

Análisis Conjunto de Configuraciones de Caso: una introducción al Pensamiento Configural

Conjunctive Analysis of Case Configurations: an introduction to Configural Thinking

Timothy C. Hart 

Griffith University

Asier Moneva¹ 

Universidad Miguel Hernández de Elche

RESUMEN

El presente artículo constituye la primera guía paso a paso en castellano para llevar a cabo un Análisis Conjunto de Configuraciones de Caso (Miethe, Hart, & Regoeczi, 2008). Este se define como una técnica exploratoria para el análisis multivariado de datos categóricos que supone una alternativa a otros métodos tradicionales. Mediante su desarrollo lógico se pretende mostrar un método sistemático de preparación de datos que permite realizar con ellos una amplia gama de análisis, mostrando tanto las ventajas que brinda como las limitaciones que presenta. Para ilustrar claramente su proceso de aplicación se ha recurrido a dos bases de datos gestionadas por el Instituto de Investigación Social de la Universidad de Michigan: *The National Incident Based Reporting System* y *The Law Enforcement Management and Administrative Statistics*. La interpretación de los resultados de esta técnica permite alcanzar conclusiones que pueden tener importantes implicaciones en los distintos procesos de toma de decisiones basados en evidencias.

Palabras clave: CACC, análisis exploratorio de datos, análisis multivariado, análisis conjunto, perfil situacional.

¹ La correspondencia debe dirigirse a: Asier Moneva. Centro CRÍMINA para el estudio y prevención de la delincuencia, Universidad Miguel Hernández de Elche. Avda. de la Universidad, s/n. Edif. Hélike. 03201 Elche (Alicante). amoneva@crimina.es

ABSTRACT

This paper establishes the first Spanish step-by-step guide to carry out a Conjunctive Analysis of Case Configurations (Miethe, Hart, & Regoeczi, 2008). It is defined as an exploratory technique for multivariate analysis of categorical data, which is an alternative to other traditional methods. Through its logical development, we intend to show a systematic method for data preparation that allows a wide range of analysis to be conducted on them, showing both the advantages it offers and its inherent limitations. To clearly illustrate its application process, two databases managed by the University of Michigan Institute for Social Research have been used: The National Incident Based Report System and The Law Enforcement Management and Administrative Statistics. The interpretation of results obtained by this technique allows reaching conclusions that may have important implications for several evidence based decision-making processes.

Keywords: CACC, exploratory data analysis, multivariate analysis, conjunctive analysis, situational profile.

1. Introducción

Cuando se pretende dar respuesta a algunas de las cuestiones que, por su ámbito y alcance, estudia la Criminología, en ocasiones se recurre al planteamiento de investigaciones que aplican el método científico con el objetivo principal de dotarlas de validez. A la hora de trabajar con diseños cuantitativos es frecuente ejecutar determinados análisis preliminares o exploratorios para realizar una aproximación general a los datos que permita obtener una visión amplia del fenómeno objeto de estudio. Puesto que el volumen de información que se maneja suele ser grande, este proceso sirve para simplificar la futura toma de decisiones. Una vez que el análisis de datos ha superado las primeras fases, la tendencia general consiste en calcular los coeficientes de correlación o regresión entre dos o más variables (p. e. Bautista & Sitges, 2016; Miró-Llinares, 2013), o realizar comparaciones de medias entre grupos (p. e. Rodríguez, Fernández, & Bautista, 2017).

Uno de los objetivos del investigador cuando realiza estos cálculos inferenciales puede ser extraer conclusiones generalizadas a partir de un caso particular mediante el contraste de una hipótesis que, en última instancia, será evaluada según su significación estadística. También es frecuente que los criminólogos traten de averiguar cuál es el factor causal del crimen y, para ello, se embarquen en complejas tareas de cálculo estadístico tratando de

construir modelos explicativos de la delincuencia con los que medir las variaciones que experimentan las variables incluidas en el mismo. Pero los métodos cuantitativos no están exentos de críticas (p. e. Young, 2011 y su ya célebre datasaurio) y como cualquier otra metodología también presentan problemas.

Por ejemplo, y atendiendo a posturas críticas razonadas, es posible que surjan problemas a la hora de escoger el estadístico más adecuado para comparar dos coeficientes de regresión a través de muestras independientes, puesto que algunos estimadores sesgan el contraste de hipótesis (Paternoster, Brame, Mazerolle, & Piquero, 1998). De hecho, los estadísticos no consideran el método de regresión como una técnica demasiado robusta porque al trabajar con tamaños muestrales grandes tienden a encontrar significancia entre variables cuya relación es extremadamente débil, una relación que no arrojaría resultados significativos mediante la aplicación de otras técnicas (Champion & Hartley, 2010). Entre otras funciones, las técnicas de regresión son útiles para localizar las variables con mayor peso específico dentro de un modelo, es decir, permiten identificar cuál es la variable más importante para la aparición del fenómeno que se está midiendo con la variable dependiente; pero cuando dos predictores se comportan de forma similar es importante examinar su relación en profundidad para diferenciar claramente ambos fenómenos, puesto que puede existir interacción entre ellos. En estos casos se recurre al estudio del efecto conjunto de dos variables, sin embargo, cuando lo mismo ocurre con tres o más predictores es difícil interpretar de qué modo interactúan y el efecto que esto tiene en el modelo. Además, por su propia naturaleza, estas técnicas tienden a suavizar los valores atípicos cuando se comparan individualmente con la media en la ecuación de regresión. Al hacerlo, la importancia del caso típico se vuelve menos representativa debido a la influencia de los valores atípicos y estos valores se vuelven menos extremos debido a la influencia de las puntuaciones promedio. En otras palabras, la regresión hace que los datos analizados “encajen” en el modelo construido. En definitiva, aquel modelo que incluya las variables adecuadas, y no necesariamente un gran número de ellas, tendrá un poder explicativo mayor, reflejando de forma más precisa la realidad del fenómeno que estudia.

