

LOS ANÁLISIS CLUSTERS DE TIPO JERÁRQUICO Y LOS DENDROGRAMAS. UNA VISIÓN PARA LA TRIANGULACIÓN METODOLÓGICA EN LOS ESTUDIOS COMPARATIVOS REGIONALES EN AMÉRICA LATINA

Juan José Marín Hernández (Universidad de Costa Rica)

En las últimas décadas en las Ciencias Sociales se ha incrementado el interés por realizar estudios comparados que permitan combinar tanto los elementos cuantitativos como cualitativos de los espacios, procesos o actores que se analizan. No obstante, el intento llega siempre a las mismas interrogantes teórico-metodológicas, tales como: ¿qué comparar?, ¿para qué?, ¿cómo?, ¿qué elementos deben considerarse?, ¿qué tipo de fuentes se pueden usar?, ¿qué cantidad de fuentes se requieren para poder realizar un análisis comparativo relativamente exitoso?, ¿qué estrategias metodológicas se pueden seguir para considerar los factores cualitativos y cuantificables de los espacios comparados? (Sartori, 1999: 11) Estas interrogantes llevan implícito una serie de problemáticas, como por ejemplo qué estrategias se pueden usar para valorar tanto grandes cantidades de elementos a comparar como el número de categorías, la temporalidad que se debe tomar en cuenta y, desde luego, cómo se puede sopesar la incidencia de los elementos cuantitativos y cualitativos; y si estos se pueden o no triangular datos, fuentes, enfoques, perspectivas y metodologías.

El presente artículo pretende valorar el denominado análisis clusters jerárquicos y uno de sus correlatos más significativos: los dendrogramas como estrategias de investigación validas para las ciencias sociales con perspectiva comparada. Debe indicarse que ambas han sido muy frecuentes utilizadas en la investigación biológica, pero en las últimas décadas se han difundido en otra áreas. Así, por ejemplo, en la historia su vinculación ha sido por la vía arqueología, antropología y la historia antigua. Como se verá a lo largo de este artículo la lógica del procedimiento es relativamente sencilla, se parte de que “tantos grupos iniciales como individuos se estudian, se trata de conseguir agrupaciones sucesivas entre ellos de forma que progresivamente se vayan integrando en clusters los cuales, a su vez, se unirían entre sí en un nivel superior formando grupos mayores que más tarde se juntarán hasta llegar al cluster final que contiene todos los casos analizados” (Pérez, 2001: 468). Con ese procedimiento se puede visualizar fácilmente que la gran potencialidad de los análisis clusters jerárquicos y los dendrogramas se halla en su capacidad de crear tipologías y agrupamientos que faciliten la comparación a través del análisis de los mismos elementos (con sus características intrínsecas) que se estudian, facilitando el diálogo entre teoría y realidad social, generando a su vez nuevas problemáticas y con ello convirtiendo a nuestras categorías en herramientas de análisis, antes que conceptos duros que se imponen a la realidad estudiada.

El análisis cluster jerárquico puede utilizar perfectamente variables de tipo cuantitativo como cualitativo, como se verá más adelante, y esto abre una serie de posibilidades para crear estrategias

enmarcadas en la lógica del multi método y la combinación de lo cualitativo con lo cuantitativo¹. En el ámbito específicamente de lo regional facilitan el estudio de distintas dimensiones como son las de los actores sociales, sus vivencias y sus condiciones, así como los terrenos de lo espacial facilita la investigación de las interacciones entre el lugar, el espacio y la regionalización así con sus vinculaciones trans locales, y trans nacionales y trans regionales.

La vinculación entre lo cuantitativo y cualitativo en los análisis cluster jerárquicos se da gracias a que pueden usar tanto las variables nominales, ordinales; de intervalo y de razón; así como las denominadas categóricas, es decir aquellas que no aparecen en forma numérica, sino como categorías o atributos (sexo, profesión, color de ojos).

Es importante subrayar que el análisis cluster jerárquico acepta la posibilidad de usar tanto las variables cualitativas como cuantitativas, aspecto que no se puede lograr con el procedimiento de cluster *no jerárquico*, el cual no puede interpretar las variables de tipo binario y las de recuento².

El presente artículo es solo una invitación a revisar esta estrategia metodológica, por lo cual se abordarán, de forma general e introductoria, los diferentes dimensiones de esta metodología. Asimismo este trabajo se divide en tres partes: la primera procurara visualizar los usos que se le han dado a los dendrogramas en diferentes disciplinas y sus vías de aproximación con las ciencias sociales y la historia; la segunda parte procurará describir los pasos fundamentales de las técnicas de conglomerados y de dendrogramas, a través de un ejemplo sencillo; y finalmente, se incorpora un epílogo que se encaminará en presentar un análisis prospectivo para la historia y los estudios regionales.

1. LOS DENDROGRAMAS EN LAS CIENCIAS SOCIALES

Desde hace mucho tiempo, el uso de las estrategias de agrupamientos y dendrográficos han sido utilizados por las ciencias naturales y la estadística. Los puntos de encuentro con las ciencias humanas y ciencias sociales han sido básicamente a través de los análisis lingüísticos, demográficos, psicológicos y geográficos, aspecto que asemeja el descrito por Marc Bloch en los inicios del desarrollo del método comparado³.

¹ Convencionalmente se entiende por triangulación al uso de múltiples métodos en el estudio de un mismo objeto. No obstante es conveniente concebir la triangulación envolviendo variedades de datos, investigadores y teorías, así como metodologías (véase Samper, 2001).

² Los análisis cluster no jerárquicos son un procedimiento que intenta identificar los grupos de caso relativamente homogéneos basándose en las características seleccionadas y utilizando un algoritmo que puede gestionar un gran número de casos, para ello se debe seleccionar y especificar el número de conglomerados o grupos que se quieren estudiar. Según César Pérez una pregunta típica puede ser cuáles son los grupos identificables de países con población, densidad de población y población urbana similares, lo cual deja de lado las variables de recuento, entendidas estas como el número de veces que un fenómeno se repite en un periodo de tiempo determinado. En todo caso, esta vía es útil para comparar casos análogos o equivalentes (Pérez, 2001: 470).

³ Una versión en castellano del trabajo clásico de Marc Bloch se puede localizar en el siguiente sitio web http://jmarin.izonecr.com/documentos/Bloch_metodo_comparado.pdf Vale indicar que en Centroamérica y especialmente, en Costa Rica, se ha dado una aceptación casi dogmática de los principios esbozados por Bloch, en especial su máxima de “comparar solo lo comparable”, ello ha inhibido el desarrollo de otras estrategias metodológicas como las provenientes de la antropología, la lingüística y las denominadas ciencias duras.

Una rápida revisión de los trabajos que han usado las estrategias cluster y de dendrogramas parece confirmar esa visión. Pero, ¿cómo y para qué se usan? En esencia se aprovechan para establecer clasificaciones, taxonomías y distribuciones de procesos, espacios y actores sociales. La zoogeografía, la biología y la microbiología parecen ser las que mejor han aplicado los atributos de estas técnicas para la elaboración de taxonomías y clasificaciones, las cuales son la base para realizar procesos de análisis más complejos, como pueden ser la determinaciones de patrones de evolución; el establecimiento y descubrimiento de mecanismos de diferenciación genética, la localización de procesos de especialización; el develamiento de distintos grados de dispersión, entre otros, que a su vez les permite a los investigadores establecer los desarrollos de linajes; los grados de biodiversidad, y las magnitudes de procesos como la extinción o la dispersión de especies⁴.

La geografía física ha hecho un empleo parecido a la biología utilizando los análisis cluster y dendrogramas para coaligar los diferentes parámetros vegetales y variables climáticas con la distribución y la abundancia de la fauna en distintos tipos de hábitat. Así el proceso de clasificación determinaba los parámetros del hábitat, evidenciando como se tenía en estos espacios una mayor relación con la distribución y abundancia especies permitiendo comprender la estructura vegetal de los sitios estudiados. De esta forma, tanto el análisis cluster como los dendrogramas son muy apreciados para clasificar suelos, hábitats y ecosistemas, entre otros (García, Castro y Cárdenas, 2005).

La lingüística ha desarrollado diversas estrategias de agrupamiento por conglomerados y su visualización en dendrogramas. Un interesante uso de esto es el desarrollado por Martha de Alba (2004) para analizar las representaciones sociales del espacio urbano en la ciudad de México. Tal vez, lo más útil del método es que permitió a la autora aproximarse a los mundos lexicales, no sólo a través del análisis de la organización y distribución de las palabras principales co-ocurrentes en los enunciados simples de un texto, sino que además le facilitó descubrir la distribución estadística de sucesiones de palabras que componen los enunciados de un texto los cuales se agrupaban en clasificaciones basadas en las frecuencias de asociación de palabras del texto, para con ello analizar la forma en que el discurso se estructura bajo un análisis jerárquico descendiente de manera general. Según la autora, esto permite analizar una gran “cantidad de texto” bruto relativizando las preconcepciones del investigador⁵.

