOS-DIHC	0	0	0	0	0	0	0	0	16	1	0	0	0	0	0	35	35	0
K LESIONADOS-VPS	0	0	0	0	0	0	0	0	0	0	0	2	44	0	0	46	46	0
L DETENCIÓN ARBITRARIA-DH	0	1	0	29	28	0	0	2	0	0	0	0	0	0	0	70	70	0
M AMENAZA-DH	3	1	0	48	23	0	1	251	0	0	0	0	0	0	1	328	328	0
N AMENAZA-DIHC	0	0	0	0	0	0	0	70	63	7	1	0	0	0	0	149	149	0
Ñ AMENAZA-VPS	0	0	0	0	0	0	0	0	0	0	0	12	91	0	0	103	103	0
O ATENTADO-DH	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	4	4	0
P ATENTADO-VPS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	12	0
S VIOLENCIA SEXUAL-DH	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	7	7	0
U VIOLENCIA SEXUAL-VPS	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
W SECUESTRO-VPS	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	10	10	0
X RAPTO-VPS	0	0	0	0	0	0	0	0	0	0	0	0	1	14	0	15	15	0
AD RECLUTAMIENTO DE MENORES-DIHC	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	4	0
Total columnas																1436	1352	84

3.4.3 Decisiones finales acerca de la tabla de contingencia

Las decisiones acerca de las categorías que finalmente quedaron en la tabla de contingencia que fue objeto del Análisis de Correspondencias simple, fueron de tres tipos: la primer referente a la simplificación de la complejidad en términos del número de categorías a considerar, la segunda referente al conocimiento implicado en las disciplinas sociales: estas marcan diferencias entre categorías, que se salen del ámbito del conocimiento de los autores; y la tercera que obedece a no considerar categorías donde su ocurrencia haya sido muy insignificante comparadas con otras categorías. Las decisiones que se explican en lo que sigue se encuadran dentro de algunas de las razones anteriores.

Si bien en las clasificaciones de la revista Noche y Niebla, existen categorías y variables más detalladas para las violaciones de DDHH, dependiendo de la acción y el autor, que involucran distinciones entre el DH, DIH, DIHC y VPS (como se vio en la sección 3.3), para los propósitos de este trabajo y dada la complejidad que implica trabajar con tales distinciones, hemos optado por simplificar las categorizaciones con el propósito de ayudar en la interpretación

de los datos. Por ejemplo, en lugar de hablar de Muertos atribuidos a DH, o a DIH, o a VPS, simplemente agrupamos los tres en una sola categoría llamada MUERTOS (ver tabla 12)..

Parte de la complejidad reside en el amplio conocimiento que hay que tener en áreas sociales, filosóficas y especialmente jurisprudenciales al momento de hablar de los derechos fundamentales de las personas. Concebir conclusiones relacionadas con estos últimos involucran la utilización de conocimientos profundos del orden constitucional, así como de sus técnicas de aplicación e interpretación y esto no hace parte del ámbito de los propósitos definidos para este trabajo de grado.

Aunque no se analicen todas las categorías, esto no implica que se pierde información, porque el conteo de los casos continuará en las categorías agrupadas. Además, analizar datos que se encuentran en un nivel de desagregación elevado, en ocasiones no permite avanzar en la comprensión de patrones que se puedan observar en estos.

Por otra parte, con el Análisis de Correspondencias simple, aplicado a una tabla de contingencia para las variables que se decidió agrupar, se generará una representación geométrica en una dimensión posible de observar, la cual posibilita interpretar los datos más fácilmente que con la simple observación de las tablas o polígonos que reportan las frecuencias.

Otra decisión de importancia para trabajar con los datos, para el caso de las filas (categorías de victimización) fue tener en cuenta solo las que tuvieran en total más de 20 registros. Entonces, las categorías, ATENTADO DH, ATENTADO VPS, VIOLENCIA SEXUAL DH, VIOLENCIA SEXUAL VPS, SECUESTRO VPS, RAPTO VPS y RECLUTAMIENTO DE MENORES, no se tuvieron en cuenta en la tabla de contingencia final.

Finalmente, en la Tabla 1 (sección 3.4.1), se pueden ver todos los códigos numéricicos de los presuntos responsables tenidos en cuenta en la edición 62 de la revista Noche y Niebla.

Algunos de estos no fueron considerados, debido a que la frecuencia no era significativa. Los que quedan en la tabla de contingencia final, son: PR2: Fuerza Pública, PR4: Ejército, PR7: Policía, PR14: Paramilitares, PR25: FARC-EP (disidencias), PR28: ERP (Ejército Revolucionario del Pueblo, disidente del ELN), PR33: Otros (Bandas criminales parcialmente identificadas), PR35: Fiscalía, PR37: AGROMAR S.A (relacionado con minas antipersonales).

En la Tabla 12, se puede ver la tabla de contingencia final que se tendrá en cuenta para el Análisis de Correspondencias simple.

Tabla 12. Tabla de contingencia final

	PR2	PR4	PR7	PR14	PR25	PR28	PR33	PR35	PR37
A MUERTOS	0	10	33	59	39	10	12	224	9
D DESAPARICIÓN	0	4	0	6	12	0	0	0	0
F TORTURA	0	11	16	16	1	0	1	10	0
I LESIONADOS	6	11	70	6	16	1	2	44	18
L DETENCIÓN ARBITRARIA	1	39	28	2	0	0	0	0	0
M AMENAZA	1	48	23	251	63	7	12	91	0

Fuente: elaboración propia

El álgebra lineal como fundamento matemático del análisis estadístico multivariado

Aunque en este trabajo no pretendemos profundizar en el funcionamiento del álgebra lineal en los análisis estadísticos, si consideramos de gran importancia poder abordar algunas ideas acerca de temas de esta rama de las matemáticas, involucrados en los análisis que serán utilizados aquí.

Como ayuda para comprender los temas discutidos, se usará de guía el acercamiento intuitivo para el álgebra lineal presentado en la lista de reproducción “Essence of linear algebra” (Sanderson, 2016) del canal de YouTube 3Blue1Brown. Este fue creado por el matemático de

Stanford, Grant Sanderson y de allí se tomarán ideas textuales y visuales realmente claves para entender por qué el álgebra lineal es de utilidad. En resumidas cuentas, se espera poder transmitir algo de la comprensión conceptual que entrega este recurso audiovisual. La idea es darle mayor importancia a esto último en lugar de, por ejemplo, los cálculos algorítmicos presentados en la mayoría de los textos sobre el tema.

Las ideas se extraen principalmente de los siguientes episodios de la lista de reproducción:

Capítulos de interés
Combinaciones lineales, sistema generador y bases de vectores
Transformaciones lineales y matrices
Multiplicación de matrices como composición
El determinante
Matrices inversas, espacios de las columnas y núcleo
Cambio de bases
Valores y vectores propios

Se anima al lector a revisar estos videos. Además, puede que sea necesario, dependiendo del estado en el cual se encuentren los conceptos del álgebra lineal, revisar algunos otros de la lista de reproducción. No cabe duda de que esto ayudaría en la lectura y la comprensión de manera más profunda de los conceptos del análisis estadístico multivariado.

Es de importancia señalar que estos videos proporcionan animaciones de gran utilidad para entender el álgebra lineal como herramienta en el análisis estadístico multivariado u otras

áreas. Infortunadamente estos fragmentos de los videos son difíciles de replicar en un documento escrito. Sin embargo, usando descripciones de lo animado, esta sección propende por develar la simplicidad de las ideas del álgebra lineal y su gran utilidad al considerar espacios vectoriales de más de dos dimensiones.

4.1 Combinaciones lineales, sistema generador y bases de vectores

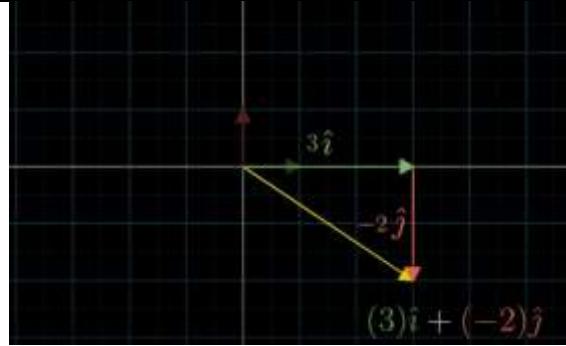
Una buena manera de conceptualizar las coordenadas de un vector, por ejemplo $\begin{bmatrix} x \\ y \end{bmatrix}$ en dos dimensiones (2d de aquí en adelante), es pensar que cada una de ellas es un escalar que se multiplicará por un vector base y luego se sumarán los resultados. En el caso de dos dimensiones, los vectores unitarios \hat{i} y \hat{j} (ver fig.1), serían los vectores de la base y la representación del vector amarillo, en términos de estos, es la combinación lineal de la fig. 2: allí el vector \hat{i} se multiplicó por 3, el vector \hat{j} cambió su dirección al escalarlo por un factor de -2 y ambos se sumaron.

Figura 1. Vectores unitarios en dos dimensiones



Fuente: 3Blue1Brown

Figura 2. Vector en términos de vectores unitarios o vectores base del sistema coordenado xy



Fuente: 3Blue1Brown

Parafraseando y en concordancia con Sanderson (2016) cada vez que se describan vectores numéricamente, hay una decisión implícita acerca de qué vectores base usamos. El autor también aclara que “el conjunto de todos los posibles vectores que se pueden alcanzar con una combinación lineal de un par de vectores, se denomina el sistema generador de esos dos vectores”. Identificar o determinar un sistema generador, es una manera de preguntar cuáles son todos los posibles vectores que se pueden alcanzar solo usando la suma entre vectores y la multiplicación por escalar.

En resumen, un conjunto de vectores linealmente dependientes, como vectores de la base, no es suficiente para describir un espacio vectorial. Cuando ello sucede existirán vectores que no se pueden generar a partir del par de vectores linealmente dependientes. Por ejemplo, se podrían limitar a la extensión de una única recta (en 2d) como en el caso de $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ y $\begin{bmatrix} -1 \\ -3 \end{bmatrix}$. Escalar y sumar estos vectores, deja solo combinaciones lineales que caerán en la recta $y = 3x$. En cambio, basta con tener un sistema generador: en 2d, sería un conjunto de dos vectores linealmente independientes (como es el caso de los vectores unitarios $\hat{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $\hat{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, en donde cada uno aporta una dimensión al sistema) para tener de manera completa un espacio vectorial de dos

dimensiones. Es decir: cualquier vector de dos dimensiones se puede construir como una combinación lineal de los vectores de la base.

4.2 Transformaciones lineales y matrices

Sanderson (2016), de manera conveniente cuando se refiere a las matrices, las denomina transformaciones lineales. Primero, la palabra transformación, menciona el autor, sería una manera elegante de decir “función”; en este caso sería una función que tiene como entradas y salidas vectores. Que sean “lineales” tiene que ver con que, aplicar una transformación lineal, por ejemplo en 2d, va a “mantener todas las líneas del espacio paralelas, equidistantes y dejará el origen del sistema coordenado fijo” (Sanderson, 2016). En el lenguaje formal de las matemáticas del álgebra lineal, esto se traduce en satisfacer las propiedades de linealidad, es decir, una transformación L que vaya de un espacio vectorial en otro, debe satisfacer:

$$L(\vec{v} + \vec{w}) = L(\vec{v}) + L(\vec{w}) \text{ y } cL(\vec{v}) = L(c\vec{v}).$$

Numéricamente, cada columna de una matriz indica a dónde mover cada uno de los vectores de la base de un sistema generador. En la fig. 3, se ilustran los dos casos en donde los vectores unitarios \hat{i}, \hat{j} y el vector amarillo, se pueden ver antes y después de aplicar una transformación lineal. Hay que considerar que solo es necesario mover los vectores de la base y todos los posibles vectores del espacio cambiarán de acuerdo con esto. De esta manera es posible ver a la transformación lineal como un todo: cada posible vector del sistema coordinado de la izquierda (ver fig. 3) va a tener una equivalencia en el de la derecha, es decir que cada punto en el espacio se movió a otro, pero manteniendo las líneas paralelas, equidistantes y el origen fijo. A manera de ayuda, en la fig. 3 se deja en gris el sistema coordinado original como una guía para notar qué le pasó al espacio.

La transformación del espacio en 2d de la fig. 3, quedaría descrita solamente usando los cuatro números de una matriz 2×2 , en donde cada columna indica hacia dónde se movió cada vector de la base. La primera para \hat{i} y la segunda para \hat{j} :

$$\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}$$

Adicionalmente, la multiplicación de la matriz anterior con la coordenada que representa al vector amarillo inicial: $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, es como si este hubiera entrado en una “función” la cual como salida entrega un nuevo vector. Dicha salida se consigue de manera similar a cuando los vectores base eran unitarios: a partir de la combinación lineal de cada columna de la matriz (que es la nueva posición de \hat{i} y \hat{j}), con las coordenadas del vector:

$$\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

De manera general se puede ver en la fig. 4, donde para una determinada transformación lineal del espacio en dos dimensiones (descrita por una matriz), sería posible establecer las posiciones en donde cualquier vector aterrizaría luego de aplicarla.

Figura 3. Transformación lineal en 2d

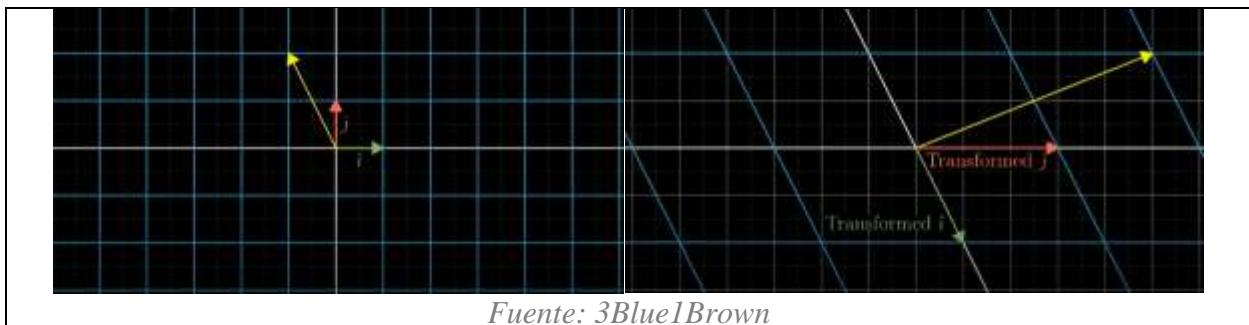


Figura 4. Multiplicación matriz-vector

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Fuente: 3Blue1Brown

4.3 Multiplicación de matrices como composición

Al construir sobre las ideas de los apartados anteriores, si se tiene en cuenta que una matriz describe una transformación lineal del espacio, habría que considerar los casos en los cuales es necesario aplicar más de una transformación. “La multiplicación entre matrices tiene el significado geométrico de aplicar una primera transformación y luego otra” (Sanderson, 2016).

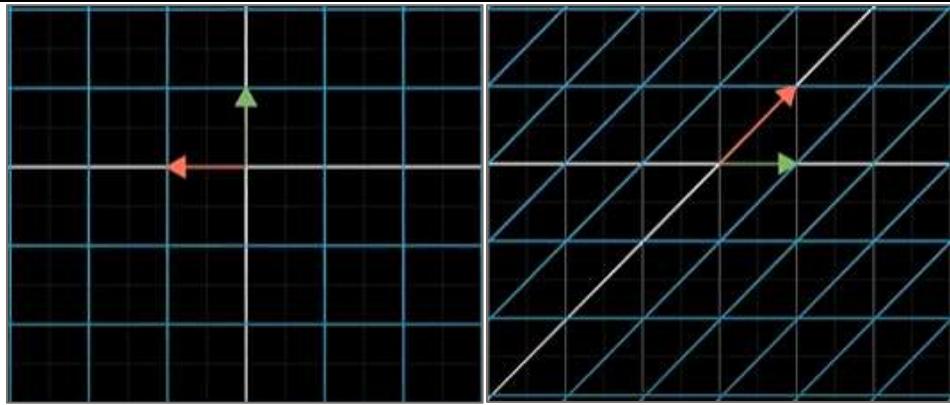
Lo anterior se acerca mucho a la concepción de componer funciones f y g como

$(f \circ g)(x) = f(g(x))$. Incluso la lectura de derecha a izquierda en este caso se transfiere también para la multiplicación de matrices.

Un caso específico en 2d sería una rotación de 90° en sentido antihorario. Dicha transformación estaría representada por la matriz $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Otra transformación, denominada “shear”⁴ es descrita por la matriz $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (ver fig. 5). Una transformación lineal que describa estos dos movimientos en el espacio, uno después del otro y empezando con la rotación de 90° , estaría dada por $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$ (ver fig. 6).

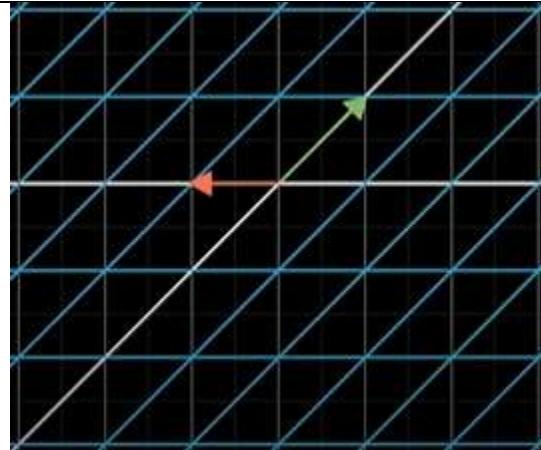
⁴ Esta transformación se refiere a un movimiento que hace alusión a considerar unas tijeras abiertas en ángulo de 90° que se cierran. Para el caso de la fig. 5, lo hacen hacia un ángulo de 45°

Figura 5. Izquierda: 90° en sentido antihorario, derecha: transformación “shear”



Fuente: 3Blue1Brown

Figura 6. Rotación 90° en sentido antihorario seguida de la transformación “shear”



Fuente: 3Blue1Brown

Obtener la matriz que representa la transformación lineal de la fig. 6, se conecta con el apartado anterior, porque luego de aplicar la primera rotación al espacio y obtener nuevas posiciones para los vectores de la base (ver fig. 5, izquierda), solo sería necesario hacer dos multiplicaciones de matriz-vector para saber cuál sería el vector de salida del \hat{i} con la segunda transformación:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

y de manera análoga para saber las nuevas coordenadas de \hat{j} :

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

(“shear”, seguida de 90° en sentido antihorario), no da el mismo resultado, ya que en general la multiplicación de matrices no es comutativa.