Parece entonces conveniente explorar métodos alternativos que hagan replantear la forma de entender el proceso de análisis de datos cuando se afronta una nueva investigación

empírica. El enfoque que aquí se presenta es el Análisis Conjunto de Configuraciones de Caso (CACC, por sus siglas en inglés), desarrollado por Miethe, Hart y Regoeczi en 2008. La intención no es presentarlo como una técnica superior de análisis de datos respecto a otras, como la regresión de Mínimos Cuadrados Ordinarios (OLS, por sus siglas en inglés), los Modelos Jerárquicos Lineales (HLM, por sus siglas en inglés) o la regresión logística, sino contribuir con una visión general del método paso a paso, que permita entender su proceso de aplicación y valorar su utilidad. Y es que el CACC es un método que ya se ha utilizado en distintas investigaciones relacionadas con la Criminología o la Justicia Penal, como las dirigidas por Hart y Miethe sobre victimización (2008), microentornos criminógenos (Hart & Miethe, 2015), miedo al crimen (Hart, 2016), o las aplicadas por Summers y Caballero (2017) en robos callejeros desde una perspectiva espacial. Con estos antecedentes el objetivo final de este trabajo es persuadir a la comunidad española de criminólogos para que apliquen este método configural de análisis en sus propios datos y reflexionen sobre las posibilidades que ofrece.

1.1. Fundamento: el surgimiento de un enfoque alternativo

Muchos investigadores han sido formalmente entrenados en lo que Ragin (2013) describe como “la plantilla convencional para la investigación social” (p. 172). A continuación, se presenta un esquema que, a modo de resumen, muestra los conceptos básicos de la plantilla convencional diseñada por Ragin para, al mismo tiempo, compararlo con lo que él denomina plantilla alternativa (Tabla 1).

Tabla 1.

Comparativa entre las plantillas convencional y alternativa para la investigación social (Ragin, 2013, p.172)

Convencional	Alternativa
Variables	Conjuntos
Operativización	Calibración
Variables dependientes	Resultados cualitativos
Poblaciones determinadas	Poblaciones construidas
Correlaciones	Relaciones teóricas de conjuntos

Matriz de correlación	Tabla de la verdad
Tamaño del efecto	Fórmulas causales

En primer lugar, la plantilla convencional, orientada a la variable, adopta un enfoque sencillo para comprender la causación, que, generalmente, comienza con la especificación de una variable dependiente y algunas variables independientes cuidadosamente seleccionadas para explicar la variación de la primera, cuya elección dependerá de su adecuación para dar respuesta o solución a un determinado problema. Después, el investigador se dispone a probar sus teorías recogiendo los datos que estima relevantes sobre una población determinada. Es entonces cuando diseña uno o más modelos estadísticos con el objetivo de conocer el poder explicativo relativo de cada una de las variables independientes para identificar el factor que ha provocado la variación de la variable dependiente o el tamaño del efecto del modelo, siempre y cuando esté fundamentado sobre la base de un planteamiento teórico sólido.

Este modo convencional de plantear una investigación social se puede contrastar con uno configural o alternativo. En lugar de poner el foco en las variables, este método se orienta a un caso concreto para analizar los conjuntos de perfiles definidos por la coincidencia de todos los atributos de las variables predictoras. Al contrario que el proceso convencional de operativización basado en el uso de indicadores que, como mínimo, deben reflejar el constructo teórico subyacente que se cree que representan, los registros o casos estructurados mediante el sistema que propone el CACC se evalúan entre sí, a menudo con estadísticas descriptivas como la media o la desviación estándar específicas de la muestra. Del mismo modo, mientras que la variable dependiente es el punto focal de la plantilla convencional, en la plantilla configural o alternativa se pone el énfasis en un resultado cualitativo, centrando el interés del investigador en la presencia o ausencia de un fenómeno particular que, a su vez, queda definido por cada uno de los conjuntos de atributos de las variables, o perfiles, observados en los datos. A diferencia de los métodos convencionales, los métodos configurales como el CACC centran su análisis en las combinaciones de variables que componen estos perfiles situacionales mediante la construcción de tablas de la verdad, identificando así los perfiles únicos que existen en un conjunto de datos determinado, en

lugar de poner el foco en la existencia de correlaciones entre variables. Por tanto, en lugar de estimar probabilísticamente el tamaño del efecto en un modelo estadístico, el CACC identifica las fórmulas causales que dan lugar a cada resultado.