La arqueología y la antropología también han encontrado un uso valioso de las técnicas cluster y de dendrogramas. Los investigadores José Cocilovo *et al.* (2001) y Francisco Rothhammer *et al.* (2003) han combinado en sus análisis sobre el origen de las poblaciones antiguas la información lingüística y la proveniente de la etnohistoria clásica con el uso de análisis ADN mitocondrial antiguo. De este modo, Rothhammer y sus colegas usan el dendrograma como una manera de evidenciar las relaciones entre grupos humanos agrupándolos y patentizando las vinculaciones ancestrales. Por su parte, José Cocilovo y sus colaboradores hicieron un procedimiento similar para analizar las afinidades biológicas entre grupos prehistóricos empleando distancias biológicas –aunque no únicos, pues se usaron también factores culturales junto con los macro evolutivos–.

⁴ Algunos ejemplos pueden ser Huaylinos Villalva, Quispitúpac Quispitúpac y Martínez Luján (2003), Vilanova Prieto y Stutz (2006), Neto y Moita (1998), López Caballero y Pérez Suárez (1999) o Chinelatto Neto y Castro (1999).

⁵ Otro trabajo que usa ALCESTE es el de Dos Santos Aléssio (2004).

Las vías de comunicación entre la antropología y la historia fueron muchas, pero particularmente interesante es rescatar las desarrolladas con la historia antigua. Así por ejemplo, el Grupo CEIPAC del Área de Historia Antigua de la Universitat de Barcelona y, en particular, su investigador Antonio Aguilera Martín (1998) han evidenciado las bondades de la aplicación del análisis multivariable y de conglomerados para la historiografía. En efecto, en un trabajo sobre la caracterización de la cerámica para determinar las redes de comercio en la antigüedad, Aguilera evidenció como esta técnica se podía utilizar en el análisis físico-químico de las pastas cerámicas (estudio epigráfico) y como no sólo se alcanzaba una clasificación tipológica de las cerámicas a través de parámetros numéricos, sino que también se podía determinar la procedencia de artefactos, permitiendo deducir así las posibles redes comerciales.

Un buen ejemplo de la aplicación de los análisis cluster al estudio de lo regional es el trabajo conjunto de Díaz Díez, Cruces Pastor y Morillas Raya (1997). En este artículo los autores utilizan la información estadística referente a los indicadores regionales europeos y realizan una ingeniosa clasificación y caracterización de los espacios socioeconómicos. Para ello utilizan primero un análisis cluster con algunos indicadores para evaluar las posibilidades de clasificación así como de establecer el número de grupos a formar. Los autores parten del hecho que los avances en la informática se han acrecentado tanto, que han permitido a los investigadores sociales aplicar diversas estrategias multivariantes de análisis de datos. Para ellos, el análisis de conglomerados o cluster es un nombre genérico de técnicas que sirven para identificar una gran variedad de procedimientos muy beneficiosos para formalizar clasificaciones de cualquier tipo. Piensan entonces que el procedimiento es riguroso, pues supone procesos de definición teórico metodológicos tanto de las variables como de los indicadores a utilizar. Los autores demuestran también la posibilidad de complementar los análisis con indicadores estadísticos o su presentación más gráfica en mapas, lo cual permite restar aridez al lector y ganar con ello una mayor comprensión de los fenómenos descritos. En la actualidad los manuales internacionales de estadística para historiadores si bien no han inserto los análisis cluster sí han incluido otros tipos de análisis de clasificación como los escalímetros Guttman.

En suma, en la actualidad la socio-demografía, la economía, la psicología recurren a las técnicas cluster y dendrográficas⁶ lo cual evidencia diversas prácticas, procedimientos y estrategias metodológicas que se acercan a diversas inquietudes que poseen los investigadores de lo regional. Este breve recuento permite visualizar otras aproximaciones y escenarios de trabajos para abordar la complejidad de lo regional. En efecto, los historiadores, más que cualquier otro tipo investigador dentro de las ciencias sociales, se caracterizan por tener diversas dificultades de tipo operacional, como son la fragmentación de las fuentes, la escasez de las mismas para ciertos hechos sociales, el predominio de los datos cualitativos sobre los de tipo cuantitativo, en especial cuando más se retrocede en el tiempo y el número de variables intervinientes en el análisis histórico, ante ello debe recurrir a diversas estrategias que le permitan juzgar y valorar las posibilidades de las problemáticas estudiadas para iniciar procesos más complejos asociados a la interpretación.

⁶Consúltense, por ejemplo, los trabajos de Kageyama y Troncoso Leone (1999), Barreiro, López, Losada y Ruzo (2002), o Carvalho, Accioly Júnior y Raffin (2006).

Los trabajos cluster y de dendrográficos reseñados evidencian posibles vías de operacionalización de los diversos problemas que enfrenta no sólo el historiador sino también el científico social y con ello introducirse en las posibilidades de triangulación metodológica con otros elementos de tipo contextual. De este modo, tanto, el historiador como el investigador de lo regional podrán contar con una herramienta más, en su aspiración de realizar análisis comparativos, que al menos les permitan proyectar ciertas generalizaciones básicas para poder comprender las relaciones entre los procesos estudiados y con ello ofrecer interpretaciones más abarcativas que las ofrecidas por la historia positivista tradicional.

2. PROCEDIMIENTOS Y DESARROLLO DEL ANÁLISIS DENDROGRÁFICO

El siguiente apartado tomará la forma de una descripción técnica del uso de los análisis cluster y de los dendrogramas. Dado que es una herramienta poco usual en el taller del historiadores y los denominados científicos sociales procuraremos ir delineando las fases, pasos, problemas, ventajas y límites que deben tenerse a la hora de usarla en los análisis de lo regional, en especial, por que con ella se desea construir una estrategia teórico metodológica que permita entender e interpretar las diferencias entre lugares, territorios, y región, aspectos que se constituyen en un elemento fundamental para los investigadores actuales, máxime con el desarrollo de espacios móviles y el entrecruzamiento de fenómenos sociales.

La interpretación de los espacios-tiempos múltiples en forma comparada no es sólo un reto para el historiador y el estudioso de lo regional actual en Centroamérica y en Iberoamérica, sino un gran desafío en la construcción de instrumentos y herramientas analíticas que permitan una discusión de base interdisciplinaria y de la creación de un proyecto colectivo de análisis de las grandes disparidades regionales en Iberoamérica en general y de Centroamérica en particular, por lo que implica la unión de fuerzas entre los diversos estudiosos. Hoy se requiere convocar investigadores, problemáticas, estudios, enfoques y estrategias en todas las ciencias sociales, para ir delineando no sólo los procesos económicos y sus repercusiones inmediatas en el espacio, sino también las diversas imágenes y representaciones del mundo y sus partes, el cómo se enfrentan y organizan los espacios, las relaciones sociales y los distintos elementos naturales, económicos, sociales y culturales; así como las causas y los porqué se han constituido históricamente las disimetrías entre las relaciones sociales y entre la relación del territorio con la sociedad que lo utiliza y transforma.

Los análisis de conglomerados jerárquicos o análisis cluster junto con los dendrogramas son estrategias metodológicas que trabajan bajo una serie de principios, entre ellos que todo fenómeno es afectado por diversos procesos que obligan al investigador a considerar múltiples variables y, por ende, considerar los procedimientos de análisis multivariantes para detectar el impacto que tienen tanto en el proceso como con los actores sociales que están inmersos en ellos.

Los objetivos de los análisis de conglomerados jerárquicos y los dendrogramas son sencillos: proporcionar un método que considere los datos multivariantes; tomar decisiones sobre como procesar la información disponible, en especial para construir tipologías primarias que nazcan de los mismos procesos estudiados; realizar interacciones entre un número elevado de variables (ya sea cualitativas o cuantitativas) con un número relativamente alto de casos; clasificar los datos o terri-

torios en grupos de observación más pequeños, pero pertinentes con el fin de discriminar la relevancia de los procesos sociales y con ello facilitar la interpretación; suministrar diversos criterios de para el agrupamiento (cluster) de personas, empresas, productos, procesos y espacios, entre otros, de forma tal que cada conglomerado este ordenado según un criterio de jerarquización. Los objetivos anteriores evidencian cuales son los límites y posibilidades de estas estrategias metodológicas, por lo que podríamos afirmar que encausan al investigador en lo que pueden o no conseguir con este tipo de estrategias. En esencia, con estos procedimientos se busca crear agrupamientos que lleven a la definición de tipologías y categorizaciones.

Tanto los análisis de conglomerados jerárquicos como dendrogramas recurren a diversas rutinas o procedimientos que se pueden ejecutar en la mayoría de los paquetes estadísticos de empresa como son el SPSS, SAP, STAHPGRAPHIC y los de libre acceso como OpenStat4, R o Epiinfo⁷.

3. PASOS Y ETAPAS DE LA INVESTIGACIÓN

3.1. *La estrategia metodológica no es el centro del proceso de investigación*

En el caso de las ciencias sociales y la historia el primer elemento a considerar es el de contar con una muy buena problemática. Logrado esto se pueden ir dando otros pasos, tales como discernir los enfoques teóricos a utilizar, el tipo de fuentes requeridas y desde luego ir aclarando cuáles van a ser las estrategias de investigación que den cuenta del problema planteado. Así la máxima para iniciar cualquier estrategia metodológicas es el de aceptar que esta no es el centro del proceso de investigación, pero que si ayuda a conseguir los propósitos perseguidos⁸.