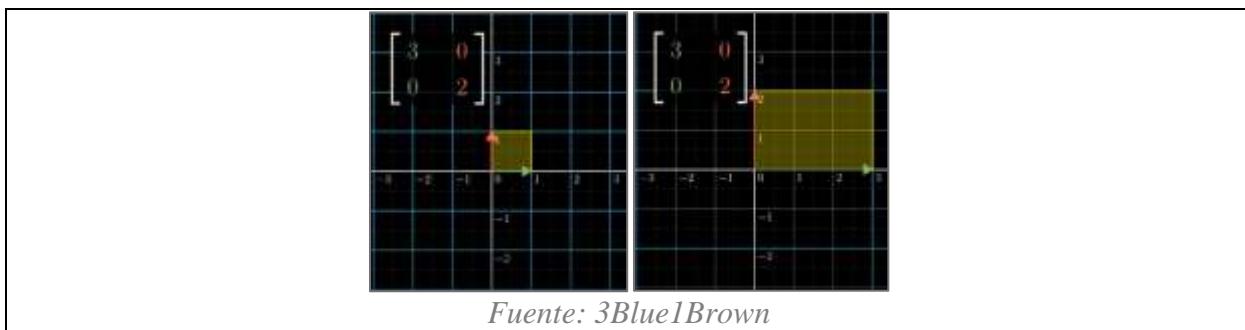
Al considerar la comparación de la multiplicación de matrices con expresiones de la forma $f(g(x))$, es conveniente pensar que también un vector en el espacio, sin aplicar ninguna transformación (ver fig. 3, izquierda), tendrá una traducción luego de aplicar una transformación lineal seguida de otra. En el caso de la fig. 6, y para cualquier vector, sería:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

4.4 El determinante

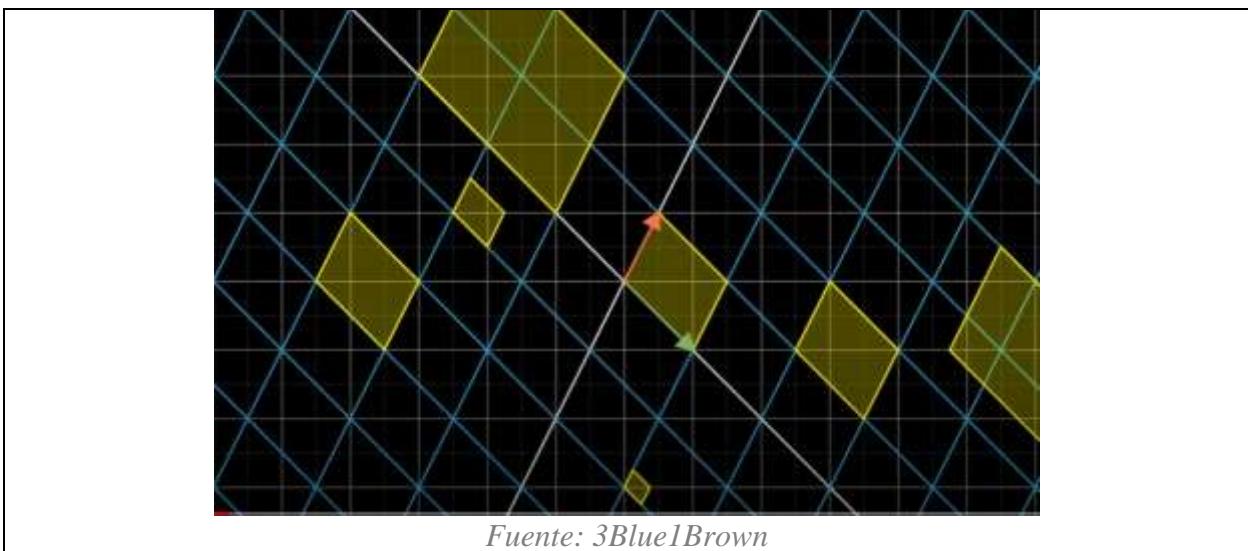
Al aplicar una transformación lineal, desde el punto de vista geométrico, generalmente se piensa que el espacio se estira o se comprime. En palabras de Sanderson (2016) valdría la pena entonces “medir el factor por el cual el área de una región dada aumenta o disminuye”. Para el caso de la fig. 7, antes de enviar a los vectores de la base a las coordenadas descritas por la matriz, se tiene un cuadrado de lado 1, y luego de aplicar la transformación, el área original se estiró en un factor de 6. Este valor es el que comúnmente se conoce como determinante de una matriz.

Figura 7. Determinante visto como área, antes y después de aplicar la transformación lineal.



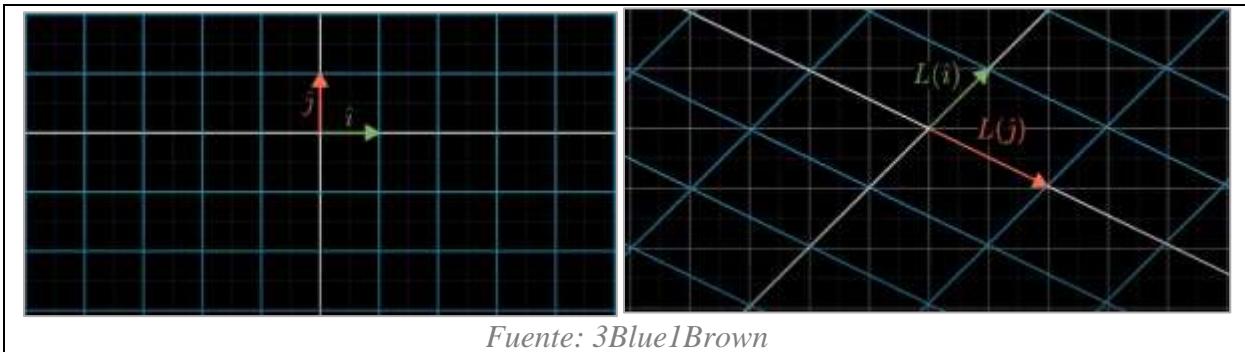
Para resumir la idea, solo habría que saber qué le pasa al área determinada por los vectores de la base, porque, nuevamente, al tener todas las rectas paralelas, equidistantes y el origen fijo; luego de aplicar una transformación, el área de cualquier posible región también será afectada por el determinante (ver fig. 8). Estás áreas incluyen superficies que no necesariamente tengan lados rectos, ya que estas pueden “ser reproducidas bastante bien con aproximaciones arbitrariamente buenas, si se usan cuadrados lo suficientemente pequeños” (Sanderson, 2016).

Figura 8. Ejemplos de algunas áreas resultantes, luego de aplicar una transformación lineal específica.



Con esta visión, vale la pena señalar los casos en los que el determinante de una matriz es mayor a cero, menor que cero e igual a cero. Para el primero, la fig. 8 sugiere que, si el valor del determinante es positivo, simplemente el área aumenta o disminuye. En el segundo caso, en el cual se obtiene un signo negativo en el determinante, este indica que se invierte la orientación del espacio, pero las áreas continúan siendo escaladas por el valor absoluto del número. Respecto al signo del determinante, en la fig. 9 se puede ver que luego de la transformación lineal, \hat{j} queda a la derecha de \hat{i} , cuando inicialmente estaba a la izquierda. Esto es un indicador del cambio de orientación.

Figura 9. Determinante menor que cero, invierte orientación del espacio



Fuente: 3Blue1Brown

Para el caso en el cual el determinante es igual a cero, indica que todas las posibles áreas pasaron a tener este valor. En otras palabras, si el determinante es igual a cero (en 2d), tiene que ver con el hecho de pasar de tener áreas, a una recta o un punto (que tiene área igual a 0); o, en tres dimensiones, se relaciona con pasar a regiones con volumen igual a 0, como un plano. De manera importante Sanderson (2016) recalca que verificar “si el determinante de una matriz dada es igual a cero, entregará una manera de calcular si la transformación asociada con esa matriz reduce todo a una dimensión menor o no”.

4.5 Matrices inversas, espacios de las columnas y núcleo

La utilidad más reconocida para las matrices es la solución de sistemas de ecuaciones lineales. Es posible simplificar de manera relevante la notación, al representar en una sola línea un sistema de ecuaciones usando matrices, como se puede ver en la fig. 10.

Figura 10. Matrices para solucionar sistemas de ecuaciones

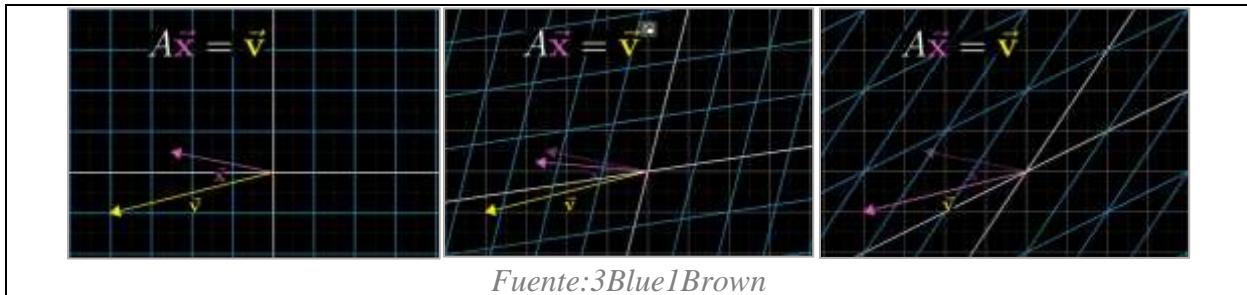
$$A\vec{x} = \vec{v}$$

$$\begin{array}{l} 2x + 5y + 3z = -3 \\ 4x + 0y + 8z = 0 \\ 1x + 3y + 0z = 2 \end{array} \rightarrow \underbrace{\begin{bmatrix} 2 & 5 & 3 \\ 4 & 0 & 8 \\ 1 & 3 & 0 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}}_{\vec{v}}$$

Fuente: 3Blue1Brown

En el lenguaje de la intuición geométrica presentada por el autor, la matriz A corresponde con una transformación lineal. Entonces, resolver $A\vec{x} = \vec{v}$, tiene que ver con el hecho de buscar un vector \vec{x} el cual, luego de aplicar la transformación (al entrar en la “función”), aterriza en \vec{v} . Tres etapas de este proceso pueden verse en la fig. 11 (nótese el desplazamiento de los ejes coordenados representados en color blanco).

Figura 11. El vector \vec{x} aterrizando en el \vec{v}

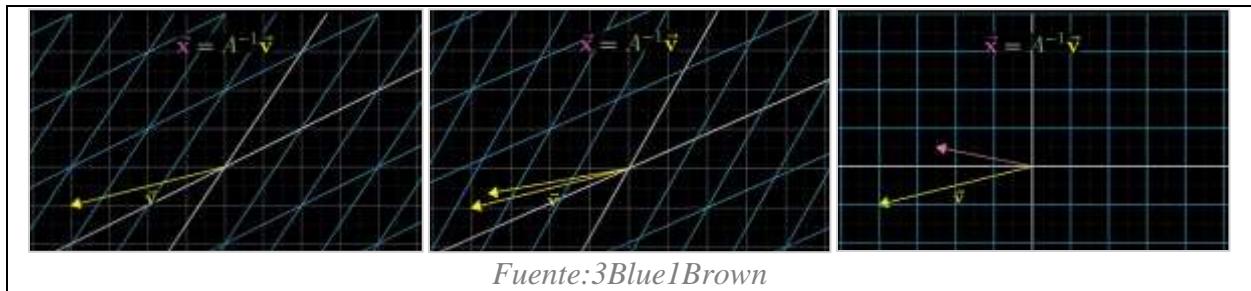


Encontrar las soluciones para las incógnitas dependerá entonces de si la transformación lineal descrita por A , hace que el espacio tenga una menor dimensión, o si continúa generando todo el espacio. En otras palabras, se subdivide en los casos en los cuales el determinante de A es igual o diferente de cero.

El caso más probable: $\det(A) \neq 0$, indica que solo habrá un único vector que aterrizará en \vec{v} , el cual se puede encontrar “reproducido la transformación en reversa” (Sanderson, 2016). Esto corresponde con la transformación A^{-1} (la inversa de A).

Luego de encontrar A^{-1} , que puede realizarse con la ayuda de un software, sería posible solucionar el sistema multiplicándola por la izquierda a ambos lados del igual: $A^{-1}A\vec{x} = A^{-1}\vec{v}$. En la izquierda de la igualdad $A^{-1}A = I$, se tiene I como identidad, porque $A^{-1}A$ describe una transformación lineal que no hace nada: primero se aplica A seguida de revertir lo que se hizo, con A^{-1} . En la derecha se tiene la reproducción en reversa a A y el propósito, con este acercamiento geométrico, sería entonces seguir a \vec{v} , para encontrar a \vec{x} . En la fig. 12 se pueden ver tres momentos de la aminación que representan lo descrito anteriormente, en donde ahora es \vec{v} el que entra en una “función” descrita por A^{-1} .

Figura 12. Reproduciendo en reversa de la transformación lineal dada por una matriz A



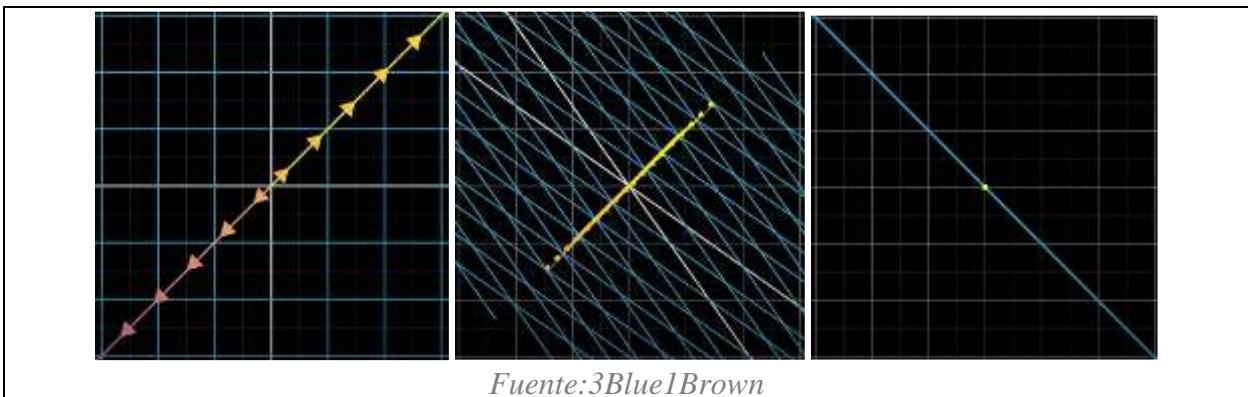
Pero cuando se tiene $\det(A) = 0$, A reduce el espacio a una dimensión menor, por lo tanto, la inversa no existe. En términos del autor, si existiera la inversa para el caso de 2d, esto implicaría “descomprimir” una recta o un punto para obtener un plano, lo cual es imposible de realizar a través de una función, dado que, a un solo vector, no se le pueden asociar dos vectores simultáneamente (es decir en su imagen). Sin embargo, aún podrían existir recursos para dar una solución al sistema de ecuaciones en este caso, pero queda más reducido el ámbito para

encontrarla. Por ejemplo, habría que tener suficiente suerte como para que \vec{v} , luego de ser transformado por A , caiga en la recta a la cual el espacio de dos dimensiones se redujo.

Estas posibilidades de solución, al estar en el caso del determinante igual a cero, se resumen en pensar en el *rango de la transformación*: este tiene que ver con “el número de dimensiones en la salida de una transformación” (Sanderson, 2016). Como una matriz representa una transformación lineal, el número de columnas indicaría el rango, por eso al conjunto de todas las posibles salidas, se le denomina el *espacio de las columnas*: que es el sistema generador de lo descrito por la transformación. Por ende, el rango se puede definir como “el número de dimensiones en el espacio de las columnas” (Sanderson, 2016). Entonces, por ejemplo, una transformación con rango 2, tiene que ver con el hecho de que los vectores base continúan generando el plano xy luego de moverlos a dónde indican las columnas de una matriz. Si se continúa generando el mismo espacio que había antes de aplicar la transformación, se le denomina de rango completo. Otro ejemplo es que se tenga una con rango de 1: esta implica que ahora solo hay una dimensión en la salida y el espacio pasó a ser una recta o un punto, en el caso de 2d.

Para terminar esta parte, es de importancia pensar que, por ejemplo, el vector $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, siempre va a estar en el espacio de las columnas. Luego de aplicar una transformación, puede que el espacio tenga rango completo y a $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ no le ocurrió nada. Pero en el caso contrario, puede haber un montón de vectores que caen en $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ luego de modificar los vectores de la base. Para una transformación que pasa de 2d a una recta, es decir de rango de 1, se pueden ver en la fig. 13 tres imágenes de cómo varios vectores ubicados en una recta diferente terminan en $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Figura 13. Transformación lineal que pasa de 2d a 1d y envía vectores a $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$



El conjunto de vectores que terminan en $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ luego de aplicar la transformación lineal, es

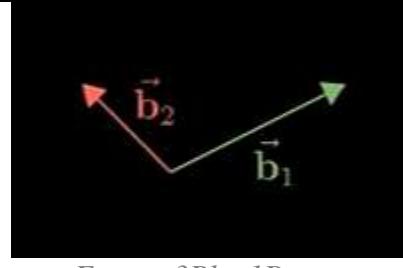
lo que denominamos como *el núcleo*: este es el conjunto de todos los vectores que se vuelven nulos. En términos de un sistema de ecuaciones, cuando se tiene $A\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, “el núcleo entrega todas las posibles soluciones para el sistema” (Sanderson, 2016).

4.6 Cambio de base

Como se ha mencionado anteriormente, hay suposiciones implícitas cuando se describen vectores en sistemas coordenados: en los casos de 2d, implican escalar y si es necesario cambiar la dirección de los vectores de la base; además, se requiere tener una convención para la unidad de distancia que se toma de referencia al hacer estos movimientos.

Una idea interesante sería la posibilidad de escoger otros vectores para la base, como por ejemplo los de la fig. 14, que no son los ejes ortogonales presentados en los ejemplos de 2d de los apartados anteriores. No obstante, para alguien que no esté familiarizado con el uso estandarizado de bases ortogonales, pueden llegar a ser concebidos como vectores apropiados para hacer un sistema generador, e incluso tener otra convención para la unidad, como se puede intuir de la fig. 14.

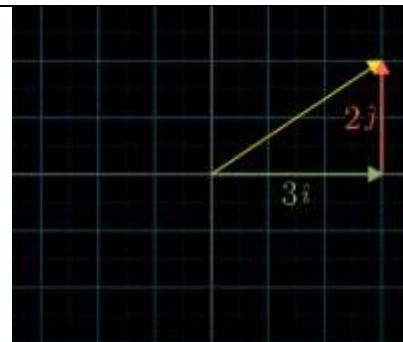
Figura 14. Vectores alternativos para una base en 2d



Fuente: 3Blue1Brown

Si consideramos el vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$, en el sistema coordenado común de 2d, lo describiríamos escalando y sumando los vectores de la base, como en la fig. 15.