Cabe destacar que este no es el único método que aporta una visión configural; existen otros enfoques que poseen ciertas similitudes con el CACC, como el Análisis Cualitativo Comparativo (QCA, por sus siglas en inglés) desarrollado por el propio Ragin (1987), que ha ganado una visibilidad notable en la investigación aplicada al campo de la Ciencia Política y que aplica una metodología similar o, en menor medida, el Conjoint Analysis ideado por Green (Green & Srinivasan, 1978) e implementado principalmente en la investigación de mercados para conocer las preferencias de los consumidores. Como el CACC, ambos métodos centran el interés de su análisis en el de estudio de las combinaciones de un número limitado de atributos que se asocian a un determinado resultado. No obstante, desde que se introdujera en el campo de la criminología como una técnica para el Análisis Exploratorio de Datos hace ya una década, el CACC ha trascendido su utilidad original. Desde entonces, los investigadores han ampliado el alcance de su aplicación, ayudando a comprender y explicar mejor la variedad contextual que rodea a un determinado resultado relacionado con algunos de los temas que suscitan mayor interés en la disciplina, como el contexto situacional que define los nodos de actividad que transitan los agresores (Hart & Miethe, 2014), la puesta en conocimiento de incidentes violentos a la policía (Rennison, 2010), la disparidad en las sentencias de encarcelamiento (Lockwood, Hart, & Stewart, 2015), los patrones microespaciales del crimen (Hart & Miethe, 2015) o incluso la piratería (Bryant, Townsley, & Leclerc, 2014).

2. Método

Tras presentar una visión general del pensamiento configural, y contextualizar el origen del CACC sobre la plantilla de Ragin (2013), se proporcionan a continuación las instrucciones para llevar a cabo este método de agregación de datos en una matriz que los prepara para su

posterior análisis². Para ello, es conveniente la utilización de un *software* de análisis estadístico (p. e.: SPSS, STATA, SAS, R)³ e imprescindible un conjunto de datos que cumpla los siguientes requisitos:

- Los datos deben ser categóricos (nominales, ordinales o de intervalo).
- Las filas o configuraciones de caso deben representar la unidad de análisis.
- En las columnas se deben situar tanto la variable dependiente como las variables independientes que se espera estén asociadas al resultado de la primera.
- El resultado de la variable dependiente se debe medir con puntuación dicotómica.

Para realizar un CACC es necesario seguir tres pasos. El primer paso consiste en la construcción de una tabla de la verdad (ver Ragin, 2013) o matriz de datos con las observaciones contenidas en un conjunto de datos existente donde las filas representen la nueva unidad de análisis. Ahora, esta matriz de datos contiene todas las combinaciones posibles de los valores que toman las variables en el conjunto de datos o, es decir, consiste en una población construida de resultados que muestran la “verdad” en lugar de una estimación basada en una muestra obtenida de una población determinada. En segundo lugar, es necesario completar la matriz con la variable dependiente además de las variables independientes y de control seleccionadas. La variable dependiente se debe dicotomizar (0, nunca, y 1, siempre) para poder obtener la probabilidad media del resultado. Como se puede anticipar, se trata de una tarea relativamente sencilla en la que se agregan configuraciones de caso homogéneas, definidas por combinaciones de atributos idénticas, y se determina el número de registros asociados a cada perfil, además de la proporción de perfiles que arrojan un resultado positivo. El tercer y último paso consiste en aplicar un criterio de decisión para definir cuándo un perfil situacional es dominante, una circunstancia que a menudo se determina según el tamaño muestral, puesto que con este proceso se pierde diversidad de perfiles. Por este motivo, se han establecido diferentes criterios para identificar estos perfiles dominantes atendiendo al tamaño del set de datos: para muestras pequeñas ($N < 1.000$) son

² Para consultar otras publicaciones en las que se desarrolla esta metodología, ver Hart (2014); Hart, Rennison & Miethe (2017); y Miethe et al, (2008).

³ La sintaxis necesaria para generar una matriz de datos conjuntos con SPSS, STATA y SAS se encuentra originalmente en Miethe et al. (2008, p.240).

necesarias, al menos, 5 configuraciones idénticas, mientras que conjuntos de datos mayores requieren 10 o más configuraciones iguales (Hart et al., 2017). La tabla de la verdad resultante tendrá un registro para cada perfil situacional dominante empíricamente observado en los datos originales. Una vez completados los tres pasos, el investigador estará preparado para analizar sus datos del modo que estime oportuno en busca de patrones que le permitan responder a las distintas preguntas de investigación que puedan suscitar su interés mediante el uso de métodos descriptivos o pruebas estadísticas más formales, como una prueba de bondad de ajuste chi-cuadrado o una ANOVA de dos factores (ver Hart & Miethe, 2015, p.116).

3. Aplicación y resultados

Con el objetivo de facilitar la comprensión del método presentado, a continuación se ilustran algunos de los conceptos clave relacionados con la plantilla configural y se presenta una explicación paso a paso para aplicar un CACC de forma sistemática.