De ese modo, antes de plantearse el análisis de conglomerados jerárquicos y los dendrogramas debe saberse para qué se utilizará, así por ejemplo, si se desea tener una visión comparada de un elemento desde la totalidad, si se espera probar una teoría con muchos casos, o por el contrario definir una a través de pocos casos con muchas variables o simplemente crear tipologías o clasificaciones que producirán nuevas preguntas o problemas.

Si bien es cierto, el proceso anterior no es automático, el solo hecho de interrogarse sobre el problema y la forma en que puede abordarse ahorra una considerable cantidad de tiempo. Entre más se esté familiarizado el investigador con la complejidad del problema mejor podrá discernir sobre los pasos siguientes que deberá dar en el proceso investigativo y con ello apreciar cuáles serán las fuentes de información que debe buscar, cuáles serán las preguntas más atinentes, cuáles serán los procesos de recolección de información más idóneos, qué métodos serán necesarios o los más adecuados y cómo se podrá analizará los datos que se obtengan.

⁷Dado el carácter abierto proporcionamos las direcciones electrónicas para poderlos bajar <http://www.statpages.org/miller/openstat/> ; <http://www.r-project.org/> o <http://www.cdc.gov/epiinfo/epiinfo.htm>.

⁸Sobre esto es importante señalar que entre más se conozca un tema mejor. El proceso de afinar la idea y problema de investigación será más eficiente y rápido si contamos con una perspectiva amplia de enfoques, aportes, logros, estrategias metodológicas y fuentes utilizadas en el tema de nuestro interés, para así visualizar cuáles serán los aportes propios a esa temática. Los temas ya investigados siempre requieren de planteamientos más específicos y buscar de nuevos documentos escritos o enfoques que reporten avances en el conocimiento. Ello permitirá detectar cuales cuestiones están más estructuradas y formalizadas y cuales no, así como los problemas poco investigados, mal tratados o del todo no investigados. Es primordial definir en cuál de las categorías anteriores puede caer la investigación que se esta planteando.

Cuadro # 1 :
Tabla de variables e indicadores

| Variable | Explicación e indicadores propuestos |
|---|---|
| País | Se refiere al país que se estudiara |
| Índice de desarrollo democrático 2005 | Índice desarrollado por la Fundación Konrad Adenauer y Polilat.com que considera 31 indicadores agrupados en 4 dimensiones (1) que miden las siguientes variables: "la legalidad del régimen democrático", el "respeto de los derechos políticos y libertades civiles", la "calidad institucional y el grado de eficiencia política" y el "ejercicio del poder efectivo para gobernar". (0 el más bajo) |
| Tipo gobierno | Indicador del tipo de República (Unitaria o Federal) |
| Tamaño de país | Rango de los países según su extensión territorial (rango 1 = más grande y 3 = más pequeño) |
| Tasa de mortalidad de menores de cinco años (2005) | Tasa que se estableció por 100.000 en el año 2005 |
| Población (2005) | Número de habitantes medido por miles |
| % de la pobl línea de extrema Pobreza (1999-2005) | Porcentaje de la población en línea de pobreza. Dado que los datos eran parciales se consideraron los promedios entre 1999 y el 2005 |
| Balanza de Pagos Reserva de Divisas 2005 según FMI (2005) | Registra las operaciones comerciales, de servicios y de movimientos de capitales llevadas a cabo durante un período de tiempo determinado. La diferencia entre ingresos y pagos de una determinadas sub balanza se denomina saldo. |
| Acceso electricidad 1998-2005 | Registra el acceso a este servicio para la totalidad del país |
| Acceso de agua potable a las casas 2000-2005 | Registra el acceso a este servicio para la totalidad del país |
| Índice de Libertad económica 2005 | Indicador elaborado por la Fundación Heritage para analizar la forma como el gobierno estimula o restringe la capacidad de las empresas del sector privado para operar independientemente |
| Índice Libertad Eco - y factor de corrupción (2005) | Además de medir el nivel de interferencia del gobierno en la economía evalúa el impacto de la corrupción. |
| Índice libertades civiles (2005) | Índice creado por la Freedom House el cual integra dos componentes: derechos políticos y libertades civiles (1 es el más alto e indica un país libre, el 2 establece los países parcialmente libre y el 3 los no libres). |
| Sostenibilidad energética (2006) | Índice muestra la participación de fuentes renovables en la oferta total de energía |
| Los Riesgos de muerte maternal (año 2004) | Indica la posibilidades de una muerte entre el número de casos probables que se registre esta |
| El nivel de alfabetización adultos (005) | índice o tasa de alfabetización donde se destaca el porcentaje de la población que sabe leer o escribir después de determinada los 15 años |

Fuente: Elaboración propia a partir de los datos de la USAID (http://quesdb.usaid.gov/cgi-bin/broker.exe?program=lacprogs.dg_2.sas&service=default&sscode=HER210011) y Anuario Estadístico de América Latina y el Caribe de la CEPAL (2006 http://websie.eclac.cl/anuario_estadistico/anuario_2006/esp/index.asp).

3.2. De la definición del problema a la operacionalización

Una vez planteado el problema de investigación se debe buscar una correspondencia entre los objetivos que persigue la investigación, las preguntas de investigación que darán pie a las hipótesis, y entre estas y las variables e indicadores que se usarán, para ello siempre se recomienda hacer uso del denominado cuadro de concordancia⁹. A falta del mismo se puede usar una tabla que variables con sus indicadores (Cuadro 1) que dará pie a la recolección y sistematización de datos, los cuales a su vez harán posible la creación de la base de datos (Cuadro 2).

El Cuadro 1 nos indica la necesidad de ir operacionalizando las variables, como muy bien señala María de los Ángeles Cea, en este punto el investigador al plantearse un problema, al recolectar los datos y al sistematizarlos en variables debe asegurarse de cumplir con una serie de condiciones para que estas variables puedan ser correctamente procesadas tanto en la estrategia metodológica adoptada como en el logro de la rigurosidad en la interpretación. El Cuadro 1 nos evidencia que las variables a utilizar deben ser precisas y concretas. Como bien indica Cea (2001: 126-7) las condiciones básicas que deben poseer las variables son: *exhaustividad* es decir que se comprendan el mayor número de atributos para que ninguna observación quede sin poder clasificarse; *exclusividad* en este caso que los distintos atributos que componen una variable sean mutuamente excluyentes y *precisión* con lo que se logra realizar el mayor número de distinciones posibles.

Cuadro # 2:
Base de Datos, gobernabilidad en América Latina.

| Pais | Índice de desarrollo democrático 2005 | tipo gobierno | tamaño de país | Tasa de mortalidad de menores de cinco años en el 2005 | Población en miles de hab | % de la polid. línea de extrema Pobreza 1999-2005 | Balanza de Pagos Reserva de Divisas 2005 según FMI | acceso electricidad 1998-2005 | acceso de agua potable a las casas 2000-2005 | Índice de Libertad económica 2007 | Índice Eco- y Libertad factor de corrupción | Índice libertades civiles | Sostenibilidad energética 2006 | Riesgos de Maternal jaño (2004) | El nivel de alfabetización adultos |
|---------------|---------------------------------------|---------------|----------------|--|---------------------------|---|--|-------------------------------|--|-----------------------------------|---|---------------------------|--------------------------------|---------------------------------|------------------------------------|
| Argentina | 4,337 | Federal | 1 | 18 | 37,032 | 3.4 | -7,172.4 | | | 57.5 | 28 | 2 | 77.8 | 410.00 | 97.200 |
| Bolivia | 3,528 | Federal | 2 | 69 | 8,329 | 9.7 | -1,284.5 | 69.1 | 67.7 | 55.0 | 25 | 3 | 63.7 | 47.00 | 86.700 |
| Brazil | 3,82 | Federal | 1 | 34 | 176,406 | 4.3 | -36,665.3 | 96.6 | 80.7 | 60.9 | 37 | 2 | 86.6 | 180.00 | 88.600 |
| Chile | 16,435 | Republica | 2 | 8 | 15,741 | 2.1 | -5,079.0 | 98.6 | 92.1 | 78.3 | 73 | 1 | 74.6 | 160.00 | 95.700 |
| Colombia | 2,993 | Federal | 1 | 21 | 42,105 | 8.3 | -23.2 | 94.9 | | 60.5 | 40 | 3 | 82.4 | 240.00 | 92.800 |
| Costa Rica | 8,51 | Republica | 3 | 13 | 4,024 | 2.9 | -1,015.6 | 98.9 | 98.5 | 65.1 | 42 | 1 | 86.0 | 690.00 | 94.900 |
| Ecuador | 3,658 | Republica | 2 | 26 | 12,646 | 7.9 | 212.3 | 89.0 | | 55.3 | 25 | 3 | 66.4 | 210.00 | 91.000 |
| El Salvaor | 5,053 | Republica | 3 | 28 | 6,278 | 9.4 | -71.6 | 79.5 | 57.0 | 70.3 | 42 | 3 | 82.5 | 180.00 | 86.600 |
| Guatemala | 1,648 | Republica | 3 | 45 | 11,385 | 18.5 | -252.3 | 76.1 | 76.2 | 61.2 | 25 | 4 | 82.1 | 74.00 | 65.100 |
| Honduras | 4,332 | Republica | 3 | 41 | 6,417 | 26.3 | -300.1 | 61.9 | 75.7 | 60.3 | 26 | 3 | 73.2 | 190.00 | 80.000 |
| México | 5,522 | Federal | 1 | 28 | 98,872 | 3.6 | -2,152.0 | 98.5 | 89.6 | 65.8 | 35 | 2 | 67.6 | 370.00 | 91.600 |
| Nicaragua | 4,032 | Republica | 3 | 38 | 5,071 | 49 | -131.7 | 66.7 | | 62.7 | 26 | 3 | 71.3 | 88.00 | 76.700 |
| Paraná | 6,918 | Republica | 3 | 24 | 2,856 | 6.9 | -152.5 | | | 65.9 | 35 | 2 | 69.5 | 210.00 | 91.900 |
| Paraguay | 4,493 | Republica | 2 | 24 | 5,496 | 13.1 | -381.8 | 94.8 | 58.9 | 56.8 | 21 | 3 | 69.9 | 120.00 | 93.500 |
| Perú | 3,126 | Republica | 2 | 29 | 25,662 | 9.2 | -3,065.6 | 72.3 | 64.9 | 62.1 | 35 | 3 | 83.5 | 73.00 | 87.900 |
| R. Dominicana | 3,823 | Republica | 3 | 32 | 8,373 | 10.4 | -295.5 | | 71.7 | 56.7 | 30 | 2 | 69.3 | 200.00 | 87.000 |
| Uruguay | 8,355 | Republica | 3 | 17 | 3,337 | 1.0 | 10.8 | | | 69.3 | 59 | 1 | 84.5 | 1300.00 | 96.000 |
| Venezuela | 2,581 | Federal | 1 | 19 | 24,170 | 7.4 | -4,672.0 | 99.1 | 92.6 | 47.7 | 23 | 4 | 32.1 | 300.00 | 93.000 |