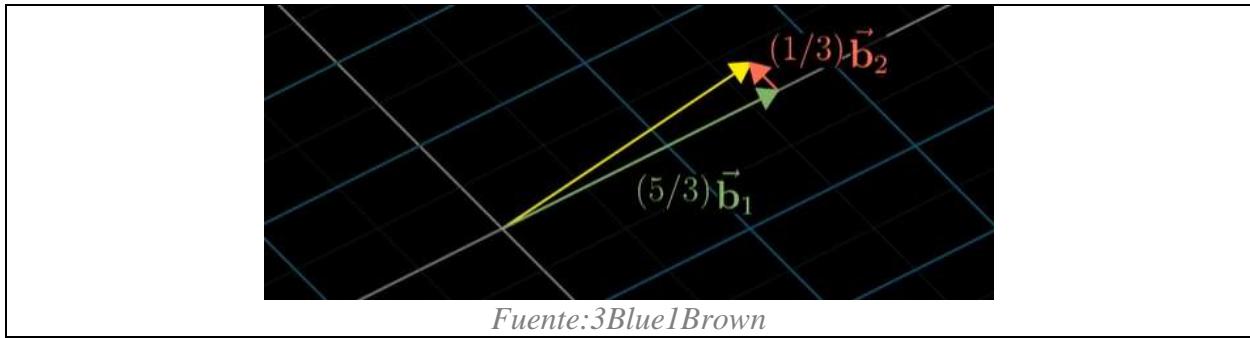
Figura 15. Vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ en 2d



Fuente: 3Blue1Brown

Pero si se considerarán los ejes coordinados de la fig. 14, explicaremos más adelante que $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ luego de realizar la combinación lineal de los vectores de la base, luciría como en la fig. 16 y en realidad sus coordenadas, escalando y sumando \vec{b}_1 y \vec{b}_2 , serán $\begin{bmatrix} 5/3 \\ 1/3 \end{bmatrix}$, las cuales, en el sistema alternativo, son las instrucciones para ir del origen a la punta de $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ en el espacio generado con \hat{i} y \hat{j} unitarios.

Figura 16. Vector $\begin{bmatrix} 5/3 \\ 1/3 \end{bmatrix}$ en sistema coordenado alternativo



Intuitivamente las diferentes elecciones de vectores para la base se pueden comparar con la forma en la cual se interpretaría un determinado vector en diferentes idiomas: en los casos anteriores se habla del mismo vector, pero se están usando diferentes “palabras” para describirlo.

Justo como es posible traducir palabras entre idiomas, también habrá una “traducción” entre sistemas coordinados, la cual se puede alcanzar considerando la idea de las matrices como transformaciones lineales. Primero, *la matriz del cambio de base*, que representa las posiciones de \vec{b}_1 y \vec{b}_2 en la fig. 14 es $\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$ para el sistema coordenado común en 2d. Geométricamente, aplicar esta matriz en el plano xy , ubica a \hat{i} y \hat{j} en las posiciones de \vec{b}_1 y \vec{b}_2 ; por otro lado, numéricamente, una multiplicación de tipo $\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$, como vimos anteriormente, permitirá “traducir” un vector del sistema coordenado que usa \vec{b}_1 y \vec{b}_2 como base, a uno que usa a \hat{i} y \hat{j} en su lugar. Esto último se debe a que es una multiplicación matriz-vector, la cual entregará una imagen de $\begin{bmatrix} x \\ y \end{bmatrix}$ sometida a la transformación $\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$, donde las columnas representan \vec{b}_1 y \vec{b}_2 , en el lenguaje usual del plano xy .

Con lo anterior en mente, sería posible entonces ir en la otra dirección de estas concepciones geométrica y numérica si se considera $\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix}$: ahora, geométricamente se pasaría de \vec{b}_1 y \vec{b}_2 a \hat{i} y \hat{j} (porque se reprodujo en reversa la transformación); y, numéricamente, permite ir del lenguaje que usa como base a \hat{i} y \hat{j} al que usa a \vec{b}_1 y \vec{b}_2 . Usando este último hecho, fue posible encontrar las coordenadas del vector de la fig. 16:

$$\begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5/3 \\ 1/3 \end{bmatrix}.$$

Hasta aquí se ha descrito cómo es posible traducir, de un lado para el otro, la descripción de vectores individuales entre sistemas coordinados, pero también sería relevante pensar en cómo se verían transformaciones del plano con ejes alternativos. Para eso se considera la multiplicación entre matrices como herramienta.

El ejemplo para la multiplicación de matrices como composición de un apartado anterior, tiene que ver con la rotación de 90° en sentido antihorario, descrita para los vectores unitarios comunes como $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ (ver fig. 5). La manera de averiguar cómo se vería esta rotación en un sistema con vectores base alternativos, tiene que ver con convenientemente realizar multiplicaciones entre matrices para poder alcanzar el efecto deseado.

A continuación, se describen los pasos para alcanzar esto, siempre pensando en la idea de multiplicación de matrices como composición:

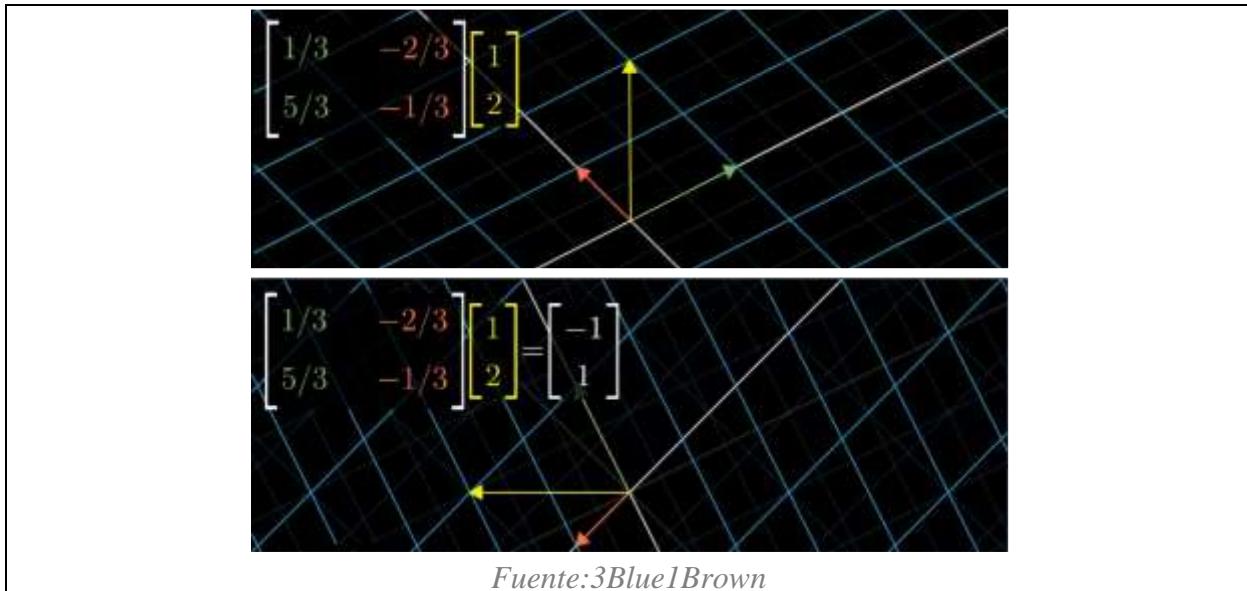
- Primero se comienza con un \vec{v} en el lenguaje de los vectores \vec{b}_1 y \vec{b}_2 , el cual al realizar $\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \vec{v}$, permite ver el mismo vector, pero en el lenguaje de los vectores unitarios comunes.

- Luego se aplica la transformación a lo obtenido: $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \vec{v}$, esta multiplicación todavía indica dónde \vec{v} aterriza en el lenguaje de \hat{i} y \hat{j} , luego de hacer la rotación.
- Finalmente se aplica la inversa de la matriz de cambio de base:

$$\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \vec{v} = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \vec{v}, \text{ para obtener el vector transformado, pero ahora en el lenguaje de } \overrightarrow{b_1} \text{ y } \overrightarrow{b_2}.$$

Esa composición de tres matrices, luego de realizar los cálculos, entrega la rotación en el lenguaje de $\overrightarrow{b_1}$ y $\overrightarrow{b_2}$. Por lo tanto, si se hace una operación como $\begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \vec{v}$, se entregará la versión de \vec{v} , rotada y expresada en las coordenadas del sistema generado por $\overrightarrow{b_1}$ y $\overrightarrow{b_2}$. En la fig. 17, se puede ver un ejemplo de cómo se vería un vector luego de realizar la rotación de 90° con la matriz descrita en las líneas anteriores.

Figura 17. Rotación de 90° en sentido antihorario para un espacio generado por vectores alternativos.



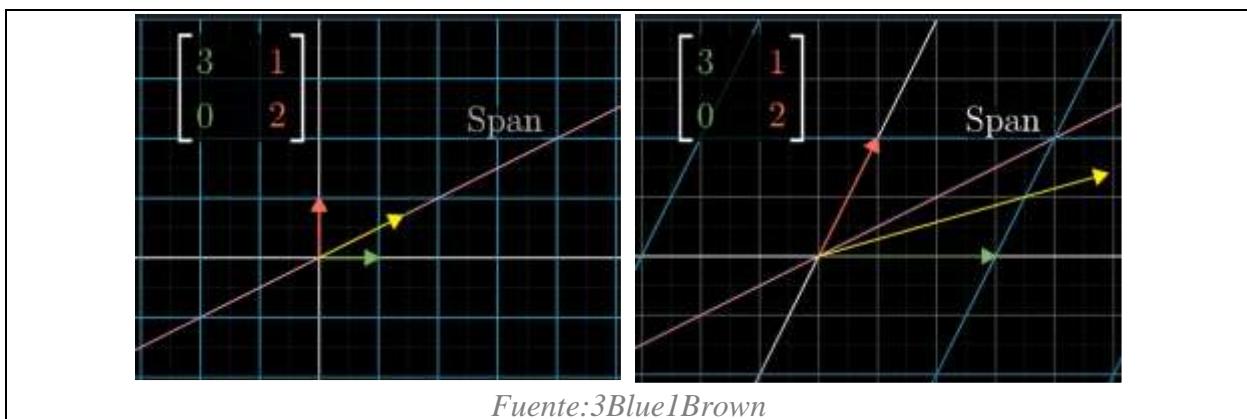
En general “cuando se tiene una expresión como $A^{-1}MA$, sugiere un tipo de “empatía matemática” (Sanderson, 2016), la matriz M representa una transformación de algún tipo como se vería normalmente en el plano xy , y las matrices de los lados, representan esa empatía, el cambio de perspectiva, y “el producto completo, representa esa misma transformación, pero como alguien más la vería” (Sanderson, 2016).

4.7 Valores y vectores propios

Generalmente la comprensión de los valores y vectores propios genera problemas en cuanto a encontrarle sentido a qué significan o para qué sirven. Este apartado, apoyado en los anteriores propende darles sentido.

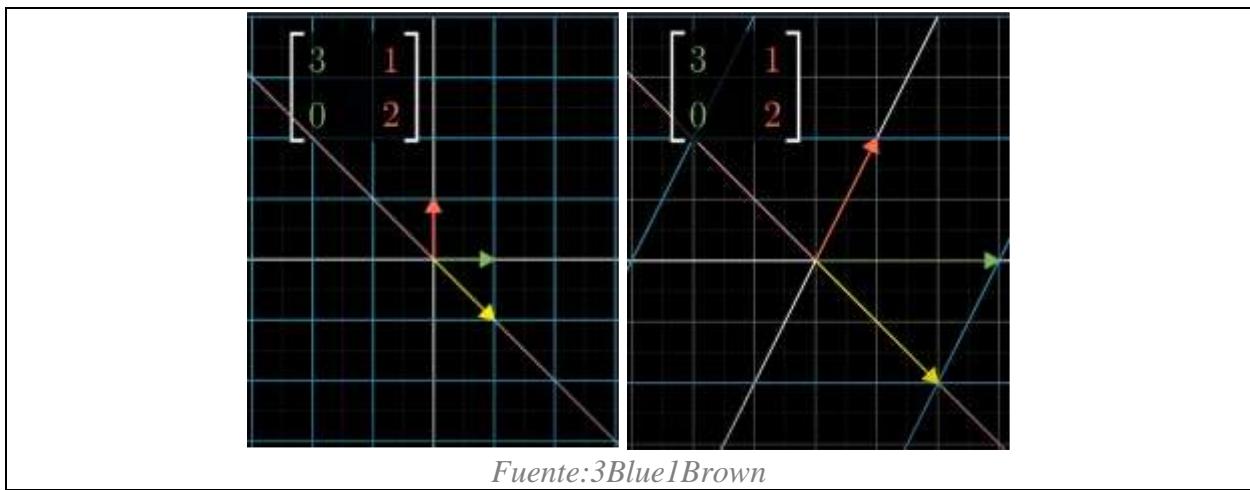
Si consideramos una transformación lineal, representada por la matriz $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$, luego de aplicarla sobre los ejes de la base del plano xy , muchos de los vectores saldrán de su “span” o sistema generador. En palabras del autor, algunos vectores rotarán fuera de la recta que los contiene (en 2d). Un ejemplo es el vector amarillo de la fig. 18: en la imagen de la derecha, luego de aplicar la transformación, podemos ver cómo la salida de ese vector ya no está en la recta en la que inicialmente estaba siendo generado.

Figura 18. Vector se sale de su “span” luego de aplicar la transformación en el sistema coordenado en 2d.



Por otro lado, existirán vectores que luego de ser modificados por la transformación, permanecerán en su propio “span”; simplemente se estirarán o encogerán. A estos vectores se les denomina vectores propios y a los factores por los cuales se escalan, se llaman valores propios. Un vector propio de la transformación discutida se puede ver en la fig. 19: nuevamente, a la derecha, luego de aplicar la transformación, podemos atestiguar que continua en la recta de su sistema generador original y solo fue escalado por un factor de 2, entonces 2 sería el valor propio de ese vector propio. Otro vector propio es \hat{i} , que solo fue escalado por un factor de 3 (su valor propio) y permaneció en su sistema generador (el eje x positivo).

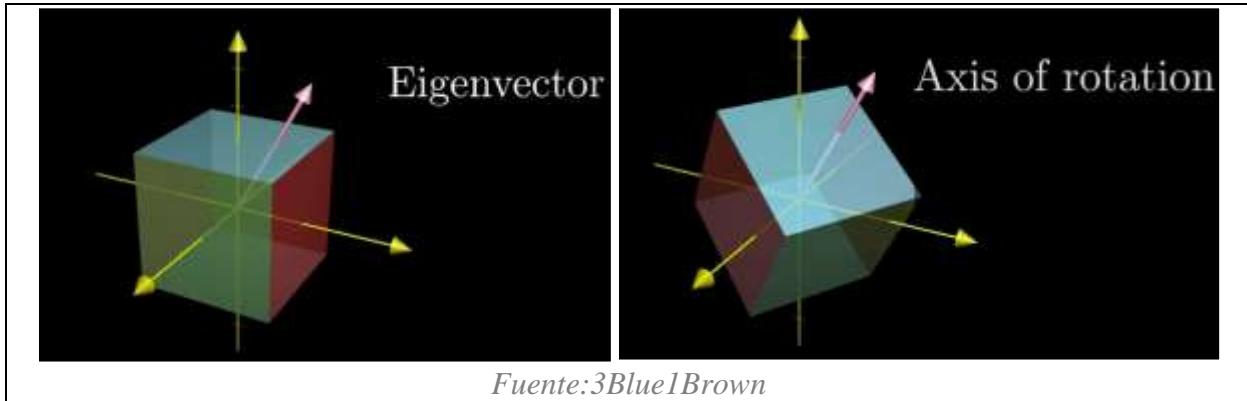
Figura 19. Vector propio de $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$.



Una mirada que propone Sanderson (2016) para entender por qué sería útil conocer estos vectores, es pensar en una rotación en tres dimensiones como la que se puede ver en las dos imágenes de la fig. 20. Un vector que permanezca en su “span” luego de aplicar una transformación para este caso, estaría indicando un eje de rotación. En la fig. 20 es el vector rosa respecto al cual el cubo rotó 30° . Ese vector y el ángulo de rotación permiten resumir qué paso

durante la transformación, en lugar de tener que describir la rotación con una complicada matriz de 3×3 .

Figura 20. Vector propio como eje de rotación en 3d



Fuente: 3Blue1Brown

El patrón descrito en apartados anteriores, en los cuales se ven a las columnas de una matriz como las coordenadas en las cuales aterrizaran los vectores de una base, es algo común en el álgebra lineal y permite entender qué está haciendo una transformación. Sin embargo, esto “le da mucho peso al sistema coordenado” (Sanderson, 2016). Entonces, “una mejor manera de llegar al corazón de lo que realmente hace una transformación lineal, menos dependiente de un sistema coordenado particular, es encontrar los vectores y valores propios” (Sanderson, 2016).

Simbólicamente, la idea se representa con la siguiente expresión: $A\vec{v} = \lambda\vec{v}$, considerando como incógnitas a \vec{v} como el vector propio y λ como su correspondiente valor propio. Esto quiere decir que la multiplicación matriz-vector, da el mismo resultado que simplemente escalar el vector por un valor λ . La línea de razonamiento común para poder encontrar los valores y vectores propios es la siguiente:

$$A\vec{v} = \lambda\vec{v}$$

Hacer el producto λI permite que a ambos lados de la igualdad se tenga una multiplicación matriz-vector.

$$A\vec{v} = \lambda I\vec{v}$$

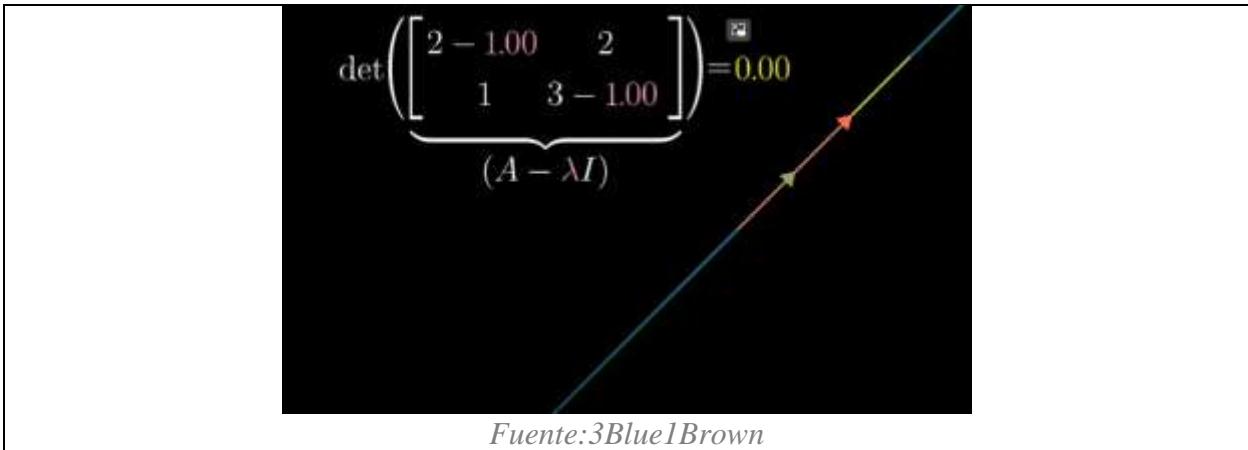
$$A\vec{v} - \lambda I\vec{v} = \mathbf{0}$$

$$(A - \lambda I)\vec{v} = \mathbf{0}$$

Como nos interesan soluciones que cumplan $\vec{v} \neq \mathbf{0}$, la única manera de que se cumpla la igualdad en este punto es considerar que, para que el producto de una matriz con un vector diferente de cero nos entregue $\mathbf{0}$, como vimos antes, es si la transformación asociada con esa matriz reduce el espacio a una dimensión menor. Esa reducción corresponde con $\det(A - \lambda I) = 0$.