Si, por ejemplo, se quiere averiguar hasta qué punto las estrategias de patrullaje y vigilancia comunitaria afectan a la cantidad de arrestos que se llevan a cabo, se puede plantear la siguiente pregunta de investigación para practicar un enfoque configural en un contexto concreto:

PI ¿Cómo afecta la presencia policial y los servicios policiales orientados a la comunidad al número de arrestos efectuados?

El investigador puede aprovechar distintas fuentes de obtención de datos para dar respuesta a esta cuestión. En este caso, se va a utilizar la información almacenada mediante dos programas estadounidenses de recogida de datos transversales: *The National Incident Based Reporting System* (NIBRS) y *The Law Enforcement Management and Administrative Statistics* (LEMAS), cuyo diseño permite su vinculación. Estas bases de datos resultan ideales para ilustrar la aplicación del CACC puesto que (1) almacenan grandes volúmenes de datos y (2) han sido utilizadas con anterioridad en diversas investigaciones que aplican métodos analíticos tradicionales.

De los datos que ofrece LEMAS, se pueden utilizar algunas de las variables que categorizan ciertas características de las organizaciones policiales. Siguiendo la estructura de la plantilla convencional, las variables independientes extraídas pueden ser, en primer lugar, un indicador de la presencia policial en un área determinada (Patrol), elaborado mediante el cálculo de una proporción simple -el número de agentes a tiempo completo asignados a patrullaje dividido entre su total⁴- y, en segundo lugar, se pueden contabilizar las actividades y servicios policiales orientados a la comunidad (COPS) para crear una escala aditiva. Ambas variables se han codificado posteriormente como categorías ordinales. Por otro lado, del NIBRS se puede extraer información dicotomizada sobre el resultado asociado al crimen, en este caso agresión agravada, medido como arresto, o no arresto, que actúa como variable dependiente. Para terminar de estructurar los datos, se han extraído algunas variables de control de la base de datos del NIBRS. Teniendo en cuenta una vez más la plantilla convencional, se pueden utilizar análisis covariados para determinar si realmente existe una relación significativa entre las variables independientes y la variable dependiente. Así, se han extraído algunas variables que covarían con el resultado y reflejan ciertas características personales, como el sexo y la edad del agresor, el tipo de arma utilizada, o si estaba influenciado, o no, por alcohol o drogas en el momento de la ofensa; y situacionales, como la hora en la que ocurrió el evento y la ubicación del mismo (si ocurrió en un domicilio o no). Como en el resto de los casos, las variables de control también se han categorizado cualitativamente (Tabla 2).

Tabla 2.

Ejemplo de matriz de datos

Registro	Arresto	Edad	Sexo	Influencia	Arma	Hora	Domicilio	Patrol	COPS
1	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
2	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
3	No	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
4	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
5	No	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
6	No	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto

⁴ Esta proporción es una medida de uso común en Estados Unidos que permite operativizar distintas categorías relacionadas con las estrategias policiales y las fuerzas del orden.

7	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
8	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Alto
9	No	50+	Hombre	Sí	Otra	18-23	Sí	Baja	Alto
10	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
11	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
12	No	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
13	No	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
14	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
15	Sí	50+	Hombre	Sí	Otra	18-23	Sí	Media	Moderado
...

Puesto que en la matriz se han vinculado datos referentes tanto a las organizaciones policiales como al incidente con posible resultado de arresto, el investigador podría ejecutar una regresión logística o un HLM para estudiar la relación existente entre ambas variables. Y aunque esta técnica le permitiría identificar el peso específico de cada variable y el tamaño del efecto explicado por el modelo, estaría aplicando nuevamente técnicas tradicionales de análisis de datos. Como alternativa, el enfoque configural del CACC se centra en analizar los conjuntos de variables que componen perfiles situacionales. De este modo, se pueden identificar tanto los patrones relacionados con la frecuencia relativa con la que se repiten estos conjuntos, como la configuración de los atributos que pueden tomar las variables asociadas a una mayor o menor probabilidad de arresto.

Para comprender mejor el trasfondo del pensamiento configural, es recomendable tomarse un instante para reflexionar una vez más sobre la matriz de datos construida en la Tabla 2. Cada fila muestra un registro de datos combinados, extraídos de las bases de datos de NIBRS y LEMAS, sobre un total de más de 72.000 registros que tienen un identificador asociado y donde cada columna representa las variables independientes. Estas son las denominadas configuraciones de caso, la nueva unidad de análisis que será agregada. Si se puede resaltar alguna característica relativa a los 15 registros que se muestran en la tabla anterior, es que la mayoría comparten los mismos atributos en sus variables independientes. A pesar de ello, es importante resaltar que diferentes registros con un mismo perfil pueden arrojar resultados distintos (p. e.: los registros 2 y 3 son idénticos en relación con los predictores, pero sólo uno de ellos resulta en arresto), y que diferentes perfiles o configuraciones pueden tener el mismo resultado. Un análisis de datos que utiliza un enfoque

configural, como el CACC, agrega los perfiles homogéneos, basándose en la repetición de atributos de cada variable, creando conjuntos y usándolos como nueva unidad de análisis. De este modo, se puede evaluar la diversidad contextual del archivo de datos, su agrupamiento circunstancial, las configuraciones de caso típicas y atípicas, e, incluso, identificar las fuentes concretas de variabilidad contextual.