Fuente: Elaboración propia a partir de los datos de la USAID (http://quesdb.usaid.gov/cgi-bin/broker.exe?_program=lacprogs.dg_2.sas&_service=default&sscode=HER210011) y Anuario Estadístico de América Latina y el Caribe de la CEPAL (2006 http://websie.eclac.cl/anuario_estadistico/anuario_2006/esp/index.asp).

⁹ Al respecto véase: <http://video.google.es/videoplay?docid=-270035968091537485&hl=es> (consultado el 10/10/2007).

No entraremos en los diferentes tipos de variables (independiente y dependiente), tan solo se indicará que la variable es una propiedad que adquiere diferentes valores y cuya variación es susceptible de ser medida a través de indicadores, como se puede observar en el Cuadro 1, a través de rangos, índices, tasas y cantidades, entre otras. No obstante, para estos efectos, del análisis de conglomerados debe tenerse claro las características que tendrán las variables, pues estas pueden ser *numéricas* (números reales en general como son los casos las variables referidas al acceso electricidad y agua); *nominales* (variables discretas pero que no tienen un orden especificado como por ejemplo el tipo de gobierno); *ordinales* (variables discretas con una relación de orden como por ejemplo alto, medio y bajo como son los casos de tamaño de país y el índice libertades civiles), o *binarias* (que solo pueden tomar dos estados posibles o dicotómicas). Por su parte los indicadores se pueden considerar tanto como un sistema adecuado de recogida de información como de especificación de la variable de estudio, por ello deben ser relevantes, objetivos, útiles, precisos, y congruentes con la problemática y con los objetivos planteados¹⁰.

Lo anterior es fundamental a la hora de sistematizar la fuente en una base de datos como la presente en el Cuadro 2, de este modo al pasarlo a cualquier paquete estadístico permitirá crear rangos de datos, discernir cuales datos son numéricos, cualidades, ordinales o dicotómicos y con ello evitar posibles incompatibilidades en las unidades de medida; mezclar indiscriminadamente variables de distinto tipo –aunque esto puede realizarse tomando ciertas previsiones que veremos más adelante en el análisis de conglomerados jerárquicos– o considerar errores con variables faltantes que no pueden ser procesadas por el paquete estadístico –en especial por tener inconsistencias o estar mal digitados aspecto que usualmente se denomina suciedad de los datos de la base–.

3.3. De la base de datos hasta su sistematización

Los análisis de conglomerados jerárquicos y los dendrogramas requieren que la base de datos este lo más limpia y ordenada posible con los criterios arriba señalados¹¹. Como se puede apreciar en el Cuadro 2 la base consta de diferentes variables, la mayoría de ellas de tipo numérica (población, porcentaje de población en línea de extrema pobreza, balanza de pagos, reserva de divisas, etc.) y solo dos de tipo textual: país y tipo de gobierno.

Una vez registrado todos los datos en la base con sus rangos, categorías o cifras se debe aplicar un análisis estadístico simple y de frecuencias, el cual permitirá ver la consistencia tanto del digitado de datos como la estabilidad de los mismos y sus magnitudes. Ello permitirá crear otros rangos o nuevas categorías o atributos con los cuales se pueda trabajar mejor con los datos. Las Ilustraciones 1 y 2 evidencian parcialmente el ordenamiento y coherencia de los datos.

¹⁰Esta de más señalarlo, pero los indicadores no pueden ser conceptualizados como mera fuente o criterio de obtención de información, sino que se deben ver en razón a su utilidad a la hora de consideración la relevancia de las fuentes, con el fin de que estas sean necesarias y oportunas al planteamiento y los objetivos de la investigación. En general, ayudan a discernir sobre la fiabilidad, relevancia, utilidad, accesibilidad, concreción, precisión, aplicabilidad, significatividad y adecuación a la problemática estudiada.

¹¹En este trabajo se parte del concepto básico de base de datos, el cual se refiere a un conjunto de datos que pertenecen a un mismo contexto y que se almacenan sistemáticamente para su uso consultivo, correlacionado o de análisis.

Ilustración # 1: Ejemplo del procedimiento de estadística descriptiva en SPSS 13.

| → Descriptives | | | | | |
|--|----|---------|---------|----------|----------------|
| Descriptive Statistics | | | | | |
| | N | Minimum | Maximum | Mean | Std. Deviation |
| Indice de desarrollo democrático 2005 | 18 | 1,648 | 10,435 | 4,84244 | 2,301857 |
| tamaño de país | 18 | 1 | 3 | 2,17 | ,857 |
| Tasa de mortalidad de menores de cinco años en el 2005 | 18 | 8 | 69 | 28,56 | 13,929 |
| Población en miles de hab | 18 | 2856 | 170406 | 27092,78 | 42579,241 |
| Indice Libertad Eco - y factor de corrupción | 18 | 21 | 73 | 34,83 | 13,334 |
| Indice libertades civiles | 18 | 1 | 4 | 2,50 | ,924 |
| El nivel de alfabetización adultos | 18 | 69100 | 97200 | 88611,11 | 7609,737 |
| Valid N (listwise) | 18 | | | | |

Fuente: Base de datos Gobernabilidad en América Latina

Como se puede evidenciar de la Ilustración 2 la cantidad de datos puede ser larga por lo que al igual que en la estadística simple se pueden crear rangos, además de limpiar en caso necesario los denominados *Missing System* o valores perdidos. Todos los programas estadísticos tienen la función de recodificar variables y agruparlas según los criterios de magnitudes que se prefieran, tal y como se observa en la Ilustración 3. Usualmente se construyen de tres a cinco grandes grupos. Para los científicos sociales en particular, estas funciones son vitales, pues permiten mantener la riqueza de los datos alfa numéricos recolectados al mismo tiempo que se pueden crear nuevas variables.

Un aspecto importante fue que los datos originales fueron convertidos en rangos para ello se crearon rangos de uno a cuatro, donde el uno siempre muestra una magnitud positiva y el cuatro la más negativa de la escala. Dado el carácter algorítmico de los análisis cluster jerárquico y dendrogramas cuyo propósito final es el de agrupar datos de un conjunto de entrada atendiendo a diferentes criterios, es que es posible introducir variables de diferente tipo (cuantitativas y cualitativas), pues al final con esta estrategia no se pretende tener una medida matemática, sino una visualización: de un agrupamiento, el cual se espera sea ordenado con el fin de descubrir características nuevas o relaciones que no se habían previsto o notado de antemano. No obstante, siempre es oportuno contar con un criterio uniforme de ahí la previsión de crear nuevas variables a través de los rangos. Lo anterior es esencial, pues la estrategia de análisis cluster jerárquico y dendrogramas desarrolla lo que Kant llamó juicio reflexionante, es decir, una operación que se distancia del supuesto objetivismo de la teoría para llevar al investigador a juzgar, valorar y explorar la relación entre realidad y teoría, y con ello estimar tanto las posibles vías de interpretación como criterios metodológicos a utilizar (cuantitativos, cualitativos o de triangulación metodológica).