Por ejemplo, en la fig. 21, podemos ver que el valor para λ , el cual genera un $\det(A - \lambda I) = 0$ es 1. Esto está diciendo que cuando $\lambda = 1$, la matriz $(A - \lambda I)$ reduce el espacio a una recta. Entonces existe un $\vec{v} \neq \mathbf{0}$ tal que $(A - \lambda I)\vec{v} = \mathbf{0}$ y devolviéndonos en la línea de razonamiento descrita, podríamos encontrar uno o varios vectores propios relacionados con $\lambda = 1$.

Figura 21. Valor propio de 1



Regresando al ejemplo con la transformación $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$, para encontrar los valores propios solo habría que realizar la resta entre matrices y solucionar la ecuación.

$$\det \left(\begin{bmatrix} 3 - \lambda & 1 \\ 0 & 2 - \lambda \end{bmatrix} \right) = (3 - \lambda)(2 - \lambda) = 0$$

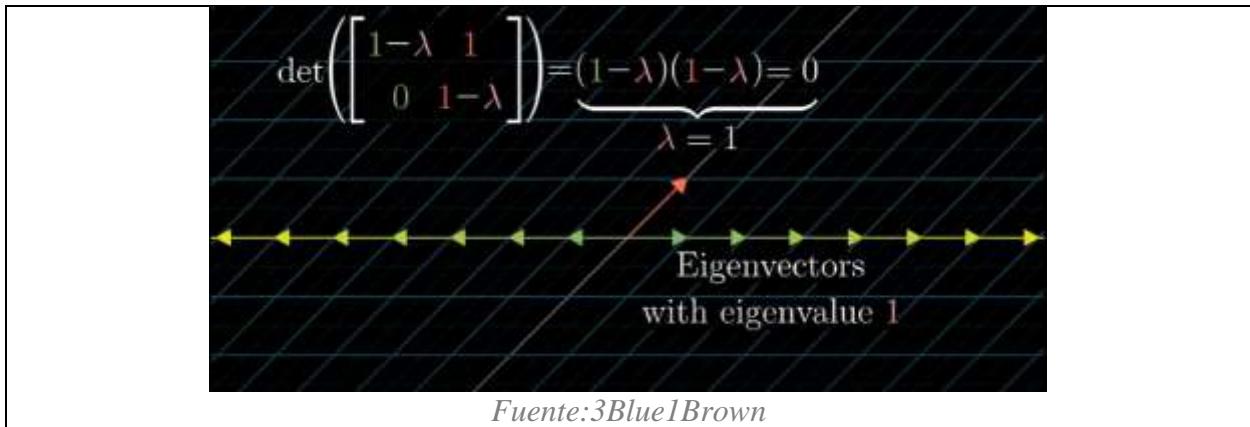
Obteniendo $\lambda = 2$ o $\lambda = 3$. Encontrar los vectores propios, por ejemplo, con $\lambda = 2$, tiene que ver con reemplazarlo en la matriz y solucionar para los vectores a los cuales esta matriz modificada en su diagonal envía a 0.

$$\begin{bmatrix} 3 - 2 & 1 \\ 0 & 2 - 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Calcular esto, de la manera como se haría con cualquier otro sistema lineal, entregaría como respuesta todos los vectores sobre la recta diagonal generada por $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Uno de estos vectores se mostró en la fig. 19. Esto corresponde también con el hecho de que la matriz sin alterar, escala todos los vectores por un factor 2.

Otro ejemplo interesante es la transformación “shear”, que se trató en una sección anterior y que es descrita por $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. En la fig. 22 se puede ver que todos los vectores propios quedan en el eje x y el único valor propio para todos estos sería $\lambda = 1$.

Figura 22. Valor y vectores propios de una transformación “shear”.



La última idea de interés para usar como herramienta en la interpretación del álgebra lineal involucrada en los análisis multivariados que se expondrán más adelante, es la de *base propia*.

Una matriz diagonal, la cual tiene entradas diferentes a cero solo en su diagonal, por ejemplo $\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$, se interpreta pensando que todos los vectores de la base son vectores propios, con las entradas diagonales siendo los valores propios. La utilidad de considerar esta idea tiene que ver con el hecho de que, si

una transformación lineal tiene muchos vectores propios, como la descrita por $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$, suficientes como para escoger un conjunto que genera el espacio completo, entonces sería posible cambiar el sistema de coordenadas de modo que estos vectores propios sean los vectores de la base (Sanderson, 2016).

Esto sería posible, considerando las ideas expresadas en el apartado de cambio de base.

Lo primero que habría que hacer es conseguir la matriz de cambio de base, cuyas columnas en este caso, serían los vectores propios escogidos para la nueva base. En el ejemplo se pueden

escoger: $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ porque son vectores propios. Cuando se usa la idea de la “empatía

matemática”, descrita como $A^{-1}MA$, teniendo a M como la transformación original: $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$, el

resultado sería esa misma transformación, pero desde la perspectiva del nuevo sistema

coordenado generado por los nuevos vectores de la base. Este proceso se puede ver en la fig. 23.

Figura 23. Consiguiendo una base propia

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

Fuente: 3Blue1Brown

Según Sanderson (2016), el punto de hacer esto con vectores propios, es que la nueva matriz si o si, será diagonal, con sus valores propios a lo largo de la diagonal de la matriz. Esto es de esta manera, porque esa matriz resultante en la fig. 23, representa trabajar en un sistema coordinado en donde lo que les pasa a los vectores de la base es que son escalados durante la transformación. Un conjunto de vectores que formen una base y que además sean vectores propios, se denomina una *base propia*.

Según Díaz (2007), en su texto sobre los métodos de inferencia multivariados, indica que los vectores y valores propios, considerando esta interpretación geométrico-estadística, son importantes en la estadística,

pues con estos vectores resulta posible identificar la dirección en que se conserva la información más importante contenida en los datos. Encontrar estos vectores significa hallar la dirección en la que se encuentra una buena parte de la información contenida en los datos (p. 493).

Para finalizar, es relevante considerar que las ideas de los apartados anteriores, generalmente expresadas para 2d, se llevan fácilmente a dimensiones más altas. Lo realmente importante sería notar que la utilidad de este instrumento de las matemáticas es poder pasar de la visión del físico, que son en esencia las representaciones con flechas; a la del analista de datos, que tiene que ver con describir los vectores como listas i.e. $\begin{bmatrix} x \\ y \end{bmatrix}$ y viceversa. Según Sanderson (2016) esto “le permite al analista de datos conceptualizar muchas listas de números de manera visual, lo cual puede clarificar seriamente patrones y entregar una visión global acerca de qué hacen las operaciones”. Por otro lado, para el físico y para quien hace animaciones, les permite “tener una manera de visualizar y manipular el espacio” (Sanderson, 2016).

Elementos del análisis estadístico multivariado

5.1 ¿Qué es un análisis multivariado?

En estadística elemental, los análisis de datos se restringen inicialmente a una sola variable y se denominan como univariados (representados en una dimensión). Es sabido que, por ejemplo, la descripción de este tipo de análisis recurre al uso de gráficas, tablas de frecuencias,

medidas de tendencia central y de dispersión. Estos resultados entregan una idea inicial acerca del comportamiento de los datos recolectados.

Es oportuno ahora resaltar que los datos bivariados involucran análisis en dos dimensiones. En este caso su descripción se apoya, por ejemplo, en medidas de correlación, análisis de distribuciones conjuntas e independencia de las variables aleatorias. Estos resultados, entregan información acerca de los datos, especialmente sobre las relaciones de las dos variables involucradas, porque estas se estudian de manera simultánea. También es cierto que es posible realizar regresiones, inferencias y análisis de varianzas, los cuales permiten entregar conclusiones con cierto grado de confiabilidad respecto a unas hipótesis.

Justo como muchas de las áreas y herramientas de las matemáticas, en la estadística también es posible generalizar los resultados conseguidos en casos particulares. Entonces, al disponer de varias variables aleatorias en un determinado experimento, se pueden tener los enfoques anteriores, pero representados en técnicas del análisis multivariable.

En el mismo sentido al caso bivariado, en los análisis multivariados las variables se estudian de manera simultánea. También está presente la búsqueda de las relaciones entre las distintas medidas, el interés de llegar a conclusiones en una determinada investigación, etc. La estadística multivariada considera un conjunto de individuos, a cada uno de los cuales se les pueden determinar o medir varias características que se configuran en variables. Tal conjunto de individuos y de variables, se puede constituir en el objeto de estudio de un análisis multivariado.

Los datos, generalmente, quedan representados a partir de filas y columnas, en las cuales se relacionan, por ejemplo, a individuos en las filas y a variables acerca de estos últimos en las columnas. Una representación de interés en el presente documento es la Tabla de Contingencia,

la cual relaciona variables categóricas en sus filas y columnas; cada entrada de la tabla representa una frecuencia absoluta para cada cruce entre las variables.

Organizar los datos de la manera como lo hace una tabla de contingencia o una que describa múltiples variables cuantitativas de individuos es muy conveniente. Cada fila y columna, representaría un vector de tamaño n , dependiendo de la cantidad de variables e individuos. Esta disposición de la tabla organiza esos vectores de manera matricial, habilitando la interpretación de los datos (resumidos en cada vector de la matriz a analizar) con las herramientas que proporciona el álgebra lineal.

Poder utilizar el álgebra lineal, garantiza una manipulación formal, estricta y confiable de los datos recolectados en una investigación, con la obvia ventaja de poder trabajar en más de tres dimensiones, en dónde varias ayudas gráficas ya no se podrían visualizar.

5.2 Metodologías y enfoques del análisis multivariado

De acuerdo con Diaz (2007) existen varios campos del conocimiento donde son requeridas técnicas multivariadas para el análisis de datos. Algunos de estos son el mercadeo, la psicología, la medicina o la sociología. El presente trabajo utiliza datos de carácter social: se intentará establecer para un grupo de individuos la relación entre diferentes tipos de violaciones a los DDHH, contrastadas con los presuntos responsables.

Como el contexto mencionado implica considerar un gran número de variables, se vuelve relevante reflexionar también de qué manera acercarse a los datos. Díaz (2007) considera dos metodologías en el análisis multivariado: los métodos factoriales y las técnicas de clasificación. La primera tiene que ver con variables ubicadas en espacios referenciados por coordenadas y la segunda se enfoca en ubicar individuos de manera espacial de acuerdo con las variables que los identifiquen.

El propósito en este trabajo es utilizar el Análisis de Correspondencias simple (AC de aquí en adelante), este es de carácter factorial, puesto que se realiza sobre la tabla de contingencia descrita en un apartado anterior. Esta tiene filas y columnas con múltiples variables que describen los datos recolectados. Al quedar estos en un arreglo matricial, se abre la posibilidad de considerar un espacio para las filas y otro para las columnas. Ambos tienen una dimensión mayor a dos, pero como se verá más adelante, el AC usa herramientas del álgebra lineal, con las cuales es posible continuar con el enfoque de coordenadas, solo que en una dimensión que no es posible visualizar.

Además de la metodología descrita, es necesario adherirse a un enfoque del análisis multivariado. Tal enfoque para el presente proceso de indagación es la *Simplificación de la estructura de datos* la cual consiste en “Tratar de encontrar una representación reducida del espacio de las variables en estudio, mediante la transformación de algunas variables a un conjunto de menor dimensión” (Díaz, 2007). Un enfoque diferente, que en todo caso no es la intención de este trabajo, sería, por ejemplo, estudiar la dependencia o interdependencia entre las variables.

5.3 Tabla de contingencia

Un arreglo de doble entrada en donde se clasifican individuos de acuerdo con dos o más variables, usualmente nominales, se denomina tabla de contingencia. Pardo (2020) menciona que para describir la asociación de dos variables cualitativas o cuantitativas se construye una tabla de contingencia (en adelante TC), que clasifica a los individuos por las categorías de las dos variables simultáneamente. Disponer los datos de esta manera (filas y columnas) corresponde a una representación matricial.

Para efectos de medición de las variables es usual referirse a las escalas nominal, ordinal, de intervalo y razón. En el caso de las variables cualitativas, las mediciones que responden exclusivamente a atributos se ubican en el nivel más elemental de medición: el nominal (Díaz, 2007). En algunas ocasiones se utilizan códigos numéricos para representar dichos atributos, pero en esos casos los números empleados se consideran solo como “etiquetas”. Estos códigos numéricos se asignan a los objetos con el propósito de clasificarlos, pero no poseen el significado numérico usual, aparte de la relación de igualdad. Lo anterior significa que tienen una naturaleza no-métrica.

Para propósitos de este trabajo, se utilizará la notación usada en Lebart, (citado por Pardo, 2020) en donde **K** (en negrilla) representa una TC; k_{ij} , su término general; $k_{i\cdot}$, la suma de su fila i (marginal fila), $k_{\cdot j}$, la suma de su columna j (marginal columna), y $k_{\cdot \cdot} = k$, la suma total de todos los datos. A continuación, se puede ver un arreglo matricial que correspondería con una TC, en donde la última fila y columna, son los marginales fila y columna respectivamente.

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} & \cdots & k_{1n} & \cdots & k_{1\cdot} \\ k_{21} & k_{22} & k_{23} & \cdots & k_{2n} & \cdots & k_{2\cdot} \\ k_{31} & k_{32} & k_{33} & \cdots & k_{3n} & \cdots & k_{3\cdot} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ k_{m1} & k_{m2} & k_{m3} & \cdots & k_{mn} & \cdots & k_{m\cdot} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{\cdot 1} & k_{\cdot 2} & k_{\cdot 3} & \cdots & k_{\cdot n} & \cdots & k_{\cdot \cdot} \end{bmatrix}$$

Al dividir cada elemento de la tabla por el número total de individuos, se tendrá la *tabla de frecuencias relativas*, que es la distribución conjunta de probabilidades de las dos variables a describir. La suma de las filas o columnas de esta tabla se denominan *distribuciones marginales*.

Entonces, en concordancia con la notación utilizada por Pardo (2020) la tabla de frecuencias relativas se resume en la matriz **F** y el término general se define como $f_{ij} = \frac{k_{ij}}{k}$.

Asimismo, la marginal fila tendrá como notación $f_{i\cdot} = \frac{k_{i\cdot}}{k}$ y la marginal columna $f_{\cdot j} = \frac{k_{\cdot j}}{k}$.

Enseguida se puede ver una representación general de cómo se vería \mathbf{F} .

$$\begin{bmatrix} \frac{k_{11}}{k} & \frac{k_{12}}{k} & \dots & \frac{k_{1n}}{k} & \dots & \frac{k_{1\cdot}}{k} \\ \frac{k_{21}}{k} & \frac{k_{22}}{k} & \dots & \frac{k_{2n}}{k} & \dots & \frac{k_{2\cdot}}{k} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ \frac{k_{m1}}{k} & \frac{k_{m2}}{k} & \dots & \frac{k_{mn}}{k} & \dots & \frac{k_{m\cdot}}{k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{k_{\cdot 1}}{k} & \frac{k_{\cdot 2}}{k} & \dots & \frac{k_{\cdot n}}{k} & \dots & \frac{k_{\cdot \cdot}}{k} \end{bmatrix}$$

5.4 Modelo de Independencia

En la sección 6.1, se trabajará con frecuencias relativas para desarrollar el AC. Cuando se tienen dos eventos independientes A y B , la probabilidad de que ocurran ambos será igual al producto de cada probabilidad. En otras palabras:

$$P(A \cap B) = P(A) \times P(B)$$

Este es el modelo de independencia, el cual es aplicable al caso de la matriz \mathbf{F} . De acuerdo con Husson (2016), la probabilidad conjunta de variables cualitativas independientes es igual al producto de las probabilidades marginales. Conectando con la notación introducida para las tablas de contingencia, la relación entre la probabilidad conjunta y las probabilidades marginales se expresaría como:

$$\forall i, \forall j, \quad f_{ij} = f_{i\cdot} \times f_{\cdot j}$$

Otra forma de escribir esto es la siguiente:

$$\frac{f_{ij}}{f_{i\cdot}} = f_{\cdot j} \quad \frac{f_{ij}}{f_{\cdot j}} = f_{i\cdot}$$

Esto quiere decir que la probabilidad condicional es igual a la marginal ya sea para el caso de las filas o las columnas y la igualdad significaría la independencia entre las variables.

5.5 Técnicas multivariables

Según Diaz (2007) las técnicas de análisis multivariado se diferencian de acuerdo con el trabajo por filas y/o columnas. Además, los enfoques que encierran a la mayoría de las técnicas son los de dependencia e interdependencia.

En el caso del enfoque de dependencias se tienen técnicas multivariadas como la regresión múltiple, el análisis discriminante, el análisis de correlación canónica y el análisis conjunto. Para el caso del enfoque de interdependencia, que también es un enfoque que se utiliza en el presente trabajo de grado, se tiene el análisis de componentes principales, el análisis de factores comunes y el AC.

Es importante distinguir acerca del Análisis de Componentes Principales (ACP) que, de acuerdo con Pardo (2020), “cada uno de los métodos en ejes principales se puede ver como un ACP de una matriz X que contiene los datos a analizar transformados de acuerdo al respectivo método”. Por lo que para poder explicar la técnica utilizada para el análisis de los datos (AC) primero hay que entender el ACP.

5.6 Consideraciones del Análisis de Componentes Principales (ACP)

Como herramienta experimental para analizar los datos que se consideraron acerca del contexto de violaciones a los DDHH, se usará el AC, sin embargo, usar este último involucra algunos elementos del ACP los cuales se comentarán a continuación.

Primero, consideraremos una perspectiva intuitiva, basada en una respuesta del grupo conocido como Cross Validated (ver Cross Validated, 2015), para explicar el ACP como un

método que permite resumir grupos de datos. El ejemplo utiliza datos de un contexto real: tiene que ver con características recolectadas acerca de vinos. Entonces, si se tienen datos de múltiples vinos, se podría crear una lista que resuma dichas características, por ejemplo, en una tabla en donde las filas son cada vino y las columnas indicarían una puntuación para las características, en los diferentes criterios que se puedan medir.