Aunque por motivos evidentes no se muestran todos, la Tabla 2 contiene más de 72.000 registros que van a ser analizados. Antes de agregar estos registros por configuraciones de caso homogéneas para el CACC, es posible calcular todas las combinaciones posibles de los atributos de las variables (Tabla 3). En este caso, el conjunto de datos analizado puede contener más de 8.600 perfiles situacionales únicos.

Tabla 3.

Listado de categorías de las variables independientes

(y ₁) Edad	(y ₂) Sexo	(y ₃) Influencia	(y ₄) Arma	(y ₅) Hora	(y ₆) Domicilio	(y ₇) Patrol	(y ₈) COPS ⁵
-15	Hombre	Sí	Personal	00-05	Sí	Baja	Bajo
15-24	Mujer	No	Cuchillo	06-11	No	Media	Moderado
25-29			Arma de fuego	12-17		Alta	Alto
30-39			Otra	18-23			
40-49			Ninguna				
50+							

$$\text{Perfiles posibles} = N_{\text{atributos}y_1} * N_{\text{atributos}y_2} * \dots * N_{\text{atributos}y_n}$$

$$\text{Perfiles posibles} = 6 * 2 * 2 * 5 * 4 * 2 * 3 * 3 = 8.640$$

Sin embargo, en la matriz de datos tan sólo se han observado 5.512 perfiles, que representan el 63,80% del total de perfiles posibles. Esto significa que los resultados de arresto, o no, en los casos de agresión agravada, están definidos por una amplia gama de contextos situacionales. Cuando se circunscribe el interés a los denominados perfiles situacionales dominantes, aquellos conjuntos homogéneos que se observan en la matriz de

⁵ Cuando la proporción calculada para esta variable era menor o igual que 0.57, se ha codificado como “Baja”, si el valor ha sido igual o superior a 0.80 como “Alta”, y como “Media” si la proporción se ha situado en el intervalo restante.

datos 10 o más veces, sigue existiendo una diversidad contextual considerable dentro de las configuraciones de caso observadas. No obstante, también se observan patrones de agrupamiento situacional. Así, el 8% de los perfiles representan el 25% de todos los resultados de arresto derivados de agresión agravada y el 21% de perfiles representa más de la mitad del total de observaciones. Cabe destacar que este patrón es, en realidad, mucho menor de lo que a menudo se observa. Habitualmente, se ven agrupaciones situacionales que se ajustan más a la regla del 80:20.

Para aquellos que aún muestren cierto escepticismo sobre la capacidad exploratoria de esta técnica, es importante subrayar que también se pueden identificar patrones típicos y atípicos después de construir la matriz de datos que contiene las configuraciones de casos. Para explorar las configuraciones de caso, se pueden definir como perfiles situacionales típicos aquellos incluidos dentro del margen establecido por ± 2 desviaciones estándar de la probabilidad media de arresto, mientras que los perfiles situacionales atípicos serían aquellos que superan ese margen de ± 2 desviaciones estándar. Si el investigador desea identificar de un modo más formal los “cortes” relevantes en los datos, puede aplicar una prueba de rangos o una de bondad de ajuste para evaluar formalmente la distribución de aquellos casos observados que son diferentes de lo esperado, tal y como recomiendan Hart, et al. (2017). En la siguiente tabla (Tabla 4) se muestran los perfiles situacionales que, con más probabilidad, acaban en arresto.

Tabla 4.

Perfiles situacionales de asalto agravado donde el agresor tiene una probabilidad significativamente alta (+2 desviaciones estándar) de ser arrestado.

Perfil	Edad	Sexo	Influencia	Arma	Hora	Domicilio	Patrol	COPS	P(A)	n
1452	50+	Mujer	Sí	Otra	18-23	Sí	Media	Alto	1	12
1207	25-29	Mujer	Sí	Cuchillo	18-23	Sí	Media	Alto	0.93	15
1323	15-24	Mujer	Sí	Personal	12-17	Sí	Media	Alto	0.92	13
1429	40-49	Hombre	No	Personal	18-23	No	Baja	Moderado	0.92	12
611	50+	Mujer	Sí	Otra	18-23	Sí	Media	Moderado	0.91	35
914	40-49	Mujer	Sí	Personal	18-23	Sí	Baja	Moderado	0.91	22
1506	25-29	Hombre	Sí	Personal	18-23	Sí	Media	Moderado	0.91	11
1508	30-39	Mujer	Sí	Personal	12-17	Sí	Alta	Moderado	0.91	11

1570	15-24	Hombre	No	Otra	12-17	Sí	Alta	Alto	0.9	10
1653	40-49	Hombre	No	Otra	18-23	Sí	Alta	Alto	0.9	10
...
Probabilidad media de arresto = 0.49 (SD = 0.18)										