**Ilustración # 2 :
Ejemplo del procedimiento de frecuencias en SPSS 13.**

| Indice Libertad Eco - y factor de corrupción | | | | | |
|--|--------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 21 | 1 | 2,7 | 5,6 | 5,6 |
| | 23 | 1 | 2,7 | 5,6 | 11,1 |
| | 25 | 3 | 8,1 | 16,7 | 27,8 |
| | 26 | 2 | 5,4 | 11,1 | 38,9 |
| | 28 | 1 | 2,7 | 5,6 | 44,4 |
| | 30 | 1 | 2,7 | 5,6 | 50,0 |
| | 35 | 3 | 8,1 | 16,7 | 66,7 |
| | 37 | 1 | 2,7 | 5,6 | 72,2 |
| | 40 | 1 | 2,7 | 5,6 | 77,8 |
| | 42 | 2 | 5,4 | 11,1 | 88,9 |
| | 59 | 1 | 2,7 | 5,6 | 94,4 |
| | 73 | 1 | 2,7 | 5,6 | 100,0 |
| | Total | 18 | 48,6 | 100,0 | |
| Missing | System | 19 | 51,4 | | |
| Total | | 37 | 100,0 | | |

Fuente: Base de datos Gobernabilidad en América Latina

No obstante, reiteramos que es oportuno usar un tipo de variable y ensayar diversos tipos de agrupamientos con resultados basados con datos cualitativos y cuantitativos, buscando establecer una triangulación de los datos, enfoques y perspectivas. Esto con el fin de que con la clasificación eventualmente se puedan hacer interpolaciones con valores numéricos. En todo caso, es fundamental comprender que los análisis cluster jerárquicos y los dendrogramas lo único que pretenden agrupar y clasificar los datos¹².

Con el propósito de ejemplificar lo anterior, podemos tomar una hipótesis de Seymour Martin Lipset que establece que “cuanto más próspera sea una nación, tanto mayores son las posibilidades de que se mantendrá una democracia” (cit. en Filgueira y Rossel, 2005). Como señalan Filgueira y Rossel, Lipset propone que entre más positivas sean las condiciones de industrialización, educación y urbanización, entre otras, más estable y viable será la democracia. Dado que es una hipótesis que se quiere comprobar y visualizar comparativamente se puede tomar la base de datos sobre gobernabilidad que aparece en el Cuadro 2. Para ello se utilizarán las siguientes variables –transformadas en rangos tal y como se explicó en los párrafos anteriores–:

¹² Hay que recordar que los cluster y los dendrogramas son una modalidad de análisis multivariado, que permiten reducir una serie de variables y casos a un conjunto menor de factores, que posibilitan a su vez, como ya se ha señalado, detectar rasgos comunes denominadores.

- País
- Índice de desarrollo democrático 2005
- Porcentaje de la población en línea de extrema Pobreza 1999-2005
- Índice de libertad económica 2007
- Índice libertad económica - y factor de corrupción
- Índice libertades civiles
- Sostenibilidad energética 2006
- Riesgos de muerte maternal
- Y el nivel de alfabetización adultos

Bien, la variables pueden ser otras pero fueron las que se consiguieron (como suele suceder en la investigación principalmente histórica). Una vez establecido lo anterior se procede a realizar el análisis cluster tal y como se observa en la Ilustración 4.

El proceso resultante es la creación tanto del dendrograma (Ilustración 5) como las membresías al cluster (Cuadro 3). De esta forma, el investigador puede ir realizando diferentes procedimientos para visualizar cuales factores se ajustan a la hipótesis planteada. Muchas veces es recomendable realizar análisis anova para comprobación de hipótesis con el fin de ir decantando variables según se ajusten o no al modelo.

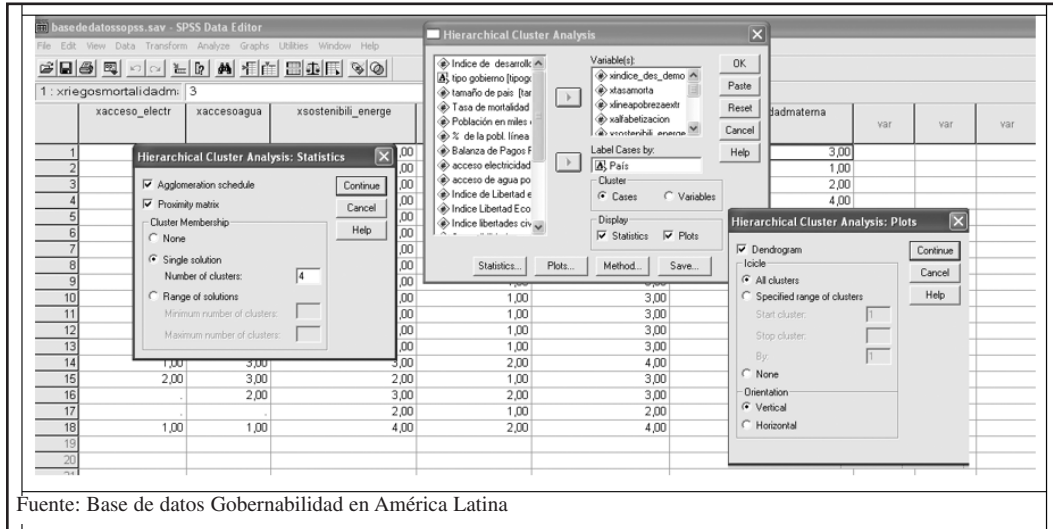
**Ilustración # 3 :
Visualización de las funciones de remodificación en SPSS 13¹³.**

| | País | Tamaño de población | Tasa de mortalidad de neonatos | Población en línea de extrema pobreza 1999-2005 | Balanza de pagos Reservas 2005 | Acceso a electricidad | Acceso a agua potable | Índice de libertad económica | Índice de libertad económica | Índice de libertades civiles | Sostenibilidad energética | Riesgos de muerte maternal | | | |
|----|---------------|---------------------|--------------------------------|---|--------------------------------|-----------------------|-----------------------|------------------------------|------------------------------|------------------------------|---------------------------|----------------------------|---------|---------|--------|
| 1 | Argentina | 1 | 18 | 37032 | 3.4 | -7,172.4 | 67.5 | 57.5 | 29 | 2 | 77.8 | 410.00 | | | |
| 2 | Bolivia | 2 | 69 | 8329 | 9.7 | -1,284.5 | 69.1 | 67.7 | 55.0 | 25 | 3 | 63.7 | 47.00 | | |
| 3 | Brasil | 1 | 34 | 170406 | 4.3 | -30,665.3 | 96.6 | 80.7 | 60.9 | 37 | 2 | 80.6 | 140.00 | | |
| 4 | Chile | 2 | 8 | 15211 | 2.1 | -2,079.0 | 96.6 | 92.1 | 78.3 | 73 | 1 | 74.6 | 1100.00 | | |
| 5 | Colombia | 1 | 21 | 42105 | 8.3 | -23.2 | 94.9 | 98.5 | 60.5 | 40 | 3 | 82.4 | 240.00 | | |
| 6 | Costa Rica | 3 | 13 | 4024 | 2.9 | -1,015.6 | 98.9 | 98.5 | 65.1 | 42 | 1 | 86.0 | 690.00 | | |
| 7 | Ecuador | 3,658 | Republica | 2 | 26 | 12646 | 7.9 | 212.3 | 89.0 | 55.3 | 25 | 3 | 66.4 | 210.00 | |
| 8 | El Salvador | 5,053 | Republica | 3 | 28 | 6278 | 9.4 | -71.6 | 79.5 | 67.0 | 70.3 | 42 | 3 | 82.5 | 180.00 |
| 9 | Guatemala | 1,648 | Republica | 3 | 45 | 11385 | 18.5 | -252.3 | 76.1 | 76.2 | 61.2 | 25 | 4 | 82.1 | 74.00 |
| 10 | Honduras | 4,332 | Republica | 3 | 41 | 6417 | 26.3 | -300.1 | 61.9 | 75.7 | 60.3 | 26 | 3 | 73.2 | 190.00 |
| 11 | México | 5,522 | Federal | 1 | 28 | 98672 | 3.6 | -2,152.0 | 98.5 | 89.6 | 65.8 | 35 | 2 | 67.6 | 370.00 |
| 12 | Nicaragua | 4,032 | Republica | 3 | 38 | 5071 | 19 | -131.7 | 66.7 | 62.7 | 26 | 3 | 71.3 | 88.00 | |
| 13 | Panamá | 6,918 | Republica | 3 | 24 | 2856 | 8.9 | -152.5 | 65.9 | 65.9 | 35 | 2 | 69.5 | 210.00 | |
| 14 | Paraguay | 4,493 | Republica | 2 | 24 | 5496 | 13.1 | -381.8 | 94.8 | 68.9 | 56.8 | 21 | 3 | 69.9 | 120.00 |
| 15 | Perú | 3,126 | Republica | 2 | 29 | 25862 | 9.2 | -3,065.6 | 72.3 | 64.9 | 62.1 | 35 | 3 | 83.5 | 73.00 |
| 16 | R. Dominicana | 3,823 | Republica | 3 | 32 | 8373 | 10.4 | -295.5 | 71.7 | 56.7 | 30 | 2 | 69.3 | 200.00 | |
| 17 | Uruguay | 8,355 | Republica | 3 | 17 | 3337 | 1.0 | 10.8 | 69.3 | 69.3 | 59 | 1 | 84.5 | 1300.00 | |
| 18 | Venezuela | 2,581 | Federal | 1 | 19 | 24170 | 7.4 | -4,672.0 | 99.1 | 92.6 | 47.7 | 23 | 4 | 32.1 | 300.00 |
| 19 | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | |

Fuente: Base de datos Gobernabilidad en América Latina

¹³ En la gráfica se observa el procedimiento de recodificación en el programa comercial SPSS. Para los análisis cluster esta función es fundamental, pues permite crear códigos nuevos en una nueva variable, respetando la base original. Ello permite tener un control en el manejo de los rangos, ensayando diferentes vías para realizarlos.