Cada fila sería entonces un vector de tamaño n que resumiría lo que se midió acerca de cada vino, y el arreglo de estos vectores fila, generaría una matriz que resume todos los vinos. Si se tiene eso, sería probable que algunos de estos rasgos de los vinos sean redundantes; por ende, valdría la pena pensar en que es posible describir cada vino con menos rasgos.

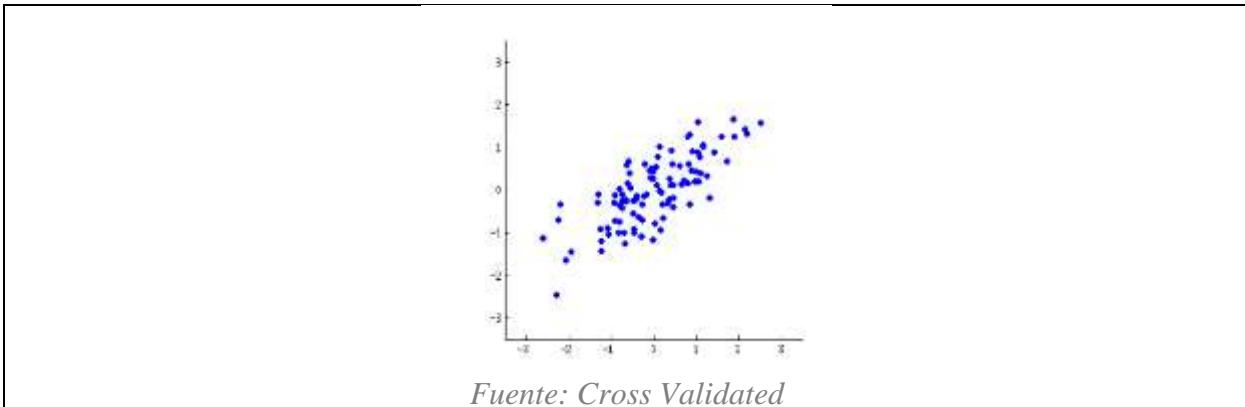
En otras palabras, al tener los datos organizados en una matriz, sería sensato pensar que la transformación lineal que representa puede cambiarse de base, modificando su espacio de las columnas para que tenga un rango menor.

Este acercamiento no implica que se estén seleccionando algunas características mientras que otras se descartan, en cambio el ACP “construye algunas nuevas características que resultan resumiendo la lista de vinos de manera adecuada” (Cross Validated, 2015). Estas nuevas características se construyen con base en los atributos iniciales, en la búsqueda de las mejores combinaciones lineales de todas las concebibles. Por ejemplo, una puede ser “la edad del vino, menos el nivel de acidez” (Cross Validated, 2015).

La mejor síntesis de los datos, en este caso de una hipotética lista de vinos, se consigue teniendo en cuenta dos criterios acerca de cuál sería el mejor resumen: garantizar la máxima variabilidad de los datos y minimizar el error de reconstrucción de los datos. Para explicarlos mejor, se usará el ejemplo presentado en Cross Validated (2015) que se visualiza en la fig. 24.

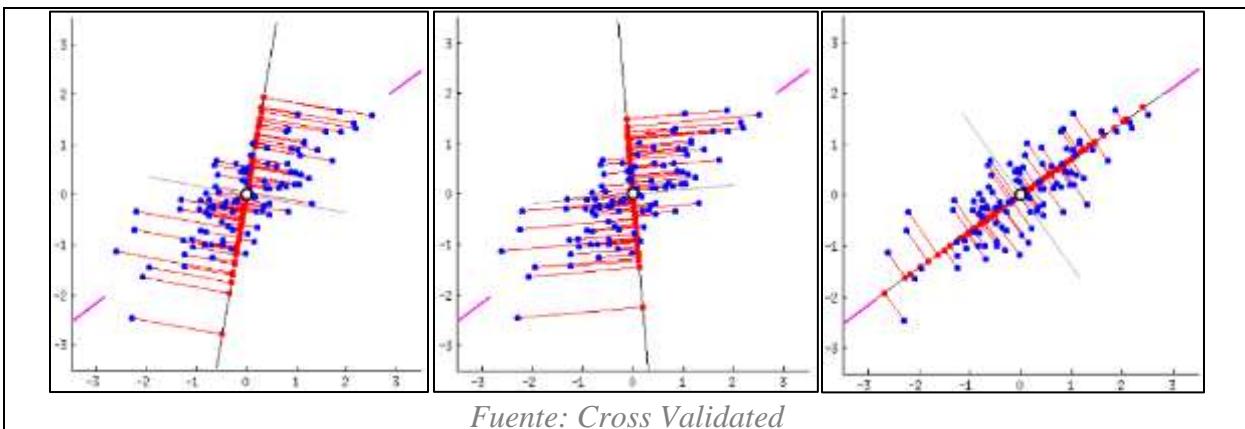
Allí se resumen dos características de los vinos, la oscuridad y el contenido de alcohol, en el eje x y el eje y , respectivamente. También se asume que están correlacionados, como se puede intuir por la organización de la nube de puntos azules, que representan cada vino.

Figura 24. Oscuridad del vino y cantidad de alcohol



“Una nueva propiedad puede ser construida dibujando una recta a través del centro de la nube de puntos y proyectando todos los puntos en esta recta” (Cross Validated, 2015). Por supuesto esa propiedad sería una combinación lineal de la forma $w_1x + w_2y$. En la fig. 25, se pueden ver varias opciones para esa recta. Los puntos rojos serían la proyección de los azules.

Figura 25. Proyectando nube de puntos en una recta, tres opciones



El primer criterio, el cual ha de garantizar que la variación de los valores a lo largo de la recta sea máxima, se cumple en la imagen de la derecha de la fig. 25. En las otras dos, los puntos rojos quedan demasiado juntos, indicando que serían opciones descartables. Por otro lado, minimizar el error de reconstrucción, dependerá de poder obtener la característica original, representada por los puntos azules, a partir de la nueva, proporcionada por los rojos. Este error de reconstrucción será entregado por el largo del segmento rojo que conecta los dos puntos. La menor distancia entre los dos puntos también ocurre en la tercera imagen de la fig. 25; es decir que acontece al mismo tiempo que la búsqueda de la máxima varianza. Esta recta, la cual cumple los dos criterios, denominada el primer componente principal, “corresponde con la nueva propiedad acerca de los vinos que será construida con un ACP” (Cross Validated, 2015).

Matemáticamente, la variabilidad de los puntos rojos en la fig. 25 es medida como la distancia promedio al cuadrado desde el centro de la nube de puntos hasta cada punto rojo; comúnmente esto se define como la varianza. Por otro lado, el error de reconstrucción total se mide como el largo promedio al cuadrado de los segmentos rojos. Debido a que la proyección de cada punto azul sobre el primer componente principal implica que cada segmento rojo forme un ángulo de 90° con la recta de la tercera imagen de la fig. 25, la suma de las dos cantidades derivadas de los dos criterios mencionados sería igual a la distancia promedio al cuadrado entre el centro de la nube de puntos y cada punto azul. Esto es precisamente el teorema de Pitágoras validando el hecho de que los dos criterios se alcanzan al mismo tiempo (Cross Validated, 2015).

Finalmente, es importante mencionar que conseguir los componentes principales por medio de un ACP, involucra, como se pudo entrever, utilizar valores y vectores propios. Debido a que las matrices derivadas de los datos e involucradas en alcanzar cualquiera de los dos

criterios del ACP, se quieren representar en menor dimensión; algo para lo que son útiles los vectores propios.

En el caso del ejemplo de Cross Validated (2015), la matriz de interés es la matriz 2×2 de covarianza, que para el caso del ejemplo citado sería la siguiente:

$$\begin{pmatrix} 1.07 & 0.63 \\ 0.63 & 0.64 \end{pmatrix}$$

Esta matriz indica que la varianza para la variable x es 1.07 y la de la variable y es 0.64. La covarianza entre las dos sería 0.63. Debido a que es una matriz cuadrada, puede ser diagonalizada, consiguiendo que los valores propios queden sobre la diagonal de una nueva matriz, la cual representaría un nuevo sistema coordenado ortogonal; determinado por los vectores propios (esto fue explicado a través del cambio de base en la sección 4.7). En este nuevo sistema coordenado la matriz luciría así:

$$\begin{pmatrix} 1.52 & 0 \\ 0 & 0.19 \end{pmatrix}$$

Aquí se vuelve claro que la varianza de cualquier proyección será entregada por el promedio ponderado de los valores propios. Consecuentemente, la máxima posible varianza (1.52) será alcanzada si simplemente se toma la proyección sobre el primer eje coordenado. La dirección de ese primer eje (primer componente principal) será entregada por el primer vector propio de la matriz de covarianza. En las tres imágenes de la fig. 25, la dirección de ese primer componente principal está indicada por los pequeños segmentos color magenta. Como se puede ver, la recta de la tercera imagen coincide con dicha dirección. Según el ejemplo citado, estos segmentos marcan la dirección del primer vector propio de la matriz de covarianza que en este

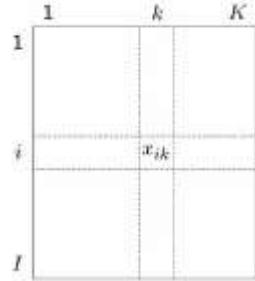
caso es igual a $(0.81, 0.58)$. Ese primer vector propio, terminaría conservando un gran porcentaje de la variabilidad original de la nube de puntos, solo que en una dimensión menor.

5.6.1 El ACP según algunos autores

El ACP siendo una técnica de análisis multivariado se aplica en tablas de datos (filas y columnas) en las cuales se tienen a los “individuos”, como comúnmente se denominan, en las filas y a las “variables” cuantitativas acerca de estos en las columnas. Según Pardo (2020, p. 41) los objetivos del ACP son: (1) comparar los individuos entre sí; las gráficas que se obtienen permiten observar la forma de la “nube de individuos”, lo que a su vez posibilita detectar patrones en ellos, (2) describir las relaciones entre las variables y (3) reducir la dimensión de la información presentada asociada al conjunto de datos (matriz). A mayor relación entre las variables mayor es la capacidad de síntesis del ACP y unos pocos ejes factoriales podrán resumir las variables originales con cierto grado de precisión.

Aquí conviene mencionar que este método multivariado descriptivo, busca una mejor representación para los datos iniciales que podrían verse en una tabla como la de la fig. 26, en la cual se consideran I individuos y K variables:

Figura 26. Tabla de individuos y variables cuantitativas ACP

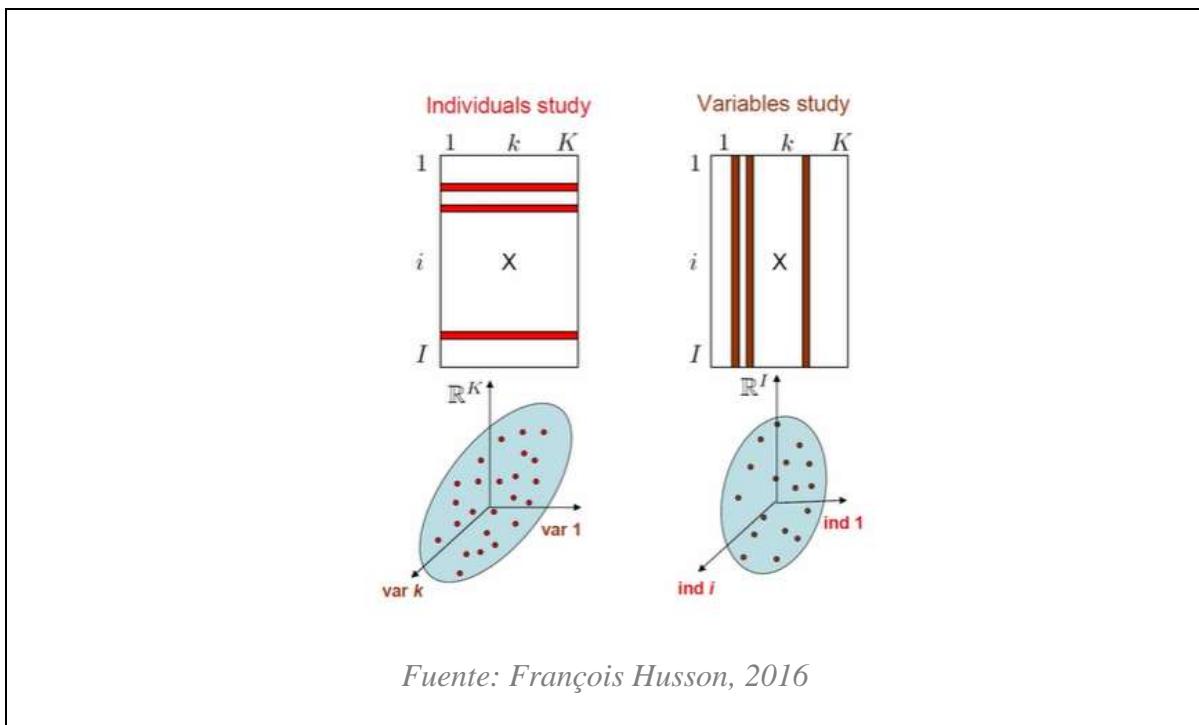


Fuente: François Husson, 2016

Algunos casos en los cuales se pueden identificar disposiciones de los datos en la manera descrita se pueden ubicar en varios campos, por ejemplo: “análisis sensorial: puntaje de los atributos k de un producto i ; ecología: concentración de contaminantes k en el río i ; sociología: tiempo gastado en una actividad k por los individuos de la clase social i ” (Husson, 2016).

Esta disposición matricial de los datos permite verlos como dos nubes de puntos: primero, una en la que a cada individuo I le corresponde un vector que pertenece al espacio \mathbb{R}^k , entonces sería posible pensar en una nube de puntos de individuos (N_I); por otro lado, tomando las variables, se tendrían vectores en el espacio \mathbb{R}^I , obteniendo una dispersión de los datos desde esta perspectiva (N_k). Ambos casos se ilustran en la fig. 27.

Figura 27. Dos perspectivas para estudiar una tabla de datos: con I individuos y K variables



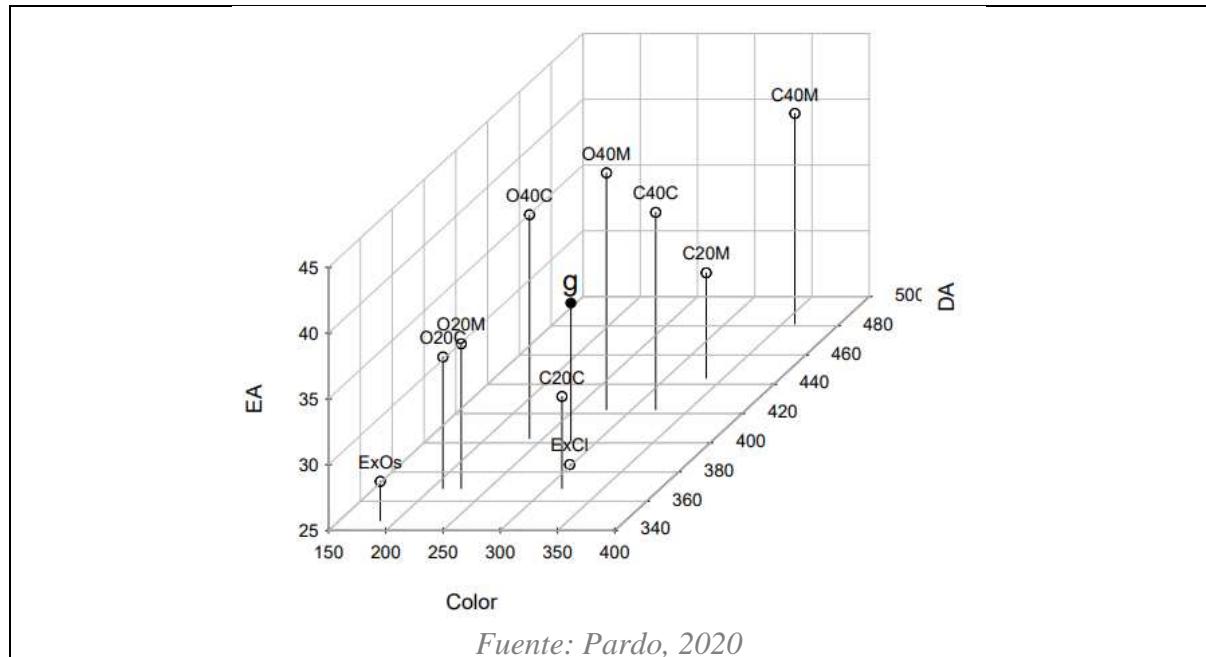
Si se analiza N_I , interesa investigar, por ejemplo, las similitudes entre los individuos teniendo en cuenta todas las variables; mientras que al hablar de N_k , el enfoque podría ser “interpretar la similitud en términos de los valores de las variables” (Husson, 2016), esto último se traduce en averiguar relaciones lineales entre estas, las cuales se verían representadas en un coeficiente de correlación entre pares de variables.

En cualquiera de los dos casos descritos, el ACP apunta a reducir la dimensión de la representación de los datos para favorecer la visualización; en otras palabras, ayudar a interpretar mejor los datos sin mucha pérdida de información. En el caso de las filas, “busca el mejor espacio resumido para una óptima visualización de N_I ”. (Husson, 2016), esto significa buscar un subespacio \mathbb{R}^p , tal que $p < k$, en el cual se retenga la mayor cantidad de

información de la nube de individuos inicial. Esto configuraría una posterior interpretación más sencilla de las relaciones entre los individuos con el uso de variables sintéticas (en combinaciones lineales)⁵.

Una manera de ilustrar esta idea de reducir la dimensión, específicamente para los individuos, es pensar que son descritos por una matriz de rango 3 ($K = 3$). En ese caso, aún sería posible visualizar los vectores que conforman N_I y tener una idea de la dispersión de los datos. De Pardo (2020, p. 43) se extrae un ejemplo en el cual se miden rasgos de diez cafés (en la fig. 28, por ejemplo, C20M es: tostación clara, 20% agregado de maíz) en tres variables: color, DA: densidad aparente, EA: extracto acuoso (ver fig. 28). Cada café sería un vector en \mathbb{R}^3 y en un arreglo matricial, se obtendría una matriz 10×3 que llamaremos \mathbf{Y} .

Figura 28. Nube de individuos, Y



⁵ La perspectiva del análisis para la nube de las variables (N_k) no se estudiará a profundidad, pues no está involucrada en el Análisis de Correspondencias simple.