Tal y como se puede observar, incluso entre los perfiles que derivan en arresto con mayor probabilidad, existe una gran diversidad de fórmulas causales. Teniendo en cuenta únicamente las variables independientes “Patrol” y “COPS”, ya se dibujan perfiles situacionales donde la probabilidad de arresto asociada es especialmente alta, incluso cuando el valor que toman estas variables es opuesto al esperado en un contexto ideal. Por ejemplo, en el perfil 1570, cuando la presencia policial (Patrol) es alta y los servicios policiales orientados a la comunidad (COPS) son también altos, la probabilidad de arresto asociada a una agresión agravada sigue siendo considerablemente alta ($P(A) = 0.90$). Estas combinaciones son dignas de mención, ya que se pueden utilizar para mejorar las estrategias policiales centradas en arrestos. No obstante, de haber utilizado métodos analíticos tradicionales, estos perfiles situacionales podrían haber pasado desapercibidos.

Otra de las posibilidades que ofrece el CACC para analizar perfiles generados a partir de la matriz de datos, consiste en mirar de cerca las fuentes de variabilidad contextual, visualizando la información mediante gráficos de caja. Siguiendo este método, los pares de configuraciones de caso que tengan perfiles idénticos, salvo por el valor del atributo que tome una de las variables, se pueden emparejar para identificar cuál es el elemento concreto que hace variar el resultado de arresto. De este modo, se puede identificar el elemento diferencial entre dos configuraciones de caso que produce el *efecto principal* en el resultado. Una vez que se calculan las probabilidades de resultado para todos los pares coincidentes, una opción es proceder a su ilustración mediante gráficos de caja. Este tipo de gráficos facilita la visualización e interpretación de la variabilidad contextual de cada variable independiente y proporciona evidencia visual de algunas de las limitaciones que tienen los enfoques convencionales para el análisis de datos basados en variables. En el gráfico 1, se ha generado un diagrama de caja para representar la variable “Patrol” a partir de 130 comparaciones de pares de perfiles. Los perfiles eran idénticos salvo por el valor asociado a esta variable. Para cada par, uno de los perfiles analizados tomaba el valor ‘alto’ en el campo “Patrol”, mientras

que el otro perfil tenía atribuido el valor ‘bajo’ en el mismo campo. Esta circunstancia permite realizar una sencilla resta de la probabilidad de arresto ($P(A)$) asociada al perfil con un valor ‘bajo’ en el campo “Patrol” sobre la probabilidad asociada al perfil con el valor ‘alto’, y la diferencia resultante es el *efecto principal* que tiene esta variable sobre la probabilidad de arresto. Este proceso se ha repetido 129 veces y se han representado los resultados obtenidos en un diagrama de caja (gráfico 1). El *efecto principal* de la variable “Patrol” que aquí se muestra es claro: mientras que en algunas situaciones en las que existe una presencia policial ‘alta’ la probabilidad de arresto aumenta en gran medida, en otras ocasiones la misma probabilidad disminuye notablemente. Es decir, parece que una alta presencia policial no determina de forma clara una mayor o menor probabilidad de arresto. Y algo similar ocurre cuando se aplica este proceso a la variable “COPS”; las diferentes estrategias policiales no presentan una relación clara con la probabilidad de arresto.

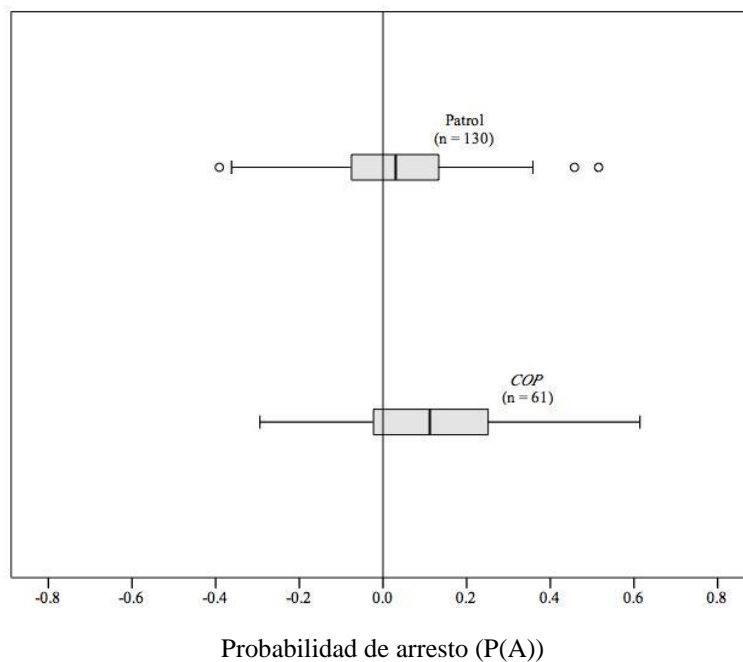


Figura 1: Fuentes de variabilidad

Los resultados ilustran la importancia del contexto en el que se estudian los diferentes predictores cuando se trata de comprender su comportamiento y, a su vez, resaltan la capacidad del CACC para detectar estos sutiles matices que las técnicas de regresión no son

capaces de revelar. El estudio de la variabilidad contextual de la presencia policial y los servicios policiales orientados a la comunidad permite desligar de estas variables el resultado de arresto para plantear que, probablemente, este se deba a la influencia de otro factor o incluso a un conjunto de predictores sociodemográficos y espaciotemporales que, de otro modo, hubieran pasado desapercibidos. Las implicaciones derivadas de los análisis que propone el CACC pueden tener un importante impacto en ámbitos como la teoría criminológica, el diseño de políticas públicas, la implantación estrategias de intervención, o la evaluación de la eficacia de los programas de prevención.