Ilustración # 4 : Análisis cluster jerárquico.



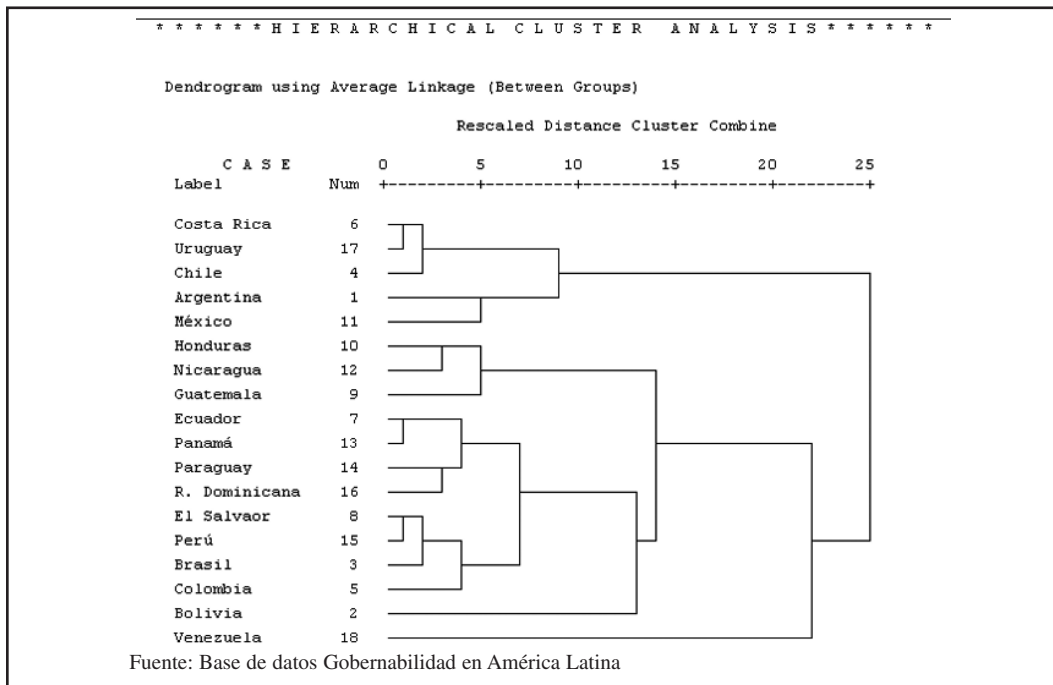
Fuente: Base de datos Gobernabilidad en América Latina

Como se observa de la Ilustración 5, el agrupamiento se basó en la búsqueda de una medida de similitud utilizada por defecto para datos de intervalos, su valor mínimo es cero, pero no tiene un valor máximo. Lo anterior posibilita vincular datos y unir elementos. En este caso, el SPSS realizó varios sucesos o procedimientos, el primero fue el de calcular las distancias entre todos los pares de objetos, asumiendo que cada objeto constituye un cluster. El siguiente paso fue el buscar los dos clusters más cercanos los cuales se juntan y constituyen uno solo, el proceso se repite hasta que no quedan pares de comparación. Y finalmente el resultado gráfico es un árbol de relaciones el cual también puede ser visualizado por países según su membresía (Cuadro 3)¹⁴.

Los datos obtenidos pueden ser fácilmente ubicados en un mapa como se visualiza en la Ilustración 6 o en un gráfico de dispersión (Gráfico 1, el cual se logra con un análisis de medidas, según se puede observar en la Ilustración 7). Como se puede observar, tanto en el gráfico como en la ilustración, los dendrogramas son la parte gráfica de los análisis de conglomerados y tienen una función de “paratexto”, es decir de recuperar los datos y procesos numéricos a través de un “texto visual” de rápida comprensión, que permite tanto al investigador como al lector percatarse de las relaciones establecidas entre los espacios (como es el caso del ejemplo utilizado) o de procesos y sujetos, que también pueden ser clasificados.

¹⁴El cálculo realizado por los análisis de conglomerados es de distancia euclídea al cuadrado, con este método se puede visualizar los clusters o conglomerados que se han ido constituyendo. Una vez logrado lo anterior se utiliza una matriz de proximidad, el cual revela el historial de la conglomeración de la distancia euclídea cuadrado, la cual muestra como las distancias aumentan conforme avanzan las etapas. Todo el proceso se visualiza en un gráfico de dendrograma el cual muestra los grados de homogeneidad interna (dentro de cada conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados) por lo que los objetos dentro de los conglomerados estarán muy próximos, y los diferentes se hallarán cada vez más alejados.

Ilustración # 5 : Cluster jerárquico usando los casos.

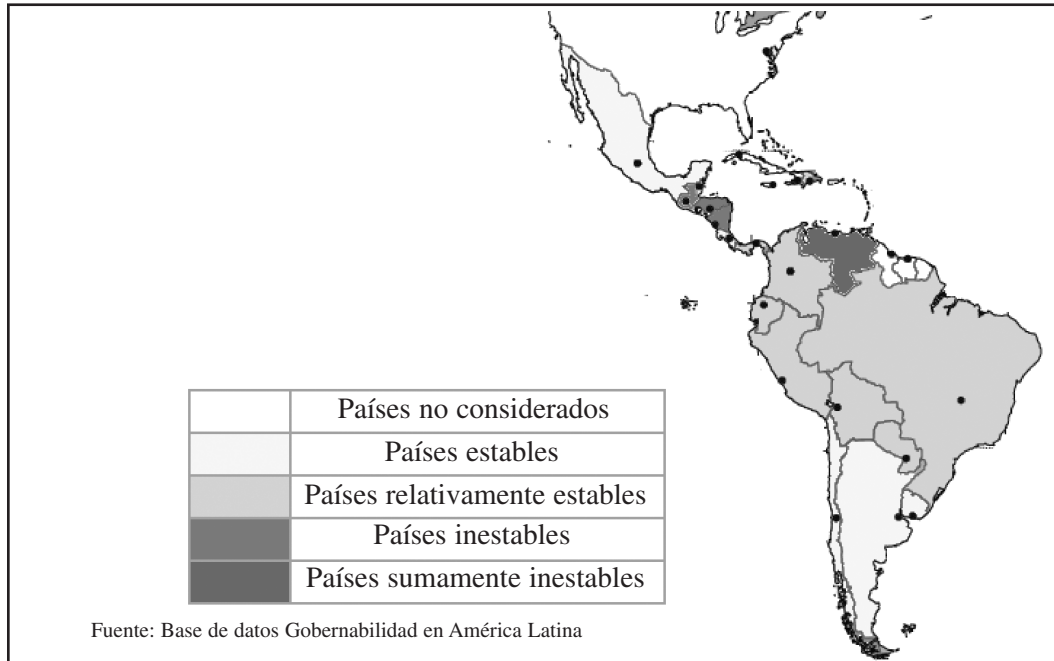


Cuadro # 3 : Membresía de los países por cluster.

| Caso | 4 Clusters |
|------------------|------------|
| 1:Argentina | 1 |
| 2:Bolivia | 2 |
| 3:Brasil | 2 |
| 4:Chile | 1 |
| 5:Colombia | 2 |
| 6:Costa Rica | 1 |
| 7:Ecuador | 2 |
| 8:El Salvador | 2 |
| 9:Guatemala | 3 |
| 10:Honduras | 3 |
| 11:México | 1 |
| 12:Nicaragua | 3 |
| 13:Panamá | 2 |
| 14:Paraguay | 2 |
| 15:Perú | 2 |
| 16:R. Dominicana | 2 |
| 17:Uruguay | 1 |
| 18:Venezuela | 4 |

Fuente Base de datos: Gobernabilidad en América Latina

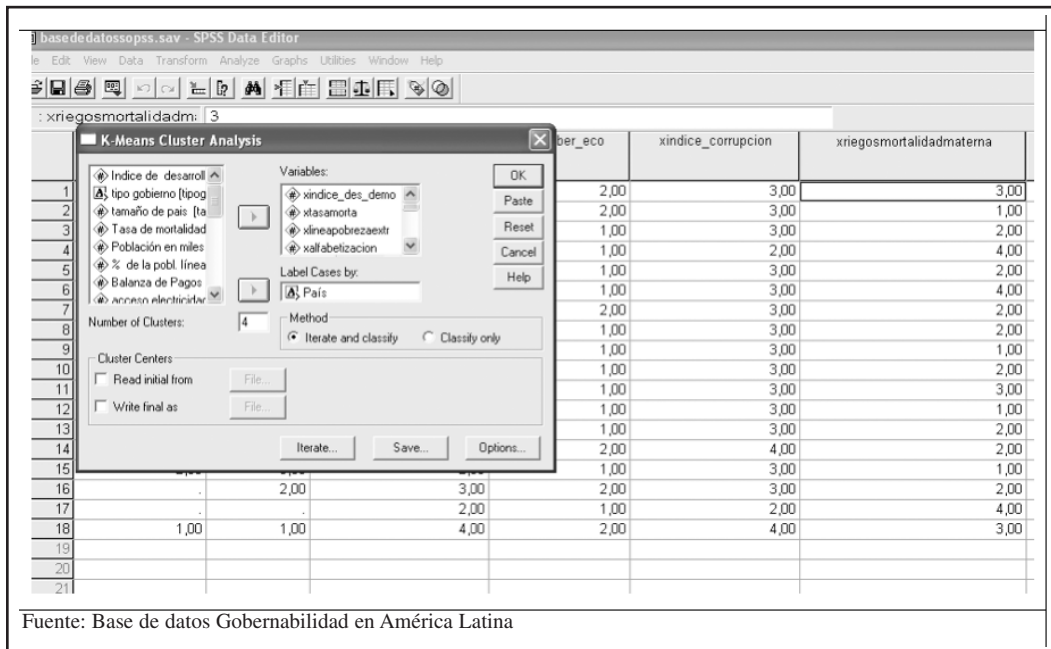
Ilustración # 6 :
Mapa de ubicación de los países con estabilidad democrática
según análisis de dendrograma y conglomerados jerárquicos.