El algoritmo común para desarrollar un ACP necesita un preprocesamiento de los datos. Esto se hace a través de los siguientes pasos:

1. Determinar el centro de gravedad de la nube de puntos (N_I): esto es la generalización del concepto de la media, como una medida de localización multivariada. Para el caso de \mathbf{Y} se tendría que el centro de gravedad \mathbf{g} es

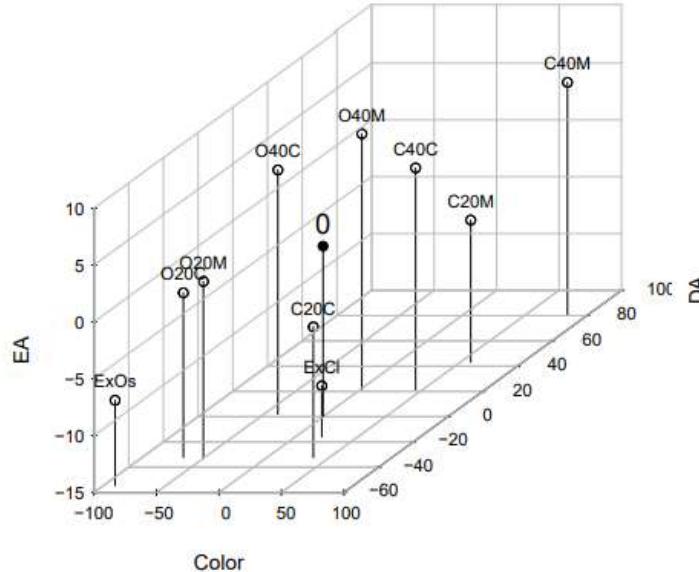
$$\mathbf{g} = \frac{1}{n} \sum_{i=1}^I \mathbf{y}_i$$

Donde \mathbf{y}_i es cada café de \mathbf{Y} y $n = 10$

2. Centrado de la nube de individuos: según Pardo (2020, p. 44) esto “permite trasladar el cero de la representación al centro de gravedad”. Las coordenadas de un individuo centrado \mathbf{yC}_i , se consiguen restándole las coordenadas de \mathbf{g} a cada \mathbf{y}_i : $\mathbf{yC}_i = \mathbf{y}_i - \mathbf{g}$.

La matriz de datos centrados sería: $\mathbf{Y}_C = \mathbf{Y} - \mathbf{1}_n \mathbf{g}'$ (ver fig. 29). Si \mathbf{Y}_C se compara con la nube de individuos de la fig. 28, la distribución no cambió, solo las coordenadas. Esto se debe a que \mathbf{Y}_C tiene la misma distribución de \mathbf{Y} , pero “las coordenadas en cada eje representan la diferencia de un café con respecto al café típico” (Pardo, 2020, p. 44) representado por \mathbf{g} .

Figura 29. Nube de puntos centrada \mathbf{Y}_C



Fuente: Pardo, 2020

3. Determinar la inercia de la nube de individuos: esta se utiliza como medida de dispersión de la nube de puntos N_I . Para conseguirla se habla primero de la distancia euclíadiana al cuadrado entre dos individuos: $d^2(i, l) = \sum_{j=1}^k (\mathbf{y}_{ij} - \mathbf{y}_{lj})^2$; esto es un resultado de poder generalizar a más de 2d la distancia entre dos vectores. El valor de la distancia de cada vector centrado a \mathbf{g} , permitiría calcular la inercia a partir de los datos centrados así:

$$\text{Inercia } (N_I) = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^K (\mathbf{y}_{cij} - \mathbf{g})^2$$

Como el centro de gravedad para datos centrados es igual a 0, entonces se tiene

$$\text{Inercia } (N_I) = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^K \mathbf{y}_{cij}^2 = \sum_{j=1}^K \frac{1}{n} \sum_{i=1}^I \mathbf{y}_{cij}^2 = \sum_{j=1}^K \sigma_j^2$$

“Esta fórmula permite ver que la inercia es la suma de las varianzas de las variables, por lo que estás influyen en el análisis en proporción a su varianza” (Pardo, 2020, p. 48). Al tener los datos centrados la nube de puntos está en su forma original, como se mencionó, pero el hecho de tenerlos centrados permite describir la inercia. Esto habilita la posibilidad de descomponerla en nuevos ejes coordenados con el ACP, siguiendo el criterio de maximizar la varianza.

4. El último paso se realiza si las diferentes características medidas de los individuos están en diferentes escalas o unidades de medidas. Según Pardo (2020, p. 49) si esto ocurre, como en el ejemplo citado, sería necesario estandarizar las medidas. La matriz normalizada \mathbf{X} , que sería la matriz \mathbf{Y} centrada y estandarizada, tiene como término general:

$$x_{ik} = \frac{y_{ik} - \bar{y}_k}{\sigma_k}$$

Donde \bar{y}_k y σ_k son respectivamente la media y la desviación estándar de la variable k . En términos matriciales \mathbf{X} se obtiene mediante

$$\mathbf{X} = \mathbf{Y}_c \mathbf{D}_\sigma^{-1}$$

Donde $\mathbf{D}_\sigma = diag(\sigma_k)$. Si se estandarizan los datos, “el valor que un individuo asume para una variable es la diferencia con respecto al promedio, pero medida en el número de desviaciones estándar” (Pardo, 2020, p. 49).

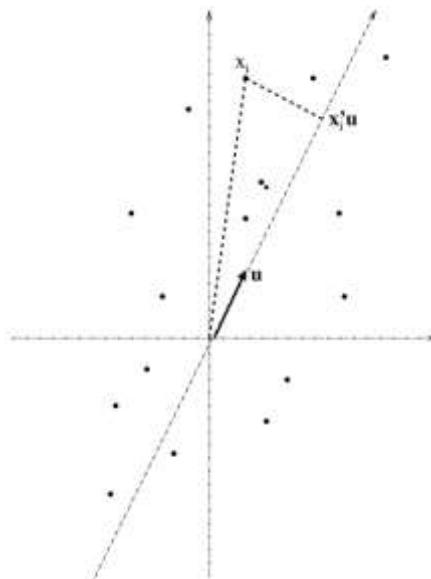
Ahora, para formalizar la noción mencionada en el acercamiento intuitivo y llegar a encontrar el mejor eje sobre el cual proyectar los datos (ver fig. 25), los autores consultados escogen el camino que garantiza la máxima varianza de las proyecciones. Para esto, interesa

encontrar un vector \mathbf{u} (unitario, es decir $\mathbf{u}\mathbf{u}' = 1$) y su dirección, de tal manera que cuando se proyecten los datos sobre \mathbf{u} , en esa dirección, los puntos proyectados varíen lo máximo posible.

Según Pardo (2020, p.52) “el objetivo geométrico de los métodos en ejes principales es buscar un nuevo sistema de ejes, de tal manera que la mayoría de la inercia se concentre en los primeros ejes”. Esto tiene que ver con descomponer la inercia que estaría representada por \mathbf{X} en ejes ortogonales, de tal manera que la mayor inercia esté en el primer eje; en el segundo, la mayor inercia residual, etc.

Si denota \mathbf{u} al vector unitario que da la dirección del nuevo eje, la coordenada de un vector individuo \mathbf{x}_i (de la matriz \mathbf{X}), sobre este eje, es el producto punto $\langle \mathbf{x}_i, \mathbf{u} \rangle = \mathbf{x}_i' \mathbf{u}$ (Pardo, 2020, p. 52). En la fig. 30, se puede ver una aproximación a cómo luciría tal proyección.

Figura 30. Proyección del vector \mathbf{x}_i' sobre \mathbf{u}



Fuente: Pardo, 2020

Ahora, la contribución a la inercia del individuo i sobre el eje \mathbf{u} es $\frac{1}{n}(\mathbf{x}_i' \mathbf{u})^2$. La inercia total de la nube de individuos, proyectada sobre el eje \mathbf{u} es entonces.

$$\sum_{i=1}^I \frac{1}{n} (\mathbf{x}_i' \mathbf{u})^2 = \sum_{i=1}^I \frac{1}{n} (\mathbf{u}' \mathbf{x}_i) (\mathbf{x}_i' \mathbf{u}) = \frac{1}{n} \mathbf{u}' \left[\sum_{i=1}^I (\mathbf{x}_i) (\mathbf{x}_i') \right] \mathbf{u}$$

Dado que \mathbf{X} representa los datos centrados y estandarizados, sería valido expresar la última sumatoria así:

$$\mathbf{u}' \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u}$$

De acuerdo con Pardo (2020, p. 53) “encontrar el eje de mayor inercia proyectada equivale a encontrar la dirección \mathbf{u} que maximice $\mathbf{u}' \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u}$ sujeto a la restricción $\mathbf{u} \mathbf{u}' = 1$. Una manera de resolver este problema es introduciendo un multiplicador de Lagrange⁶ λ . Entonces se debe maximizar:”

$$f(\mathbf{u}) = \mathbf{u}' \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u} - \lambda (\mathbf{u} \mathbf{u}' - 1)$$

Esto se hace a través de la derivación y la búsqueda de puntos críticos⁷, que en este caso serían los puntos máximos de la función, es decir, según Pardo (2020, p. 53):

$$f'(\mathbf{u}) = 2 \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u} - 2\lambda \mathbf{u} = 0$$

Entonces se tiene:

$$\frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u} = \lambda \mathbf{u}$$

⁶ Para una revisión rápida del multiplicador de Lagrange en el análisis de datos, se puede consultar el siguiente video del usuario ritvikmath en YouTube: [Lagrange Multipliers : Data Science Basics - YouTube](#)

⁷ Para una revisión rápida de la derivación de matrices en el análisis de datos, se puede consultar el siguiente video del usuario ritvikmath en YouTube: [Derivative of a Matrix : Data Science Basics - YouTube](#)

Las soluciones de $\frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u} = \lambda \mathbf{u}$ son los vectores propios unitarios asociados a los valores propios de $\frac{1}{n} \mathbf{X}' \mathbf{X}$. Para saber qué valores propios escoger, se premultiplica por \mathbf{u}'

$$\mathbf{u}' \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u} = \lambda \mathbf{u}' \mathbf{u} = \lambda$$

Lo cual nos deja a la izquierda precisamente el valor que queríamos maximizar. Esto significa que habría que escoger el mayor valor de λ , siendo los vectores propios derivados de este, la primera dirección en la cual se puede descomponer la inercia. Este primer componente principal se denota como λ_1 y su vector propio \mathbf{u}_1 . Las coordenadas de los individuos sobre el eje generado por \mathbf{u}_1 , se denotan por \mathbf{F}_1 y son $\mathbf{F}_1 = \mathbf{X} \mathbf{u}_1$ (Pardo, 2020, p.53).

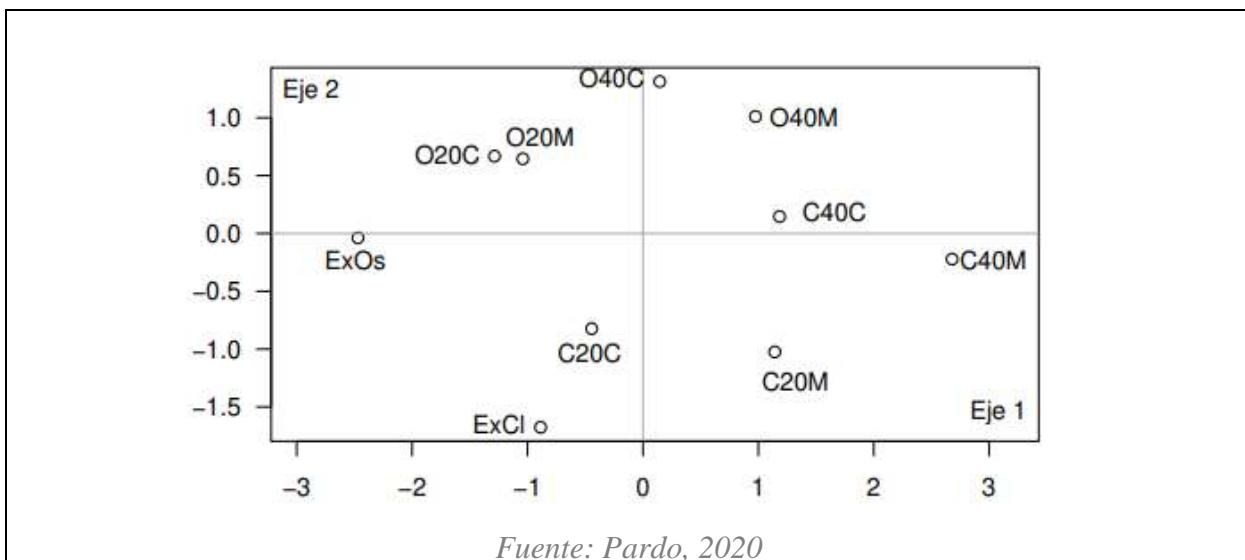
Un proceso similar se realiza para encontrar “el mejor plano de proyección de la nube de puntos” (Pardo, 2020, p. 53), tomando nuevamente la tarea de maximizar $\mathbf{u}' \frac{1}{n} \mathbf{X}' \mathbf{X} \mathbf{u}$ sujeto a la restricción $\mathbf{u} \mathbf{u}' = 1$ y $\mathbf{u}' \mathbf{u}_1 = 0$, lo cual garantiza que los ejes sean ortogonales. Esto implica introducir otro multiplicador de Lagrange y realizar un proceso análogo al descrito, del cual se concluye que ese segundo componente principal, sería entregado por el segundo valor propio más grande de $\frac{1}{n} \mathbf{X}' \mathbf{X}$.

La obtención de valores y vectores propios es un problema básico de álgebra lineal, pero para su cálculo es necesario utilizar métodos numéricos. Los programas de cálculo matemáticos y estadísticos incluyen funciones para obtener los valores y vectores propios de una matriz. En R se calculan con la función *eigen* (Pardo, 2020, p. 54).

En el libro de Pardo (2020), hay ayudas y paquetes de R para graficar las proyecciones de la matriz discutida, sobre los componentes principales. Estás se usarán en el AC simple de los datos de violaciones a los DDHH.

El resultado de aplicar un ACP en el ejemplo del autor acerca de los cafés está en la fig. 31. Allí se puede ver, comparado con la fig. 28, que ahora cada café está representado por un vector en dos dimensiones. Esto permite ver más claramente las distancias entre cada uno, y gracias a esto, describir el parecido entre individuos.

Figura 31. Primer plano factorial del ACP acerca de los diez cafés



Finalmente es importante mencionar que en el texto de Pardo (2020, p. 89) hay un capítulo dedicado a generalizar el ACP. Allí se determinan las condiciones para establecer la siguiente tripleta ACP (X, M, N). Según Pardo (2020, p. 89) se nota M a “la matriz diagonal de métrica en el espacio de las filas y de pesos en el de las columnas”, y N “a la matriz diagonal de pesos de las filas y de métrica en el espacio de las columnas”. La matriz X está centrada con los pesos dados en N .

Por ejemplo, el ACP extraído de Pardo (2020) para las explicaciones de esta sección, se podría expresar de manera generalizada así: ACP (Y_c , $diag\left(\frac{1}{\sigma_j^2}\right)$, $\frac{1}{n}I_n$). La matriz por analizar es la de datos centrados Y_c , la métrica $diag\left(\frac{1}{\sigma_j^2}\right)$ es la matriz diagonal de las inversas de las varianzas y $\frac{1}{n}I_n$ es la matriz de pesos, donde I_n es la matriz identidad de dimensión n .

A lo largo del capítulo para explicar el ACP generalizado, el autor, propendiendo por expandir el método multivariado a cualquier caso, introduce términos de álgebra lineal como el producto interno. Estos son necesarios en las justificaciones de los pasos del preprocesamiento de los datos descritos anteriormente, junto con la posterior búsqueda de los ejes de proyección de la nube de individuos y de variables. Estás explicaciones no hacen parte de lo que se espera abarcar con el presente trabajo. Sin embargo, es ineludible utilizar los resultados del ACP generalizado para la nube de individuos en el AC de la siguiente sección. Por este motivo se extrae la fig. 32 del libro en la cual se resumen los elementos principales para realizar un ACP (X, M, N) sobre el espacio de los individuos.

Figura 32. Fórmulas para el ACP (X, M, N)

Espacio	\mathbb{R}^p
Nube	N_n
Coordenadas	Filas de X : x^t_j
Pesos	Diagonal de N : p_i
Métrica	M
Distancias al cuadrado	$d^2(i, l) = \sum_{j=1}^p m_j(x_{ij} - x_{lj})^2$
Inercia	Trazo($X'N X M$)
Valor propio	λ_s
Vector propio	u_s
Fórmula valor-vector propio	$X'N X M u_s = \lambda_s u_s$

Fuente: Pardo, 2020

Resultados

5.7 Análisis de Correspondencias simple (AC)

El acercamiento de cualquier texto que trabaje análisis multivariados, en la mayoría de los casos se apoya de un ejemplo para realizar las explicaciones. Siguiendo esa línea, en esta sección se explicará el AC usando como base los datos de la TC final de la sección 3.4.3.

De acuerdo con Pardo (2020, p. 109) “El objetivo del AC es describir las asociaciones entre las variables fila y columna, a través de sus perfiles. Para ello se considera:

1. Comparar los perfiles fila.
2. Comparar los perfiles columna.

3. Estudiar las correspondencias entre perfiles fila y columna”.

Al ver la posibilidad de organizar los datos de la revista en una TC, el método más apropiado resultó ser el AC, porque ayuda a crear una representación geométrica de las columnas y las filas. Con este tipo de representación, se busca poder describir las asociaciones y poder llegar a conclusiones a partir del contexto. Utilizando la notación introducida en la sección 5.3, K es la matriz que representa la TC y se puede ver en la Tabla 13.

Tabla 13. TC de violaciones a los DDHH y presuntos responsables con marginales

	PR2	PR4	PR7	PR14	PR25	PR28	PR33	PR35	PR37	Marginal fila
A	MUERTOS	0	10	33	59	39	10	12	224	9
D	DESAPARICIÓN	0	4	0	6	12	0	0	0	0
F	TORTURA	0	11	16	16	1	0	1	10	0
I	LESIONADOS	6	11	70	6	16	1	2	44	18
L	DETENCIÓN ARBITRARIA	1	39	28	2	0	0	0	0	70
M	AMENAZA	1	48	23	251	63	7	12	91	0
Marginal columna		8	123	170	340	131	18	27	369	27
Marginal fila										
2426										

Fuente: elaboración propia

El siguiente paso en el AC, corresponde con la determinación de las frecuencias relativas. Esto generaría la matriz F (ver sección 5.3) y sus sumas “que representan la distribución de probabilidad conjunta y las distribuciones marginales, respectivamente, expresadas en porcentajes” (Pardo, 2020, p. 110). En la Tabla 15, se puede ver la tabla de frecuencias relativas.