4. Discusión y conclusiones

La sencilla e intuitiva metodología que ofrece el CACC ha fomentado su cada vez más frecuente aplicación en los distintos ámbitos de la investigación criminológica actual, demostrando su utilidad práctica como un potente método exploratorio de datos. Aunque, probablemente, el potencial absoluto de esta técnica esté aún por descubrir, ya se puede constatar su innegable capacidad para establecer perfiles situacionales que facilitan la toma de decisiones en diferentes contextos y que, incluso, permite identificar los elementos clave dentro de cada perfil situacional, lo que supone una enorme ventaja a la hora de implementar medidas específicas con una eficiencia mayor y evita el problema de la asimetría causal. Respecto a esto último, tal y como explican Miethe et al. (2017) en un estudio sobre la disparidad en sentencias, “los métodos estadísticos utilizados en estudios previos presuponen relaciones simétricas entre las condiciones sociales y el encarcelamiento (p. e.: si las altas tasas de criminalidad están vinculadas a altas tasas de encarcelamiento, las bajas tasas de criminalidad también estarán vinculadas a bajas tasas de encarcelamiento). Sin embargo, al postular múltiples vías causales, la mayoría de las teorías del castigo tienden a implicar una asimetría causal. Por ejemplo, la presencia de altas tasas de criminalidad y encarcelamiento no necesariamente implican su opuesto (que menor delincuencia conlleve una menor tasa de encarcelamiento) debido a la influencia de otras causas conjuntas para un alto y bajo resultado de encarcelamiento” (p.156). En realidad, lo que esto implica es que se necesitan análisis separados para explicar la ocurrencia, o no, de un resultado concreto. Esto puede

hacerse efectivo con facilidad en el CACC porque todas las posibles combinaciones de variables que dan lugar a un resultado están representadas en la matriz de datos. En resumen, pensar en términos de “correlación” es un modo tradicional de razonar que debería reemplazarse por un enfoque de pensamiento “configural” si se pretende comprender plenamente y apreciar el CACC como un enfoque alternativo de análisis de datos.

Del lado puramente metodológico, una de las virtudes de esta técnica es que puede ser utilizada tanto de forma exploratoria como concluyente, de forma aislada o en combinación con otros enfoques más orientados a la variable, lo que denota su versatilidad. Además, se trata de un método inclusivo con los datos, en el sentido en que permite trabajar con todos los registros existentes, incluyendo aquellos que serían clasificados como valores atípicos u *outliers* con otros métodos, permitiendo al investigador entender completamente las sutilezas de la variabilidad contextual y evitando así los errores de estimación. Y también puede resultar de ayuda a la hora de detectar problemas de multicolinealidad entre variables puesto que presenta visualmente los datos de manera comprensiva y atractiva. En todo caso, aunque es cierto que la utilización del CACC presenta muchas ventajas, posiblemente su mayor logro sea que “este método representa el término medio entre (1) el enfoque en la especificidad y la causalidad múltiple que subyace en la mayoría de la investigación cualitativa y (2) la búsqueda orientada a la variable de patrones generales a través de diferentes contextos en la mayor parte de la investigación cuantitativa” (Miethe et al., 2008, p.239).