El dendrograma se convierte en un recurso relevante dentro de los análisis de conglomerados jerárquicos, pues facilitan tanto la interpretación de los datos como su divulgación a otros públicos, es una herramienta que brinda una capacidad explicativa por sí misma, al suministrar elementos para la conceptualización de los procesos estudiados

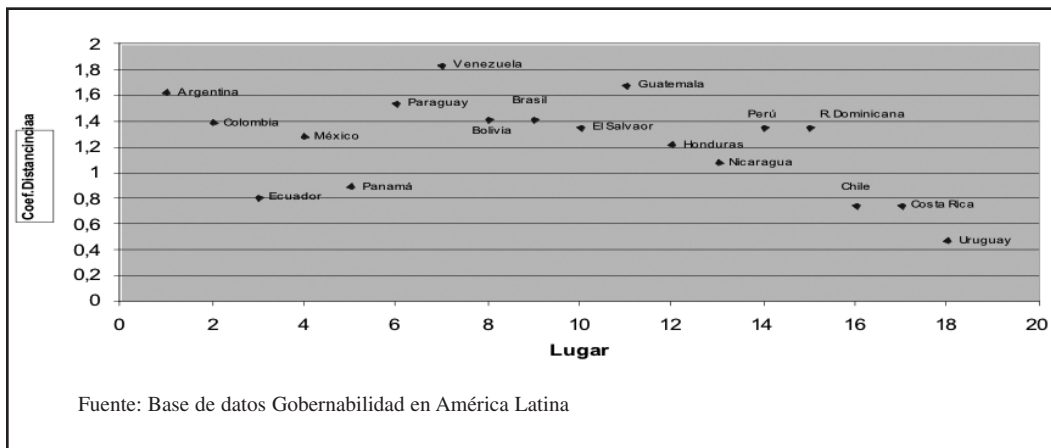
Un rápido vistazo a las conglomeraciones llevarán a cuestionarse la hipótesis lineal de Lipset la cual como se vio planteaba una relación mecánica entre bienestar económico y estabilidad democrática. En efecto, los gráficos y medidas establecidas por la estrategia de conglomerados jerárquicos y dendrogramas llamarán la atención inmediatamente de que no existe una relación directa entre desarrollo económico y democracia. Así por ejemplo, en el conglomerado de los más favorecidos (Chile, Uruguay y Costa Rica) se destacan varios factores. Por un lado, están por lejos de ser los países más ricos de América Latina. Segundo, poseen modelos económicos antagónicos. Así Chile, se rige por un voraz sistema neoliberal que genera grandes desigualdades sociales; mientras Costa Rica sobrevivió (hasta el 2007 por lo menos) con un sistema mixto de cohesión social y libre empresa. Entre tanto, en el conglomerado de menor estabilidad Venezuela se rebela con un caso también atípico de poca estabilidad a pesar de su gran riqueza. Tercero, si bien parece darse una correspondencia entre los países pobres como Nicaragua, Honduras y Guatemala con su inestabilidad democrática, también es cierto que se omite el factor externo. En efecto, el agrupamiento evidencia la

Ilustración # 7 : Análisis de medidas cluster.



Fuente: Base de datos Gobernabilidad en América Latina

Gráfico # 1 : Ubicación de los países según su grado de estabilidad política tomando en consideración las variables de gobernabilidad



Fuente: Base de datos Gobernabilidad en América Latina

necesidad entonces de retomar las variables externas como han sido los factores geopolíticos. Las intervenciones extranjeras, en especial, el rol desestabilizador de los Estados Unidos en esos países ha sido notorio. Por un lado, los gobiernos estadounidenses han sido reacios (en menor o mayor medida) a los gobiernos de corte social. Asimismo, el departamento de estado estadounidense no ha ocultado nunca su propensión a fomentar y apoyar las dictaduras en especial en Centroamérica y el Caribe, lo cual se convierten en otro parámetro a considerar al abordar el problema de los regímenes autoritarios y militares de América Latina. Finalmente, el dendrograma evidencia la necesidad de buscar otras variables internas y los procesos temporales para discernir tanto cuáles han sido los periodos que se ha logrado mayor estabilidad como el papel de los actores en su consecución.

Por último, el ejemplo incompleto que hemos abordado, abre las puertas a discusiones comparadas más profundas que amplían, rebaten y construyen propuestas de análisis de corte más interpretativo, además de revelar posibles vínculos antes no observados., los cuales rebasan los propósitos de este trabajo.

3.4. De la sistematización al análisis y del análisis a la interpretación

Los análisis de conglomerados jerárquicos y los dendrogramas individuales si bien por si solos dan lugar a relaciones e interconexiones están se hacen con un carácter sincrónico. Dado que el análisis histórico lo que le interesa es analizar el comportamiento temporal de las sociedades, buscar las transformaciones en los estados socio históricos y cambios sociales de un territorio y tiempo determinado es necesario que los dendrogramas se usen de forma similar a los censos, es decir, comparando, prospectando e infiriendo los cómo, los porqué y principalmente quiénes fueron los que posibilitaron las alteraciones entre los lapsos estudiados. Los dendrogramas deben ser el complemento empírico a las interpretaciones sobre la experiencia en sociedad a través del tiempo.

Los agrupamientos jerárquicos logrados con los cluster y los dendrogramas permiten desarrollar dos acciones simultáneas por un lado reúnen los datos heterogéneos de acuerdo a su similitudes y por otro lado, clasifican esas similitudes en asociaciones que se basan en las mismas características utilizadas. De esta forma, permite darle cuerpo a los conceptos en un diálogo dialéctico entre teoría y realidad empírica. Los conceptos vistos desde esta perspectiva permiten convertirse en herramientas de análisis, o mejor aún, en ejes problemáticos, que permiten a su vez comparar fenómenos aparentemente disímiles.

Si bien es cierto, tanto los agrupamientos jerárquicos logrados con los cluster y los dendrogramas son estrategias mediadores entre la interpretación y la explicación no por ello dejan de ser valiosas para el historiador y el científico social. Ya Kant había avanzado con el concepto de “juicio reflexionante” en su libro *Crítica de la razón pura*, en la cual lo diferencia del “juicio determinante”. Tal vez la valoración más importante es que los juicios reflexivos o reflexionantes permiten al investigador desarrollar una capacidad de juzgar los datos, si bien esta marcado por lo subjetivo, tal atributo permite entablar discusiones más universales. Por otro lado, si bien el juicio reflexivo se asoció inicialmente a la estética y a los valores, debe reconocerse que las herramientas metodológi-

cas que lo promueven parten que la realidad es cambiante, ínter subjetiva y que los cuerpos teóricos o las leyes son afectados por paradigmas, proyectos o enfoques de los mismos investigadores.

Para nosotros los agrupamientos jerárquicos logrados con los cluster y los dendrogramas son estrategias que permiten elaborar en el investigador esquemas de entendimiento y sistematización de datos que permiten enriquecer las preguntas de análisis y el diálogo entre teoría y realidad y con ello generar los juicios de tipo reflexivos. Por ello se puede indicar que esta estrategia ofrece diversas posibilidades para la comprensión abierta tanto del fenómeno social mismo como de sus relaciones internas y externas.

Además de ello visibiliza en forma más fácil los pasos seguidos por un investigador al argumentar los hallazgos encontrados durante la investigación así como de las formas o modos en que procedió este para desarrollar su discurso, aspecto que en América Central fue dejado de lado con los denominados métodos culturalistas o posmodernos, que privilegiaban la narrativa con pocos datos o en su defecto con la extrapolación con otros contextos, antes que el desarrollo de una argumentación que partiera de los mismos datos investigados. En este sentido, la estrategia basada en los agrupamientos jerárquicos logrados con los cluster y los dendrogramas es una vía más entre las existentes para intentar procesos interpretativos con los datos analizados, por ello consideramos que es un interesante esfuerzo para dar cuenta de las categorías analíticas usadas frente a la dinámica social con ello da la posibilidad de comprender tanto la pluralidad de los fenómenos estudiados como la contingencia presente en ellos.

El entendimiento logrado con los agrupamientos jerárquicos logrados con los cluster y los dendrogramas facilitan la unidad sintética de la diversidad estudiada, lo que en los análisis comparativos de tipo regional son fundamentales.

4. EPÍLOGO

Los estudios de lo regional se abren puerta como una necesidad para estudiar de y desde las regiones los problemas socio-económicos, ambientales y culturales. No obstante, para ello no podemos analizando espacios sin comprender sus relaciones, interconexiones y experiencias comunes y divergentes. La historia local y regional tradicional no pueden dar cuenta de los cambios y la movilidad de los espacios y los territorios, los cuales están en constante configuración, pluralizándose cada vez más conforme se entrelazan procesos, ideas, personas y objetos. En la actualidad, las relaciones de local con lo translocal, lo transnacional y lo transregional son cada vez más fuertes, unas no se pueden pensar sin las otras.

Los investigadores de lo regional debemos hacer uso de las diferentes herramientas disponibles para enriquecer nuestras problemáticas, pero principalmente para hacerlas más interpretativas. Los procesos de exclusión social que conllevan a mayores dosis de inequidad social nos exigen evidenciar, denunciar y proponer alternativas a la injusticia social, en especial cuando estas se pueden expresar tanto en términos espaciales, políticos, culturales y sociales.

En materia académica la comparación tiende a ser más plural. Los métodos de análisis de conglomerados jerárquicos y los dendrogramas pueden que tengan un nicho no como esnobismo sino para diversificar las vías de comparación hasta ahora existentes, en especial salirnos de la idea de “comparar solo lo comparable”.

Una vez logrado lo anterior, los análisis de conglomerados jerárquicos y los dendrogramas podrán dar paso a la creación de conceptos, enfoques y teorías, que ojalá nos conviertan a los investigadores latinoamericanos en más que simples consumidores de teoría, para comenzar a interpretar desde nuestra propia realidad, operacionalizando y discutiendo con los hallazgos de otras historiografías.

Si bien el investigador puede decepcionarse por los propósitos que poseen los agrupamientos jerárquicos logrados con los cluster y los dendrogramas debemos enfatizar que su fin último debe relacionarse en su capacidad de colaborar en la formulación de los juicios y su potencial en la reconstrucción de los procesos de mediación entre teórica y realidad estudiada, las cuales pueden usarse como un sustento básico a los que todo investigador aspira, es decir un “juicio objetivo”, que sea avalado en las comunidades intelectuales respectivas como válido, pero principalmente provocador de nuevas problemáticas e investigaciones.

BIBLIOGRAFÍA

- Aguilera Martín, Antonio (1998) “Análisis multivariable: para la caracterización cerámica una nueva vía para la caracterización de la cerámica”, *Pyrenae*, nº. 29, pp. 117-134 (versión electrónica en <http://ceipac.gh.ub.es/biblio/Data/A/0187.pdf>, consultada el 12/12/2007).
- Barreiro, J.; López, M.; Losada, F., y Ruza, E. (2002) “Análisis de las dimensiones cognoscitiva y afectiva del Comportamiento Ecológico del Consumidor”, *Revista Galega de Economía*, vol 11, nº. 2, pp. 1-21.
- Carvalho, Maria Cleide Ribeiro Dantas de; Accioly Júnior, Horácio, y Raffin, Fernanda Nervo (2006) “Representações sociais do medicamento genérico por usuários”, *Revista Brasileira de Ciências Farmacêuticas*, vol. 42, nº. 4, pp. (versión electrónica en <http://www.scielo.br/scielo.php?pid=S1516-93322006000400012&scrip>, consultada el 10/12/2007).
- Cea, María de los Ángeles (2001) *Metodología Cuantitativa. Estrategias y técnicas de investigación social*, Madrid, Síntesis (3ª. reimpr.).
- Chinelatto Neto, Armando, y Castro, Gilmar Pinheiro Cunha (2005) “Uso de análise estatística multivariada para tipificação de produtores de leite de Minas Gerais”, *Organizações Rurais & Industriais*, vol. 7, nº. 1, pp. 114-121.
- Cocilovo, José et al. (2001) “El proceso microevolutivo de la población nativa antigua de Arica”, *Chungará*, vol. 33, nº. 1, pp. 13-20 (versión electrónica en http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-73562001000100003&lng=es&nrm=iso, consultada el 20/02/2008).
- De Alba, Martha (2004) “Estudio de las representaciones sociales del espacio urbano: El caso de la ciudad de México”, *Papers on Social Representations Textes sur les représentations sociales*, vol. 13, pp. 1.1-1.20 (en http://www.psr.jku.at/PSR_2004/13_01Alb.pdf, consultada el 15/02/2008).
- Díaz Díez, B.; Cruces Pastor, E., y Morillas Raya, A. (1997) “Las Regiones Europeas: Una Tipología Basada en la Aplicación de Técnicas Multivariantes”, *Revista Asturiana de Economía*, nº. 10, pp. 249-265 (versión electrónica en http://www2.unia.es/nuevo_inf_academica/visualizar_file_Adjunto.asp?ID=540, consultada el 11/12/2007).
- Dos Santos Aléssio, Renata Lira (2004) “A representação social da violência na literatura de cordel sobre cangaço. Social representations of violence in “String Literature” about cangaço”, *Psicologia: ciência e profissão*, vol. 24, nº. 4 (en <http://scielo.bvs-psi.org.br/scielo.php?pid=S1414-98932004000400007&>, consultada el 17/02/2008).
- Filgueira, Fernando, y Rossel, Cecilia (2005) “Desigualdad, pobreza y exclusión: impotencia, fatiga y asedio en las democracias latinoamericanas”, en I. Crespo Martínez y A. Martínez Rodríguez (coords.): *Política y Gobierno en América Latina*, Valencia, Tirant lo Blanch, pp. 351-395.

- García, Juan Carlos; Castro, Fernando, y Cárdenas, Heiber (2005) “Relación entre la distribución de anuros y variables del Hábitat en el sector La Romelia del Parque Nacional Natural Munchique (Cauca, Colombia)”, *Caldasia*, vol. 27, n.º.2 (versión electrónica en http://weblogs.madrimasd.org/_universo/archive/2007/03/04/60365.aspx, consultada el 12/01/2008).
- Huaylinos Villalva, Walter; Quispitúpac Quispitúpac, Eliana, y Martínez Luján, Norberto (2003) “Variabilidad fisicoquímica y fisiográfica del ecosistema de Manglar San Pedro-Vice (Piura-Perú)”, *Revista del Instituto de Investigación de la Facultad de Geología, Minas, Metalurgia y Ciencias Geográficas*, vol. 6, n.º. 11, pp. 7-19 (versión electrónica en http://sisbib.unmsm.edu.pe/bibvirtual/Publicaciones/geologia/v_06_n11/varia_fisico.htm, consultada el 06/02/2008).
- Kageyama, Angela, y Troncoso Leone, Eugenia (1999) *Uma tipologia dos municípios paulistas com base em indicadores sociodemográficos*, texto para Discussão IE/ UNICAMP, Campinas, n.º. 66.
- López Caballero, Emilio, y Pérez Suárez, Gonzalo (1999) “Métodos de análisis en la reconstrucción filogenético”, *Bol. S.E.A.*, n.º. 26, pp. 45-56 (versión electrónica en <http://entomologia.rediris.es/sea/bol/vol26/s1/articulo/index.htm>, consultada el 12/01/2008).
- Neto, José Machado Moita, y Moita, Graziella Ciaramella (1998) “Uma introdução à análise exploratória de dados multivariados”, *Química Nova*, vol. 21, n.º. 4, pp.467-469 (versión electrónica en <http://www.scielo.br/scielo.php?pid=S0100-40421998000400016&scrip>, consultada el 03/02/2008).
- Pérez, César (2001) *Técnicas estadísticas con SPSS*, Madrid, Prentice Hall.
- Rothhammer, Francisco et al. (2003) “Análisis de ADNmt de restos esqueléticos del sitio arqueológico de Tiwanaku y su relación con el origen de sus constructores”, *Chungara, Revista de Antropología Chilena*, vol. 35, n.º. 2, pp. 269-274 (versión electrónica en http://aridaterra.uta.cl/scielo.php?script=sci_arttext&p, consultada el 12/01/2008).
- Samper, Mario (comp.) (2001) *Metodologías convergentes e historia social del cambio tecnológico en la agricultura*, San José, Progreso Editorial.
- Sartori, Giovanni (1999) “Comparación y método comparativo” en G. Sartori y L. Morlino (eds): *La comparación en las ciencias sociales*, Madrid, Alianza Universidad.
- Vilanova, Isabel; Prieto, Aldo, y Stutz, Silvina (2006) “Historia de la vegetación en relación con la evolución geomorfológica de las llanuras costeras del este de la provincia de Buenos Aires durante el Holoceno”, *Ameghiniana*, vol. 43, n.º.1, pp. 147-159 (versión electrónica en <http://www.scielo.org.ar/scielo.php?pid=S0002-7014200600100012&sc>, consultada el 03/02/2008).