Tabla 14. Tabla de frecuencias relativas F

	PR2	PR4	PR7	PR14	PR25	PR28	PR33	PR35	PR37	Marginal fila
A	MUERTOS	0,0000	0,4122	1,3603	2,4320	1,6076	0,4122	0,4946	9,2333	0,3710
D	DESAPARICIÓN	0,0000	0,1649	0,0000	0,2473	0,4946	0,0000	0,0000	0,0000	0,9068
F	TORTURA	0,0000	0,4534	0,6595	0,6595	0,0412	0,0000	0,0412	0,4122	0,0000
I	LESIONADOS	0,2473	0,4534	2,8854	0,2473	0,6595	0,0412	0,0824	1,8137	0,7420
L	DETENCIÓN ARBITRARIA	0,0412	1,6076	1,1542	0,0824	0,0000	0,0000	0,0000	0,0000	2,8854
M	AMENAZA	0,0412	1,9786	0,9481	10,3462	2,5969	0,2885	0,4946	3,7510	0,0000
Marginal columna		0,3298	5,0701	7,0074	14,0148	5,3998	0,7420	1,1129	15,2102	1,1129
Marginal fila										
100,0000										

Fuente: elaboración propia

La marginal fila de la Tabla 14 representa la distribución de frecuencias relativas de las violaciones a los DDHH consideradas y la marginal columna es la distribución de los presuntos responsables. Se puede ver, por ejemplo, que los 6 lesionados por la fuerza pública (PR2) representan el $f_{41} = 0,2473\%$ de todas las categorías de victimización; el $f_{1.} = 16,3232\%$ son Muertos y el $f_{.2} = 5,0701\%$ de los casos registrados se le atribuyen al Ejército (PR4).

Para cada categoría de victimización se tiene una distribución de frecuencias relativas entre los nueve presuntos responsables, que se denomina *distribución condicional o perfil fila*. “Se obtiene al dividir cada celda de la respectiva fila por la suma de la fila, en la TC o en la de frecuencias relativas” (Pardo, 2020). Un perfil fila i se nota $\left\{ \frac{f_{ij}}{f_{i.}} ; j = 1, \dots, p \right\}$. Matricialmente el conjunto de perfiles fila se notan y calculan así $\mathbf{D}_n^{-1} \mathbf{F}$, en donde $\mathbf{D}_n = diag(f_{i.})$ (matriz diagonal con los valores de la marginal fila). La Tabla 15 son los perfiles fila. Allí la marginal columna es la distribución de todos los presuntos responsables, en los nueve que se consideraron, sin importar la categoría de victimización.

Tabla 15. Perfiles fila

Perfil fila										Marginal fila
A	MUERTOS	0	0,03	0,08	0,15	0,1	0,03	0,03	0,57	0,02
D	DESAPARICIÓN	0	0,18	0	0,27	0,55	0	0	0	0
F	TORTURA	0	0,2	0,29	0,29	0,02	0	0,02	0,18	0
I	LESIONADOS	0,03	0,06	0,4	0,03	0,09	0,01	0,01	0,25	0,1
L	DETENCIÓN ARBITRARIA	0,01	0,56	0,4	0,03	0	0	0	0	1
M	AMENAZA	0	0,1	0,05	0,51	0,13	0,01	0,02	0,18	0
	Marginal columna	0	0,05	0,07	0,14	0,05	0,01	0,01	0,15	0,01

Fuente: elaboración propia

Mirando ahora las columnas, cada presunto responsable tiene su distribución según las seis categorías de victimización (denominados *condicionales o perfiles columna*). Un perfil columna j se nota $\left\{ \frac{f_{ij}}{f_{.j}} ; i = 1, \dots, n \right\}$. Usando la multiplicación entre matrices se calcularían así:

$\mathbf{F}\mathbf{D}_{\mathbf{p}}^{-1}$, en donde $\mathbf{D}_{\mathbf{p}} = diag(f_{.j})$ (matriz diagonal con los valores de la marginal columna). La Tabla 16 son los perfiles columna. Allí la distribución marginal fila es la distribución de todas las seis categorías de victimización, sin importar el presunto responsable.

Tabla 16. Perfiles columna

Perfil columna										Marginal fila
		PR2	PR4	PR7	PR14	PR25	PR28	PR33	PR35	PR37
A	MUERTOS	0	0,08	0,19	0,17	0,3	0,56	0,44	0,61	0,33
D	DESAPARICIÓN	0	0,03	0	0,02	0,09	0	0	0	0
F	TORTURA	0	0,09	0,09	0,05	0,01	0	0,04	0,03	0
I	LESIONADOS	0,75	0,09	0,41	0,02	0,12	0,06	0,07	0,12	0,67
L	DETENCIÓN ARBITRARIA	0,13	0,32	0,16	0,01	0	0	0	0	0
M	AMENAZA	0,13	0,39	0,14	0,74	0,48	0,39	0,44	0,25	0
	Marginal columna	1	1	1	1	1	1	1	1	1

Fuente: elaboración propia

A partir de las tablas anteriores, es posible obtener una idea inicial acerca de la cercanía de los perfiles, se consigue gracias a la representación mediante diagramas de barra. Usando uno de los paquetes para R disponibles en Pardo (2020), se obtienen la fig. 33 y 34 para ambos perfiles, en los cuales cada barra representa el 100% y mediante colores se diferencian las categorías.

Figura 33. Diagrama de barras perfiles fila

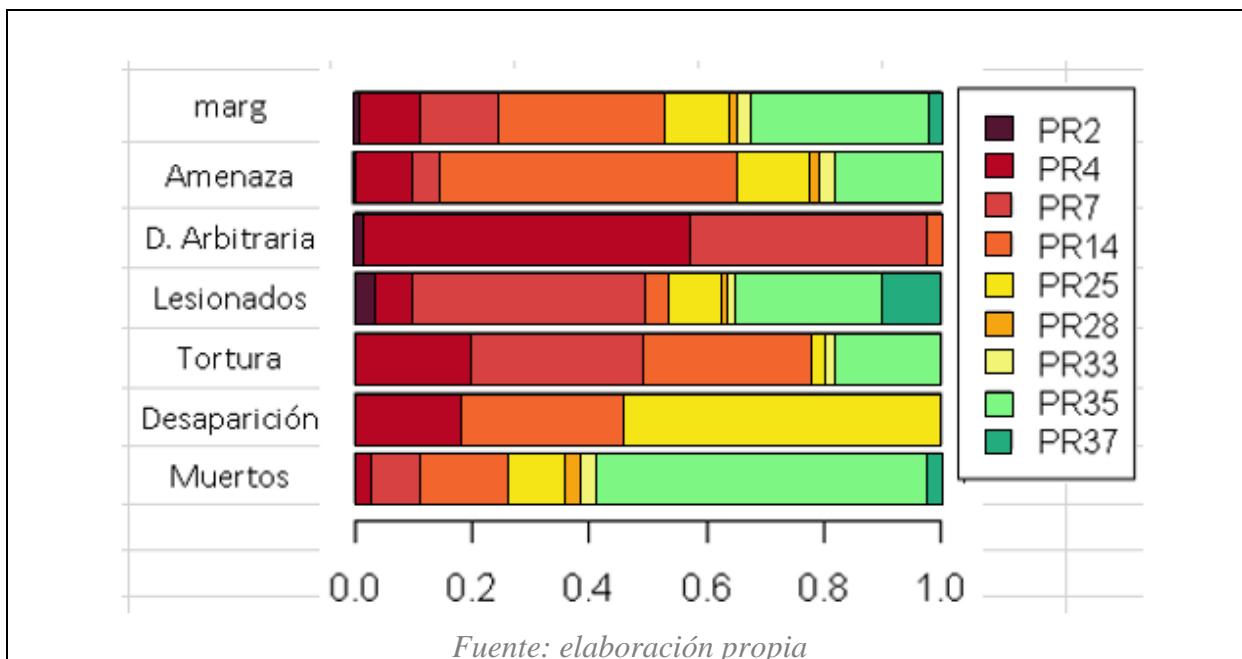
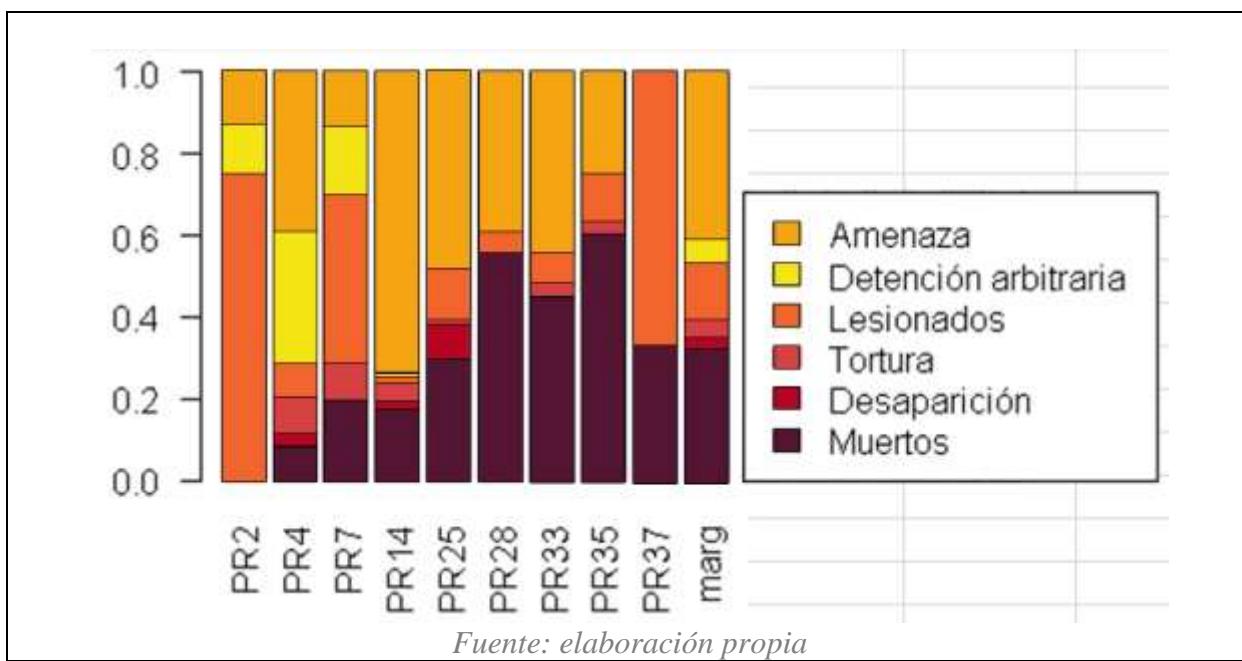


Figura 34. Diagrama de barras perfiles columna



Para la fig. 33 se puede identificar en la categoría de Muertos un alto porcentaje de presunta responsabilidad de la Fiscalía (PR35), a comparación de las demás categorías; también

se pueden resaltar las categorías de Desaparición, con un alto porcentaje para FARC-EP (PR25) y Detención Arbitraria con buena parte de la presunta responsabilidad para el Ejército (PR4).

Según la fig. 33, también se puede decir que carece de porcentaje para la Fiscalía (PR35), en comparación de las demás categorías, en el caso de Detención Arbitraria.

En la fig.34 se puede observar que la Fiscalía (PR35) tiene el mayor porcentaje registrado en la categoría de Muertos, la cual es una categoría presente en todos los presuntos responsables exceptuando a la Fuerza Pública (PR2). En comparación, PR2 parece carecer de varias de las categorías, aunque tiene un gran porcentaje de Lesionados, mayor a la que presentan los lesionados por minas (PR37). Por su parte, los Paramilitares (PR14) resaltan por tener el mayor porcentaje de la categoría Amenaza. La cual aparece en todos los presuntos responsables exceptuando a PR37.

El siguiente paso en el AC es trabajar con las dos tablas de perfiles, debido a que en este análisis multivariado se describen simultáneamente los perfiles fila y columna. Se escoge el acercamiento de hacer dos ACP (X, M, N) para cada tabla de perfiles (Pardo, 2020, p.114). Aunque inicialmente se obtendrán dos salidas, los ACP están relacionados y esto permite “representaciones simultaneas de los planos factoriales” (Pardo, 2020, p.114).

A continuación, se describen los elementos del AC para perfiles fila, considerando n filas y p columnas:

- Coordenadas, pesos: las coordenadas para los perfiles fila son: $x_i, i = 1, 2, \dots, n$ con:

$$x_i(j) = \frac{f_{ij}}{f_i}; i = 1, 2, \dots, p$$

En la Tabla 15, por ejemplo, cuando $i = 6$, se tiene AMENAZAS con coordenadas

$$x_6 = [0.002016129, 0.096774194, 0.046370968, 0.506048387, 0.127016129, 0.014112903, 0.024193548,$$

0.183467742, 0], esta es la distribución de esta categoría de victimización en los 9 presuntos responsables.

Los pesos estarían en la diagonal de D_n , la cual reúne los marginales fila de F . Para AMENAZAS es 20.44%, que sería el porcentaje de registros de violaciones a los DDHH que corresponden con esta categoría.

- Centro de gravedad: según Pardo (2020, p. 115), el centro de gravedad se calcula con los pesos de los n perfiles:

$$g_p = \sum_{i=1}^n f_i x_i$$

La coordenada j , del centro de gravedad sería:

$$g_p(j) = \sum_{i=1}^n f_i \frac{f_{ij}}{f_i} = \sum_{i=1}^n f_{ij} = f_j$$

“Es decir, el centro de gravedad es la marginal columna de la tabla F ” (Pardo, 2020, p. 116). Para nuestro caso $g_p = [0.3298, 5.0701, 7.0074, 14.0148, 5.3998, 0.7420, 1.1129, 15.2102, 1.1129]$, que corresponde a la distribución de los 2426 registros de violaciones a los DDHH entre los nueve presuntos responsables y es el valor típico para comparar los perfiles de las seis categorías de victimización.

Este sería el centro de gravedad que se sitúa en el origen de la representación, como se explicó en la sección anterior. Sin embargo, “para simplificar las fórmulas, el ACP se hace sin centrar y luego se elimina el primer valor propio (que da 1) y el primer vector propio que es el centro de gravedad, esta operación la podemos llamar *centrado a posteriori*” (Pardo, 2020, p. 116).

- Distancia entre perfiles: para este análisis la matriz que genera la métrica es \mathbf{D}_p^{-1} , cuyo elemento diagonal es $\frac{1}{f_{ij}}$. Según Pardo (2020, p. 116) dada esta matriz, la distancia al cuadrado entre dos perfiles fila i y l , es:

$$d^2(i, l) = \sum_{j=1}^p \frac{1}{f_{ij}} (x_{ij} - x_{lj})^2 = \sum_{j=1}^p \frac{1}{f_{ij}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{ij}}{f_{l.}} \right)^2$$

- Inercia de la nube de perfiles fila: está se define de manera similar al ACP debido a que ya se conoce el centro de gravedad:

$$\text{Inercia } (N_n) = \sum_{i=1}^n f_{i.} d^2(i, g_p) = \sum_{i=1}^n f_{i.} \sum_{j=1}^p \frac{1}{f_{ij}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{ij}}{f_{.j}} \right)^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(f_{ij} - f_{i.} f_{.j})^2}{f_{i.} f_{.j}}$$

Búsqueda de los nuevos ejes en el espacio de las filas: en la Tabla 15 el histograma que representa cada perfil se ve como un vector o punto en \mathbb{R}^9 . La diferencia entre dos histogramas se traduce en una distancia entre los puntos que los representan.

Según el autor (Pardo, 2020, p. 114) los pesos de los puntos fila son la distribución marginal y la suma de las filas de \mathbf{F} , que forman la matriz diagonal $\mathbf{D}_n = \text{diag}(f_{i.})$. Por otro lado, las distancias entre distribuciones condicionales descritas anteriormente se definen a partir del producto punto dado por la matriz \mathbf{D}_p^{-1} donde $\mathbf{D}_p = \text{diag}(f_{.j})$. Entonces, las imágenes para comparar los perfiles fila son los planos factoriales derivados del ACP($\mathbf{D}_n^{-1}\mathbf{F}, \mathbf{D}_p^{-1}, \mathbf{D}_n$).

Guiéndonos de las fórmulas de la fig. 32, la matriz de inercia $\mathbf{X}'\mathbf{DXM}$ sería para este caso $\mathbf{F}'\mathbf{D}_n^{-1}\mathbf{D}_n\mathbf{D}_n^{-1}\mathbf{FD}\mathbf{D}_p^{-1} = \mathbf{F}'\mathbf{D}_n^{-1}\mathbf{FD}\mathbf{D}_p^{-1}$ y su término general es

$$\sum_{i=1}^n \frac{f_{ij}}{f_{i\cdot}} \frac{f_{ij'}}{f_{j\cdot}}$$

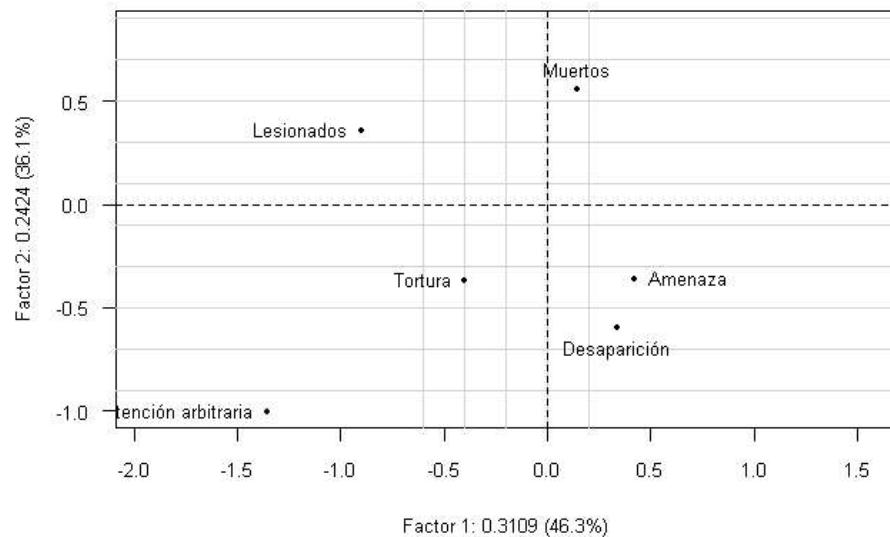
Para garantizar que $\mathbf{F}'\mathbf{D}_n^{-1}\mathbf{F}\mathbf{D}_p^{-1}$ tiene valor propio 1 asociado al centro de gravedad, se pueden obtener los valores propios de

$$\mathbf{F}'\mathbf{D}_n^{-1}\mathbf{F}\mathbf{D}_p^{-1}\mathbf{u}_s = \lambda_s \mathbf{u}_s$$

Usando la función “dudi.coa” del paquete FactoClass proporcionado en Pardo (2020) es posible aplicar el AC sobre los datos representados por la matriz. Entre las salidas se encuentran los 6 primeros valores propios derivados de la igualdad anterior: $\lambda_s = [1, 0.310925403, 0.242357214, 0.078259260, 0.037085643, 0.002847468]$. Como el primer valor propio es 1, este está asociado al centro de gravedad \mathbf{g}_p , “como \mathbf{g}_p es un vector propio, \mathbf{D}_p^{-1} es ortogonal a los demás vectores propios, se puede retirar y las coordenadas de los perfiles fila sobre los ejes factoriales no cambian” (Pardo, 2020, p. 118).