5. Referencias

- Bautista, R., & Sitges, E. (2016). Análisis de los predictores psicosociales de la transgresión de normas de tráfico en España: los casos del uso del cinturón, del teléfono móvil, los límites de velocidad y el límite de alcoholemia al volante. *Revista Española de Investigación Criminológica*, 14(5), 1-22.
- Bryant, W., Townsley, M., & Leclerc, B. (2014). Preventing maritime pirate attacks: A conjunctive analysis of the effectiveness of ship protection measures recommended by the international maritime organization. *Journal of Transportation Security*, 7(1), 69-82. DOI: 10.1007/s12198-013-0130-2
- Champion, D. J., & Hartley, R. D. (2010). *Statistics for Criminal Justice and Criminology (3rd Edition)*. Upper Saddle River, NJ: Pearson Education.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research*, 5(2), 103-123.
- Hart, T. C. (2014). *Conjunctive Analysis of Case Configurations*. Londres, Reino Unido: UCL Jill Dando Institute of Security and Crime Science.
- Hart, T. C. (2016). Using typologies of victimization worry to create strategies for reducing fear of crime. *Police Practice and Research: An International Journal*, 18(4), 407-419. DOI: 10.1080/15614263.2016.1230854
- Hart, T. C., & Miethe, T. D. (2008). Exploring bystander presence and intervention in nonfatal violence victimization: when does helping really help? *Violence and Victims*, 23, 637-651. DOI: 10.1891/0886-6708.23.5.637
- Hart, T. C., & Miethe, T. D. (2015). Configural behavior settings of crime event locations: Towards an alternative conceptualization of criminogenic micro-environments. *Journal of Research in Crime and Delinquency*, 52(3), 373-402. DOI: 10.1177/0022427814566639
- Hart, T. C., & Miethe, T. D. (2014). Street robbery and public bus stops: A case study of activity nodes and situational risk. *Security Journal*, 27(2), 180-193. DOI: 10.1057/sj.2014.5
- Hart, T. C., Rennison, C. M., & Miethe, T. D. (2017). Identifying patterns of situational clustering and contextual variability in criminological data: An overview of conjunctive analysis of case configurations (CACC). *Journal of Contemporary Criminal Justice*, 33(2), 112-120. DOI: 10.1177/1043986216689746
- Lockwood, K., Hart, T. C., & Stewart, A. (2015). First nations peoples and judicial sentencing: Main effects and the impact of contextual variability. *British Journal of Criminology*, 55(4), 769-789. DOI: 10.1093/bjc/azv005
- Miethe, T. D., Hart, T. C., & Regoeczi, W. C. (2008). The conjunctive analysis of case configurations: An exploratory method for discrete multivariate analyses of crime data. *Journal of Quantitative Criminology*, 24(2), 227-241. DOI: 10.1007/s10940-008-9044-8
- Miethe, T. D., Troshynski, E. I., & Hart, T. C. (2017). Social conditions and cross-national imprisonment rates: Using set-theoretic methods for theory testing and identifying deviant cases. *Journal of Contemporary Criminal Justice*, 33(2), 152-172. DOI: 10.1177/1043986216688815

- Miró-Llinares, F. (2013). La victimización por cibercriminalidad social. Un estudio a partir de la teoría de las actividades cotidianas en el ciberespacio. *Revista Española de Investigación Criminológica*, 11(5), 1-35.
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4), 859-866. DOI: 10.1111/j.1745-9125.1998.tb01268.x
- Ragin, C. C. (2013). New Directions in the Logic of Social Inquiry. *Political Research Quarterly*, 66(1), 171-174. DOI: 10.1177/1065912912468269b
- Ragin, C. C. (1987). *The comparative method*. Berkeley, CA: University of California Press.
- Rennison, C. M. (2010). An investigation of reporting violence to the police: A focus on Hispanic victims. *Journal of Criminal Justice*, 38(4), 390-399. DOI: 10.1016/j.jcrimjus.2010.04.007
- Rodríguez, S., Fernández, E. B., & Bautista, R. (2017). Prevención de la cibervictimización en menores de la provincia de Alicante. *Revista Española de Investigación Criminológica*, 15(2), 1-25.
- Summers, L., & Caballero, M. (2017). Spatial conjunctive analysis of (crime) case configurations: Using Monte Carlo methods for significance testing. *Applied Geography*, 84, 55-63. DOI: 10.1016/j.apgeog.2017.05.002
- Young, J. (2011). *The Criminological Imagination*. Cambridge, Reino Unido: Polity.

Agradecimientos

A la Prof. Dra. Lucía Summers de Texas State University por sus comentarios, que han contribuido en gran medida a mejorar el contenido del texto, y por su disponibilidad durante todo el proceso de elaboración del artículo.


Financiación

Este trabajo ha sido parcialmente financiado con fondos de las ayudas para la formación de profesorado universitario (FPU) del Ministerio de Educación, Cultura y Deporte del Gobierno de España, con referencia FPU16/01671.


Asimismo, el estudio también ha sido desarrollado en el marco del proyecto Criminología, evidencias empíricas y Política criminal. Sobre la incorporación de datos científicos para la toma de decisiones en relación con la criminalización de conductas, financiado por el Ministerio de Economía, Industria y Competitividad (MINECO), con referencia DER2017-86204-R

Timothy C. Hart es Doctor en Criminología y *Senior Lecturer* en la Facultad de Criminología y Justicia Penal de la Universidad de Griffith. En la actualidad, imparte docencia en investigación y análisis del delito, mapas del crimen y estadística aplicada a la investigación social. Anteriormente, ha trabajado como estadístico en la Oficina de Estadísticas de Justicia del Departamento de Justicia de Estados Unidos. Su experiencia

investigadora se centra en estadística aplicada y métodos cuantitativos, sistemas de información geográfica, encuestas y victimización. Es coautor de dos libros y ha publicado gran cantidad de artículos científicos en revistas de prestigio.

 <https://orcid.org/0000-0003-3274-3661>

Asier Moneva es graduado en Criminología por la Universidad de Salamanca y ha cursado el Máster en Análisis y Prevención del Crimen por la Universidad Miguel Hernández de Elche. En la actualidad, es Personal Investigador en Formación (FPU) en el Centro CRÍMINA para el estudio y prevención de la delincuencia de la Universidad Miguel Hernández de Elche. Sus principales áreas de interés son la criminología ambiental, el cibercrimen y los métodos de análisis de datos.

 <https://orcid.org/0000-0002-2156-0213>