La nube de perfiles fila se observa mediante las proyecciones sobre los ejes y planos factoriales. Según Pardo (2020, p. 91) “un eje factorial es la recta generada por \mathbf{u}_s , uno de los dos vectores propios \mathbf{M} unitarios asociados al valor propio λ_s . Las coordenadas del vector de proyecciones de todas las filas sobre el eje s son $\mathbf{F}_s = \mathbf{X}\mathbf{M}\mathbf{u}_s$ ”. Para nuestro caso $\mathbf{F}_s = \mathbf{D}_n^{-1}\mathbf{F}\mathbf{D}_p^{-1}\mathbf{u}_s$. En la fig. 35 se puede una de las salidas de la función “dudi.coa” en R, en la cual se observa el primer plano factorial para los perfiles fila, derivado de los dos primeros vectores propios.

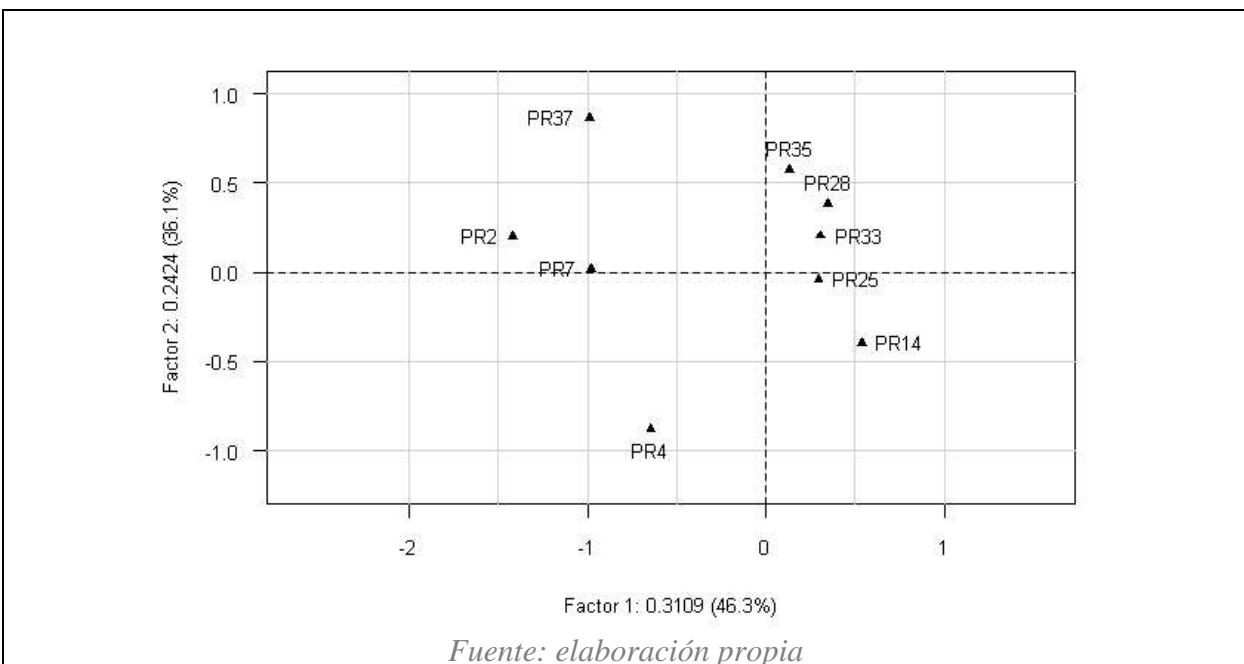
Figura 35. Primer plano factorial perfiles fila



Fuente: elaboración propia

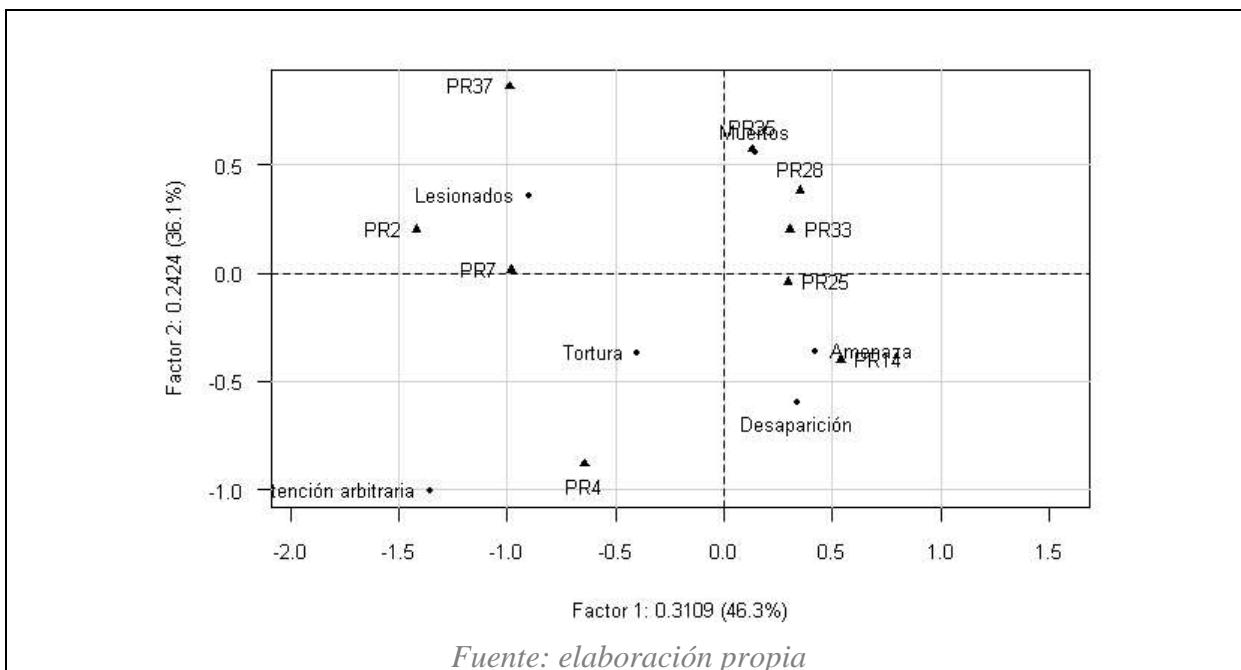
Búsqueda de los nuevos ejes en el espacio de las columnas: esta búsqueda es simétrica a la de perfiles fila y estaría asociada al siguiente $\text{ACP}(\mathbf{D}_p^{-1}\mathbf{F}', \mathbf{D}_n^{-1}, \mathbf{D}_p)$ en el cual ahora se trabaja con vectores en \mathbb{R}^6 . Como se puede ver solo cambiaron los subíndices, pero el proceso es el mismo dado que es un ACP procedente de la tabla \mathbf{F} , justo como el anterior. La salida de la función “dudi.coa” para los perfiles columna, se puede ver en la fig. 36: nuevamente un plano factorial derivado de los dos mayores valores propios y sus respectivos vectores propios, en los cuales se proyectan los perfiles columna.

Figura 36. Primer plano factorial profiles column.



De acuerdo con Pardo (2020, p. 119), dado que los dos planos factoriales para los perfiles, surgen de la misma matriz de frecuencias relativas, es posible hacer una representación simultánea (ver fig. 37). Por otro lado, en los mapas factoriales, generados a partir de la función “dudi.coa”, también se incluye el porcentaje de inercia que se acumula en cada uno de los ejes en los que se proyectaron los datos. De la fig. 37, podemos ver que en el primero se acumula un 46,3% y en el segundo un 36,1% de la inercia restante. En total se logró proyectar 82,4% de la inercia.

Figura 37. Representación simultánea de los planos factoriales



Por último, se muestran las coordenadas sobre los nuevos ejes en la Tabla 17, derivadas de $\mathbf{F}_s = \mathbf{D}_n^{-1} \mathbf{F} \mathbf{D}_p^{-1} \mathbf{u}_s$ y $\mathbf{F}_s = \mathbf{D}_p^{-1} \mathbf{F}' \mathbf{D}_n^{-1} \mathbf{u}_s$, para los \mathbf{u}_s que se obtienen de los valores propios: 0.3109 y 0.2424.

Tabla 17. Coordenadas de las categorías en la fig. 37

Categoría	Eje 1	Eje 2
PR2	-1,4207	0,2001
PR4	-0,6449	-0,8812
PR7	-0,9803	0,0132
PR14	0,5358	-0,3993
PR25	0,2932	-0,0432
PR28	0,3487	0,3822
PR33	0,3050	0,2022
PR35	0,1327	0,5701
PR37	-0,9891	0,8610
Muertos	0,1459	0,5544
Desaparición	0,3386	-0,5945
Tortura	-0,4005	-0,3697
Lesionados	-0,9002	0,3586
Detención arbitraria	-1,3565	-1,0038
Amenaza	0,4202	-0,3594

Fuente: elaboración propia

Interpretación de los planos factoriales simultáneos: en la fig. 37, se tiene el resultado de proyectar los dos perfiles en un subespacio de menor dimensión. De no poder visualizar las relaciones porque se tenían vectores en \mathbb{R}^6 y \mathbb{R}^9 , se obtiene—gracias al AC— una representación en \mathbb{R}^2 , la cual permite analizar más cómodamente las relaciones entre las categorías, debido a que el parecido entre ellas se traduce en términos de distancias.

Se puede observar la gran cercanía de las coordenadas para Amenazas y Paramilitares (PR14), lo mismo que para Muertos y Fiscalía (PR35). En el primer caso, se confirman las conclusiones que existen actualmente en la revista: revisitando una cita de la sección 3.3.3 se informó que “en las últimas décadas, las curvas de responsabilidad de la fuerza pública en violaciones graves a los derechos humanos y en crímenes de guerra solo descienden en la medida en que ascienden las de responsabilidad de los paramilitares” (Cinep/ppp, 2021, p. 9). Por otro lado, respecto a la cercanía entre las Muertes y la presunta responsabilidad de la Fiscalía, la fig. 37 termina confirmando nuevamente algo que se pudo intuir durante las lecturas de los relatos de la revista: existen motivos para creer que el abandono del estado en ciertas partes del país está relacionado con las muertes y la impunidad, a pesar de que se intente denunciar.

Por otro lado, puede observarse que, para la categoría Muertos y Amenazas, hay cercanía con, ERP (PR28), FARC-EP (PR25) y otras bandas criminales (PR33). Esto también podría dar indicios que ayudan a confirmar la atomización del conflicto luego de la firma del acuerdo de paz del 2016. En ese momento la mayor parte de las FARC abandonó los territorios que controlaron durante décadas, no hubo apoyo del gobierno para estas zonas y grupos criminales, como los incluidos en PR28, PR25 y PR33, tomaron el control y son quienes violentan los DDHH de las personas actualmente.

Con menos cercanía, pero también destacable, se puede hablar del parecido entre Desaparición con Paramilitares (PR14) y FARC-EP (PR25). Se podría intuir que la proximidad entre estos también obedece a las razones del párrafo anterior, debido a que son categorías de victimización relacionadas con el conflicto armado.

Otro caso de menor proximidad, pero relevante, es el de los Lesionados con la Fuerza Pública (PR2) y Policía (PR7). Sorprende el parecido dado que son datos del segundo periodo del 2020, en el cual estábamos en un punto álgido de la pandemia. No hay que olvidar el estallido social de 2019 que terminó teniendo mayor magnitud en 2021. Es posible que la cercanía entre estas categorías sea más evidente en análisis de datos del año pasado.

Conclusiones y reflexiones

Del presente trabajo de grado, se consideran las siguientes conclusiones y reflexiones:

- Primero, no hay que olvidar que la estadística es una herramienta para intentar explicar la realidad o los comportamientos humanos, incluyendo la posibilidad del error. Sin embargo, durante la construcción del presente documento, el propósito de caracterizar las violaciones a los DDHH, más allá de las simples descripciones anecdóticas de la revista Noche y Niebla, se alcanzó en gran medida a través de la herramienta estadística. Aunque no se consiguió una comprensión profunda acerca de los conceptos involucrados en los DDHH, debido a su complejidad, esto no impidió interpretar los resultados con los análisis utilizados. En última instancia el AC ayudó a caracterizar algunas relaciones relevantes entre los actos de violencia y presuntos responsables considerados, en gran parte gracias a la amplia atención que se le prestó a entender qué hacía este análisis y cómo darle significado a las salidas que generaba.

- Una motivación inicial para estudiar los temas de este documento obedecía a la búsqueda de otras maneras de interpretar el fenómeno de la violencia en Colombia, desde nuestra posición como profesores de matemáticas. La interpretación de este contexto tiene que ver con un problema social poco trabajado en nuestra formación en el programa, pero presente en un gran porcentaje del país; por ende, permeando las aulas cada día. Las herramientas estadísticas utilizadas terminaron dando intuiciones acerca de las violaciones a los DDHH. Esto último se evidenció en que al comparar los resultados a los que llegamos, con parte de la información consultada, las dinámicas de violencia en el país continúan en la misma línea de hace décadas: el conflicto armado y la falta de presencia del estado, afectando a gran parte de la población. El conocimiento que nos llevamos de construir este documento puede ayudar al papel que tendremos como docentes. A pesar de que sea una realidad difícil de contar, incluso desde la estadística, no se puede dejar de lado cuando se trata de educar a las siguientes generaciones.
- El álgebra lineal terminó teniendo un gran papel en comprender los análisis estadísticos multivariados que se tuvieron en cuenta. En los primeros contactos con los textos acerca del tema, creímos que el álgebra lineal involucrada estaba suficientemente clara como para entender la teoría. Fue después cuando se reveló que haber visto el curso de álgebra lineal con tanta formalidad, no colaboró a entender las ideas fundamentales de esta área; las cuales son protagonistas en los procesos del ACP y AC. Se intentó generar que el capítulo dedicado al álgebra lineal representara el descubrimiento de la simpleza de las ideas involucradas en esta materia. En nuestro caso particular, la comprensión de estas ideas permitió profundizar de mejor manera en los análisis multivariados; además, repercutiendo en el lenguaje utilizado en esta monografía. Durante nuestra formación no

tuvimos la oportunidad de conectar adecuada e intuitivamente el álgebra lineal con una interpretación geométrica, a pesar de la amplia formación en geometría que ofrece el programa. Fue gracias a lograr esa conexión, que este capítulo terminó siendo esencial para entender un poco mejor el ACP y el AC, debido a su interpretación geométrico-estadística.

- En la estructura del documento, la partes que conciernen a las temáticas que sustentan el AC (*“Elementos del análisis estadístico multivariado”*) también tuvieron gran protagonismo. En retrospectiva, el motivo de esto obedece a que se tuvo la intención de caracterizar de la mejor manera posible el AC, para que la aplicación de este sobre los datos fuera adecuada. En el curso hacía ese objetivo, fue necesario incluir explicaciones generales de estos tipos de análisis y del ACP, para que tuvieran sentido los últimos pasos del AC. Cabe señalar que quedaron cosas por fuera en esta parte comentada, pero consideramos que las explicaciones, especialmente las basadas en ejemplos, incluidas en la monografía fueron las suficientes para los propósitos iniciales que nos propusimos.
- El uso de una monografía como medio para intentar enseñar un tema en específico, representa interesantes aprendizajes para futuros licenciados en matemáticas. Como docentes de esta área, confiamos en el pensamiento matemático y las matemáticas, debido a su conocida formalidad, precisión, aplicabilidad a lo largo de muchos campos, etc. En nuestra formación, progresivamente nos volvimos aptos para entender la profundidad de las ideas relacionadas con las matemáticas, junto con la acumulación de habilidades y conocimiento para enseñarlas. A través de la experiencia de escribir esta monografía nos adentramos a explorar otras maneras de enseñar, buscando en este caso aportar a una comunidad académica que podría beneficiarse de un trabajo en esta línea.

Podemos concluir que el proceso influenció ampliamente la convicción de tener la capacidad de crear material, el cual podría funcionar como herramienta introductoria a los análisis multivariados.

- El presente documento también incluyó nuestro primer acercamiento a la investigación cuantitativa. No se pudieron tener en cuenta todas las fases debido a la naturaleza de los datos, pero resultó valioso adentrarnos en comprender las características de una investigación de este tipo. Considerar esta estructura metodológica, permitió organizar de mejor manera las consultas, estructura del documento, lenguaje utilizado, etc. Incluso permitió comprender mejor las limitaciones de utilizar la estadística en contextos reales y asimilar el funcionamiento de R como herramienta en la investigación cuantitativa.

Referencias

- Carpizo, J. (2011). Los derechos humanos: naturaleza, denominación y características. Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional, 1(25), 3-29. <https://doi.org/10.22201/ijj.24484881e.2011.25.5965>
- Centro de Investigación y Educación Popular Programa por la Paz, (2021). Anonimato. Factor dinamizador del genocidio impune. Noche y Niebla, 62. <https://www.nocheyniebla.org/wp-content/uploads/2021/04/Noche-y-Niebla-62.pdf>
- Centro de Investigación y Educación Popular Programa por la Paz, (2016). Marco Conceptual de la Red Nacional de Bancos de Datos. <https://www.nocheyniebla.org/wp-content/uploads/u1/comun/marcoteorico.pdf>
- Contreras, J. & Molina-Portillo, E. (2019) *Elementos clave de la cultura estadística en el análisis de la información basada en datos*. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), Actas del Tercer Congreso Internacional Virtual de Educación Estadística. Disponible en www.ugr.es/local/fqm126/civeest.html
- Cross Validated (2015) *Making sense of principal component analysis, eigenvectors & eigenvalues* <https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues?noredirect=1&lq=1>
- Díaz, L. (2007) *Estadística multivariada: Inferencia y métodos*. Universidad Nacional de Colombia

- Husson, F (2016) *PCA (course 1/3): description of the method in a French way*. [Video]. YouTube.
https://www.youtube.com/watch?v=IuSbb3nq4aI&ab_channel=Fran%C3%A7oisHusson
- Husson, F. (2016) *Correspondence Analysis (part 1/5): Introduction* [Video]. YouTube.
https://www.youtube.com/watch?v=Z5Lo1hvZ9fA&list=PLnZgp6epRBbTgO-l-6hLAKTstW7WSGh8A&index=5&ab_channel=Fran%C3%A7oisHusson
- Pardo, C. (2020) *Estadística descriptiva multivariada*. Universidad Nacional de Colombia.
- Sanderson, G. (2016). *Essence of linear algebra*. YouTube.
https://www.youtube.com/playlist?list=PLZHQB0bOWTQDPD3MizzM2xVFitgF8hE_ab
- Sampieri, R. & Fernández, C. & Baptista, P. (2014) *Metodología de la Investigación*. McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.
- Vargas, E. &, Pertuz, C. & Prada, M. (2020) Los otros en mí. *Estudios sobre educación para la paz. Cátedra de ciudadanía y DDHH. Tomo I* Bogotá: Universidad Pedagógica Nacional, Universidad del Rosario, Conferencia Nacional de Organizaciones Afrocolombianas, Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura.