

Mejorando la Muestra Transcultural Estándar:
Un enfoque metodológico para crear una
encuesta dirigida mediante la selección de
variables con análisis de correspondencias
múltiples

Rodrigo Manuel Martínez Olivera

2024-06-25

Índice

1	Introducción	2
2	Estado de la cuestión	4
3	Metodología	17
3.1	Preparación de los datos	17
3.2	Aplicación de MIRIMFA con la base de datos «Poison» . . .	24
3.3	Aplicación de MIRIMFA a la SCCS	27
3.4	Análisis de Factores Múltiples	38
3.5	Selección de variables	41
3.6	Elaboración del cuestionario	43

4 Reducción de variables culturales mediante análisis de componentes principales	46
4.1 Análisis de Correspondencias Múltiples (ACM)	58
4.2 Selección de variables	58
4.3 Desarrollo del cuestionario de la encuesta	58
4.4 Aislamiento de temas y refinamiento de preguntas	59
4.5 Incorporación de temas antropológicos contemporáneos .	59
4.6 Administración de la encuesta	59
4.7 Recopilación y análisis de datos	59
4.8 Consideraciones éticas	60
5 Bibliografía	60

Resumen

La Muestra Estándar Intercultural («Standard Cross-Cultural Sample», SCCS), compilada por Murdock y White, sirve como base de datos de información etnográfica de diversas sociedades, lo que facilita los estudios interculturales y la comprensión de la diversidad humana. Sin embargo, la naturaleza estática de la SCCS y los retos que plantean los datos faltantes limitan su utilidad para la investigación moderna. Esta tesina propone un enfoque metodológico para mejorar la SCCS mediante la creación de una encuesta dirigida basada en el conjunto de datos original. La metodología consiste en aplicar el Análisis de Correspondencias Múltiples (ACM) para identificar las variables más significativas y elaborar un cuestionario reducido para los investigadores. Mediante la revisión manual y el aislamiento de temas, las variables se seleccionan en función de su contribución a las dimensiones del ACM. La encuesta resultante abarca una amplia gama de aspectos culturales, incluidos temas tradicionales de la SCCS y temas de reciente in-

terés como los géneros no binarios y las dinámicas lingüísticas. La metodología propuesta pretende superar las limitaciones de la SCCS original, proporcionando una herramienta práctica para que los investigadores recojan datos etnográficos actualizados y actualicen la SCCS para análisis longitudinales y estudios comparativos. Esta tesina contribuye al avance de la metodología en los estudios transculturales y ofrece un valioso marco para comprender el cambio cultural y la diversidad a lo largo del tiempo.

1 Introducción

Los estudios transculturales desempeñan un papel crucial en la comprensión de la diversidad de las sociedades humanas y las dinámicas culturales de las distintas comunidades. Para ello, es fundamental disponer de conjuntos de datos exhaustivos que ofrezcan información sobre diversos aspectos culturales y faciliten el análisis comparativo. La Muestra Estándar Intercultural (SCCS), compilada por Peter Murdock y Douglas White (1969), es uno de esos conjuntos de datos que constituye un repositorio histórico de información etnográfica de diversas sociedades de todo el mundo.

La SCCS está formada por datos de 186 sociedades bien documentadas y proporciona una valiosa información sobre una amplia gama de ámbitos culturales, como la subsistencia, la economía, la política, los rituales y las estructuras familiares. Sin embargo, a pesar de su importancia en la investigación transcultural, la SCCS tiene algunas limitaciones. Su naturaleza estática, debida a la falta de actualizaciones desde su compilación, plantea retos para los investigadores interesados en el análisis longitudinal y la comprensión del cambio cultural a lo largo del tiempo. Además, la

dependencia del conjunto de datos de los relatos etnográficos introduce sesgos y limitaciones, lo que da lugar a información incompleta y a la frecuente ausencia de datos en la mayoría de las variables.

Para abordar estas limitaciones y liberar todo el potencial de la SCCS para la investigación transcultural contemporánea, esta tesina propone un enfoque metodológico centrado en la creación de una encuesta dirigida. Este enfoque se centra en aprovechar el conjunto de datos existente para desarrollar un instrumento de encuesta adaptado para captar información etnográfica actualizada de las sociedades representadas.

La metodología que aquí se propone se basa en los principios del Análisis de Correspondencias Múltiples (ACM), una técnica de reducción de la dimensionalidad adecuada para procesar variables categóricas. Mediante la aplicación del ACM al conjunto de datos original de la SCCS, esta tesina pretende identificar las variables culturales más significativas y priorizarlas para su inclusión en la encuesta dirigida. Mediante la revisión manual y el aislamiento de temas, se seleccionarán las variables en función de su relevancia para el discurso antropológico contemporáneo y su potencial para captar la diversidad cultural en las distintas sociedades.

El cuestionario de la encuesta resultante abarcará una amplia gama de aspectos culturales, incluidos los temas tradicionales tratados en la SCCS y los temas emergentes que reflejan los intereses de la antropología moderna. Al proporcionar a los investigadores una herramienta normalizada para la recogida de datos etnográficos actualizados, esta encuesta dirigida pretende actualizar la SCCS, permitiendo el análisis longitudinal y los estudios comparativos del cambio cultural a lo largo del tiempo.

En las secciones siguientes de esta tesina, profundizaré en la metodología

propuesta, analizando el proceso de creación de la encuesta dirigida, los criterios de selección de variables y las implicaciones del diseño de la encuesta para la investigación transcultural. Con ello pretendo avanzar en la metodología de los estudios transculturales y contribuir a una comprensión más completa de la diversidad humana y la dinámica cultural.

2 Estado de la cuestión

Imagine if Earth scientists stopped using aerial photos in the '60s due to them not showing details, flattening reality. Now, they prefer to only use ground-level photography and tangible samples. We would say that they have lost perspective. This analogy mirrors the shift in anthropological methods from large-scale monographs and databases in the '60s-'80s to more focused, urban- or nation-centric studies. While this shift had its reasons, it also risked overlooking the broader spectrum of human diversity. In response, this survey aims to bridge the gap between historical methods and modern sensitivities. There is some antecedent with the Human Relations Area Files, but by their admission, there is a “typical divide between those who cross-culturally study smaller-scale nonindustrial societies, and those who use countries as their units of analysis. Data for these two types of comparisons generally differ—ethnographic data based on participatory methods provides most of the information for nonindustrial societies; cross-national or cross-country comparisons tend to employ surveys with targeted questions on a few domains of life.” (Ember & Kalodner, 2024) This survey is a response to that claim, and so I am revisiting the Standard Cross-Cultural Sample (SCCS) from 1969, transforming it into direct questions and updating it with contemporary insights and tools. By tracking the same societies or their closest living

relatives, we aim to understand the evolution of human diversity in the 21st century. If you are working with one of these societies, your input could illuminate crucial aspects of contemporary human life. You will encounter a series of questions selected from the vast SCCS dataset, carefully curated to capture the essence of each community. Rest assured, you will not be inundated with 1780 questions; I have streamlined it to around 120, focusing on representative variables. Additionally, I have incorporated new questions to address modern research interests and cultural sensitivities, ensuring a comprehensive perspective. Your participation in this analysis is crucial, even if some questions might seem unfamiliar. Please do not skip any questions, as missing data could compromise the quality of this study. It's understandable if you don't know the answers to certain questions, as it is rare nowadays to cover so many topics. However, you should try to find those answers through a little bit of fieldwork or an interview with a local, as the missing data should not reflect your personal interests. If you need to take a break, you can save your progress and come back later. You can also use the print version of the survey to get ready in advance. If you encounter cultural barriers that prevent data collection, please use the "Other" option and provide a brief explanation. Please use "Other" whenever you see that the available categories are not relevant to your community. If that option is not available (due to Google Forms limitations), please keep notes and add them to the final question, where you will also have some space to express any commentary you deem fit. Upon completion, your responses will contribute to the study of human diversity in the 21st century. Your insights may even lead to co-authorship in an eventual publication. Together, we're striving to create a more inclusive and nuanced understanding of human cultures,

transcending traditional boundaries.

Thank you for your invaluable contribution to this endeavour.

A finales del siglo XIX, estudiosos como Sir James Frazer y E. B. Tylor intentaron responder a preguntas sobre la evolución de los humanos. Para ello, compararon descripciones de la vida tribal proporcionadas por miembros de tribus occidentalizadas y misioneros. Este enfoque, conocido como «antropología de sillón», fue criticado por ser demasiado ingenuo en su lectura de las fuentes y por ignorar los conflictos de intereses y la ignorancia de los autores. Este enfoque proyectaba prejuicios en el peor de los casos y aplanaba la realidad en el mejor. La tendencia a proyectar prejuicios quedó patente en la teoría del unilinealismo, que sugiere que algunas sociedades se encuentran en diferentes fases de progreso. El unilinealismo es un ejemplo clásico de la naturaleza altamente especulativa de esta escuela de pensamiento.

A principios del siglo XX, Bronisław Malinowski propuso que los antropólogos realizaran ellos mismos el trabajo de campo en lugar de basarse en observaciones de sillón. Subrayó la importancia de estudiar culturas extranjeras para evitar los prejuicios de la propia. Este enfoque de la investigación se conoce como etnografía.

La primera tradición etnográfica produjo documentos de alcance similar a los utilizados por los historiadores de sillón. Por ello, los sucesores de esta tradición no vieron ningún problema en incorporar estos documentos a sus análisis. El hecho de que estos documentos fueran producidos por expertos en antropología significaba que la calidad de los datos era mejor. Esto supuso una suspensión de la crítica original sobre la calidad de las fuentes. Mientras tanto, el análisis estadístico fue ganando popularidad

lentamente a principios del siglo XX. Prometía eliminar la especulación, reconocida como problemática por los primeros etnógrafos. Con este espíritu, George Peter Murdock propuso la Muestra Etnográfica Mundial y, junto con Douglas R. White, la Muestra Transcultural Estándar (SCCS) en 1969. La SCCS consiste en un conjunto cuidadosamente seleccionado de 186 culturas bien documentadas que ahora están codificadas para 1780 variables. En lenguaje cotidiano, la SCCS son dos cosas: la propia muestra de sociedades y la base de datos asociada a ellas.

La muestra se creó para abordar el problema de la autocorrelación filogenética, que se produce cuando las pruebas de las relaciones funcionales se ven confundidas por culturas que no son independientes entre sí. Esta falta de independencia puede atribuirse a que las sociedades toman prestados rasgos de culturas vecinas (influencia areal) o tienen una ascendencia común. Sin controlar estos problemas, es imposible atribuir las similitudes o diferencias a otras causas y establecer la independencia estadística. La independencia es un principio fundamental de los análisis estadísticos que incluyen mediciones de la varianza y pruebas de significación. Para resolver este problema, se pueden utilizar ajustes estadísticos que, en el mejor de los casos, pueden hacer que las pruebas de autocorrelación sean negativas. Otra posibilidad es eliminar la autocorrelación directamente de la muestra evitando la repetición de sociedades que se influyen mutuamente. Este es el enfoque que han elegido Murdock y White (Mace et al. 1994).

El proceso de muestreo de la SCCS puede considerarse una muestra estratificada de conveniencia. Murdock y White utilizaron su conocimiento de las distintas regiones para identificar 200 estratos a partir de las sociedades de las que disponían de información en su Atlas Etnográfico (George

Peter Murdock 1967). Tras seleccionar los 200 estratos, se introdujo en la muestra la sociedad mejor documentada dentro de cada estrato. A continuación, se estudiaron las sociedades de interés utilizando etnografías, diarios y otros materiales escritos por lugareños y misioneros. Estos materiales se etiquetaron línea por línea y párrafo por párrafo en función del tema. Se guardó una copia de referencia completa de la obra original, mientras que una segunda copia se cortó en sus líneas constituyentes. A continuación, se agruparon todas las líneas relacionadas con un tema concreto para codificarlas como variables en la base de datos de la SCCS.

Tras la creación de la SCCS, lo que pasó a llamarse Antropología Comparada experimentó un crecimiento y varias revistas publicaron investigaciones en este campo. Los académicos fueron motivados a seguir estudiando las sociedades cubiertas por la SCCS para ampliar y enriquecer la base de datos con información más detallada. La SCCS fue ampliamente adoptada por prestigiosas universidades, y era habitual que los departamentos de antropología albergaran una biblioteca especialmente dedicada a la SCCS. En consecuencia, la base de datos creció hasta alcanzar el enorme tamaño de 1.780 variables de que disponemos hoy en día.

La respuesta de los etnólogos, sobre todo los de las escuelas Social y Cultural, no fue positiva ante la idea de que su trabajo fuera diseccionado para un análisis cuantitativo. Malinowski y sus seguidores criticaron ese enfoque, ya que existía la creencia de que cualquier estudio comparativo aplanaría los diferentes componentes de una sociedad. Además, los elementos de una cultura sólo tienen sentido como parte de un todo. La antropología, según Malinowski y más tarde Clifford Geertz, debía convertirse en un intento de empatizar con una cultura, de hacerse local, registrar e interpretar. Se consideraba científica sólo en la medida

en que se basaba en la observación empírica, pero por lo demás, era hermenéutica: la rama de la filosofía que se ocupa del significado.

Estudiar a los seres humanos puede ser todo un reto porque presentan un alto grado de variabilidad a la escala que nos interesa y sus respuestas a los estímulos también pueden variar mucho. Además, los humanos son seres complejos, e intentar modelar esta complejidad es una tarea difícil, sobre todo cuando la entidad que realiza la modelación es otro humano. Además, los investigadores suelen imponer a sus modelos expectativas que nunca se pedirían al modelado de objetos naturales. Las consideraciones éticas que conlleva la investigación con seres humanos también pueden complicar las cosas (Andreski 1972). En consecuencia, las ciencias sociales han dependido históricamente más de los datos cualitativos que otros campos, lo que las hace más susceptibles a las influencias ideológicas (Dutton s. f.).

Esta vulnerabilidad se puso de manifiesto en las décadas siguientes, cuando, en el ámbito de los círculos académicos literarios, fue creciendo lentamente un proyecto político denominado Estudios Críticos. En él se señalaban los problemas de parcialidad y exceso de confianza en la objetividad y la ciencia que se utilizaron para justificar políticas perjudiciales en el siglo pasado. Estas críticas eran más punzantes en el caso de las ciencias sociales, sobre todo las que se basaban en gran medida en métodos cualitativos (Latour 2004; Dutton s. f.). Esto dio lugar a una división más fuerte entre la Antropología Comparada y la Antropología Etnográfica. Muchos antropólogos etnográficos empezaron a cuestionar su capacidad de ser científicos y abrazaron su subjetividad, viéndose a sí mismos como escritores literarios y activistas políticos. A medida que aumentaba su activismo político, también lo hacía su antagonismo hacia el Estado y

otras instituciones occidentales. Como eran la cara de la Antropología, muchos financiadores empezaron a retirarles su apoyo, lo que también afectó a la Antropología Comparada.

Otro factor que contribuye a la pérdida financiera es el cambio político y demográfico de las sociedades rurales a las urbanas. La urbanización ha hecho que las ciudades tengan más poder político y económico que en épocas anteriores. En consecuencia, la investigación sobre las ciudades ha recibido una mayor parte de los recursos financieros, mientras que el interés por las sociedades no industriales ha disminuido.

Del mismo modo, las etnografías cambiaron su enfoque hacia los entornos urbanos y se volvieron más especializadas. En lugar de registrar información cultural general, empezaron a concentrarse en aspectos no tangibles de la sociedad y dejaron de informar sobre algunos aspectos clave de las sociedades registradas. En consecuencia, la mayoría de las etnografías se volvieron inútiles para los antropólogos comparativos, y la proporción de datos faltantes en sus registros aumentó considerablemente.

La antropología comparada fue más resistente a los efectos del deconstruccionismo, ya que estaba muy formalizada. Sin embargo, siguió sufriendo numerosos reveses. Los financiadores la equipararon a la etnografía, lo que le causó daños de reputación, y perdió el acceso a fuentes de información cruciales, ya que ahora las etnografías no trataban sobre la mayoría de las sociedades y eran demasiado específicas.

Así, en los años 80 la SCCS tuvo que dejar de actualizarse por falta de fondos. Esto supuso un reto para los investigadores, que tuvieron que decidir si seguir utilizando la SCCS, recurrir a otras bases de datos o

recopilar sus propios datos para cada pregunta de investigación. Aunque en principio seguir utilizando la SCCS es un buen planteamiento, tiene sus limitaciones. Es como basarse en fotos aéreas de los años 80 para comprender el estado actual del planeta Tierra. Aun así, es un buen recurso, por lo que es una suerte que en la era de Internet se pueda acceder a la SCCS a través de la Base de Datos de Lugares, Lengua, Cultura y Medio Ambiente (D-Place) (Kirby et al. 2016).

El segundo enfoque se enfrenta a retos debido a las diferencias en los objetivos de las bases de datos utilizadas por los antropólogos comparativos. Las bases de datos más utilizadas en este campo son la Encuesta Mundial de Valores (WVS), el Estudio Comparativo de Sistemas Electorales (CSES), las encuestas del Centro de Investigación Pew, la Encuesta Mundial Gallup, los Indicadores de Desarrollo Mundial del Banco Mundial, el Índice de Desarrollo Humano (IDH) del Programa de las Naciones Unidas para el Desarrollo y el Índice para una Vida Mejor de la OCDE. Estas bases de datos se derivan de diversas disciplinas como la economía, la demografía, la geografía y otras, que comparan principalmente naciones. Sin embargo, las naciones son un tipo específico de sociedad que posee suficiente poder industrial, económico y político para reclamar un territorio que normalmente también contiene muchas otras sociedades, y se niega a reconocerlas, ya sea a través de la persecución o redefiniéndolas como subordinadas en alguna capacidad (como reclamarlas como manifestaciones de sí misma o a través de la subyugación política). Por lo tanto, la comparación entre naciones muestra un sesgo a favor de las sociedades industriales y predominantemente urbanas.

El último método de recogida de datos, que consiste en recopilar éstos de forma independiente, probablemente se vea restringido por los recursos

de que disponga cada investigador. Por lo tanto, debe ser específico y no exploratorio. Sólo se pueden registrar unas pocas variables antes de que resulte demasiado costoso viajar por todo el mundo.

Como diría cualquier antropólogo etnógrafo, lo que he descrito es una generalización, y todas las generalizaciones son erróneas. En este caso, hay dos puntos importantes que no he compartido hasta ahora. En primer lugar, el fin de la SCCS no significa que dejen de existir bases de datos de antropología. Sigue habiendo muchas bases de datos de antropología actualizadas. En segundo lugar, los etnógrafos no están totalmente ausentes de las sociedades no industriales ni están completamente desvinculados del compromiso con la verdad.

Existe un gran proyecto comparativo en curso llamado Los Archivos del Área de Relaciones Humanas («Human Relations Area Files», HRAF), fundado por Murdock. HRAF es una organización sin ánimo de lucro que mantiene tres importantes bases de datos: eHRAF World Cultures, eHRAF Archaeology y HRAF Explaining Human Culture. eHRAF World Cultures es la base de datos más cercana a la SCCS y contiene sus datos originales. Sin embargo, se trata más bien de un corpus de etnografías completas, con variables precodificadas escasas, y no es adecuada para el análisis cuantitativo, salvo en el caso de algunas muestras. Además, comparte el mismo problema de los valores perdidos al que se enfrentan la SCCS y otras bases de datos que utilizan el registro etnográfico. eHRAF Archaeology tiene problemas similares, mientras que HRAF Explaining Human Culture es un corpus de investigaciones anteriores que tiene el potencial de permitir la síntesis de variables u observaciones utilizadas en diferentes artículos, siempre que coincidan en sus muestras.

Como vemos, existe una importante laguna que aún no se ha llenado. El HRAF afirma que «un problema metodológico [...] es la típica división entre quienes estudian transculturalmente sociedades no industriales de menor escala y quienes utilizan países como unidades de análisis. Los datos para estos dos tipos de comparaciones suelen diferir: los datos etnográficos basados en métodos participativos proporcionan la mayor parte de la información para las sociedades no industriales; las comparaciones transnacionales o entre países tienden a emplear encuestas con preguntas específicas sobre unos pocos ámbitos de la vida.¹ Reunir los resultados de estos dos tipos de investigación que abarcan la gran variedad de culturas de la Tierra es todo un reto» (Ember y Kalodner 2024). Este trabajo es una respuesta a esta afirmación, ya que pretende reunir los hallazgos de estos dos tipos de investigación, abarcando el abanico mundial de culturas.

En el mundo actual, donde la comunicación es más accesible que nunca, podemos salvar la distancia entre analistas y etnógrafos realizando encuestas dirigidas para que los etnógrafos las rellenen. Éstos están familiarizados con el vocabulario de la antropología y tienen la experiencia necesaria para recuperar los datos que faltan. Enviándoles una encuesta dirigida, podemos obtener datos de alta calidad que antes eran inalcanzables. En el pasado, la recogida de datos dependía de que los etnógrafos aportaran sus observaciones, que luego se plasmaban en bases de datos. Sin embargo, este método era fruto de las limitaciones técnicas de la época y se basaba en cualquier dato disponible. Con las encuestas dirigi-

¹Ember y Kalodner usan la siguiente referencia en este punto, pero me fue imposible acceder a ella. Independientemente de la fuente original, el punto se mantiene: Albert, Steven M., and Maria G. Cattell. 1994. "Old Age in Global Perspective: Cross-Cultural and Cross-National Views."

das, los etnógrafos pueden ahora participar en el proceso de recogida de datos, lo que puede aliviar las preocupaciones sobre la validez ecológica de las observaciones resultantes. Además, aprovechar a la gente que ya está allí permite reducir el problema de la financiación.

Cabría hacerse dos preguntas teniendo en cuenta lo que he mencionado antes: ¿cómo pueden ayudar los etnógrafos cuando están en las ciudades y por qué querrían participar en este proyecto?

Si a un etnógrafo le preocupa el bienestar de la comunidad que investiga, las tendencias actuales son realmente preocupantes. Las sociedades no industrializadas están siendo ignoradas a un ritmo sin precedentes. Todos los campos, incluida la ciencia, les dan la espalda, y la mayoría de las descripciones modernas de la naturaleza humana siguen centrándose en las sociedades «WEIRD» («Western, Educated, Industrialised, Rich, and Democratic», occidentales, educadas, industrializadas, ricas y democráticas) (Arnett 2016; Henrich, Heine, y Norenzayan 2010). Esta tendencia afirma implícitamente que la naturaleza humana está bien representada por estos individuos tan particulares y que el proceso de urbanización se ha completado, lo que conlleva que las personas no industrializadas ya no existen. Si esto no despoja de poder a estas últimas, no sé qué puede hacerlo.

Muchas de las preocupaciones que han llevado a los etnógrafos a mostrarse escépticos ante la ciencia y a querer distanciar los estudios sociales de ella ya no son pertinentes. Como ha argumentado el teórico (post)crítico francés Bruno Latour (2004), la práctica de separar los estudios sociales de la ciencia es, en el mejor de los casos, un planteamiento anticuado que pretendía empoderar a los grupos marginados y, en el peor. En la

época posterior a la Segunda Guerra Mundial, era comprensible cuestionar la objetividad de las políticas que utilizaban la ciencia para justificar el racismo, el supremacismo, los experimentos poco éticos y las guerras. Sin embargo, en el clima político actual, el escepticismo hacia la ciencia se utiliza para obstaculizar los avances en la lucha contra el cambio climático, la aplicación de vacunas y la promoción de los derechos de los grupos marginados, incluidas las sociedades no industriales. En mi opinión, distanciar los estudios sociales de la ciencia siempre fue un error. Al fin y al cabo, si construyes castillos en el cielo, acabarán derrumbándose y cayendo sobre tu cabeza.

Una de las razones por las que este planteamiento funcionó es que es cierto que la objetividad no existe. Esto se debe a que las categorías siempre están implicadas en cualquier afirmación, y tenemos que imponerlas ya que no se encuentran en el mundo real. Nosotros decidimos dónde trazar la linde en función de su utilidad, y la utilidad viene determinada por lo que consideramos importante. Así pues, los valores y la subjetividad están siempre presentes. Pregunte a cualquier estadístico sensato y le dirá que trabaja con umbrales que no están justificados por el rigor matemático, sino por el consenso. Sin embargo, lo que los estadísticos pueden seguir diciendo es que la estadística proporciona una forma de organizar los hechos y dar a éstos un lugar en la mesa a la hora de tomar decisiones. La objetividad puede ser imaginaria, pero sigue siendo un buen ideal al que aspirar porque evita los peores efectos de la subjetividad, como la arbitrariedad, la falta de fundamento y la incapacidad de comunicar.

Sugiero que las sociedades no industriales reciban el mismo nivel de consideración que sus homólogas industriales a la hora de debatir la diversidad humana. Además, propongo que se incluya a etnógrafos en los

debates para que aporten su visión de las categorías, la aplicabilidad y los matices dentro de un entorno estructurado. ¿Simplificará este enfoque en exceso la realidad y los matices de las comunidades implicadas? Sí, lo hará. ¿Pasará por alto aspectos cruciales? Sin duda. ¿Será parcial? Sí, lo será. Sin embargo, cuando se critican las fotos aéreas por las mismas razones, no dejamos de utilizarlas. En su lugar, tomamos fotos desde ángulos diferentes o utilizamos resoluciones más altas. Incluso utilizamos lentes infrarrojos para ver más allá de las copas de los árboles, los tejados y el interior de los túneles. Si estos métodos siguen sin ser adecuados, tal vez la herramienta en sí no sea la adecuada para la tarea que tenemos entre manos. Y no pasa nada. Cada pieza forma parte del rompecabezas, y ninguna de ellas será suficiente por sí sola. Pero la belleza de las fotos aéreas es que nos ofrecen una perspectiva única.

En nuestro esfuerzo por comprender la diversidad humana, es fundamental tener la perspectiva adecuada. Aunque los etnógrafos conocen bien sus respectivos lugares, tener un marco de referencia es igualmente importante. La globalización y la modernización se han convertido en el nombre del juego. Puede que sepamos lo que se trae entre manos nuestro jugador, pero es igualmente esencial conocer el terreno, saber quién se ha modernizado y en qué momento, dónde ha mostrado resistencia y quién se ha desviado del guión tradicional. Para obtener este conocimiento, necesitamos tener una visión a vista de pájaro de la situación. ¿No es así?

Suponiendo que la respuesta sea afirmativa, los próximos capítulos tratarán de los criterios de diseño de la encuesta. Esto incluye el análisis de los datos originales de la SCCS para identificar patrones que puedan ayudarnos a resumirlos a un tamaño manejable. Al hacerlo, se representará la distribución de probabilidad conjunta, lo que a su vez identificará

aquello que caracteriza a las sociedades de la muestra.

3 Metodología

En esta sección se expone el enfoque metodológico utilizado para mejorar la Muestra Transcultural Estándar (SCCS) mediante la creación de una encuesta dirigida para el análisis longitudinal. Primero, discutiré la preparación de los datos, luego tomaré un desvío para afrontar el problema de los datos perdidos en la SCCS. En este desvío, discutiré posibles métodos de tratamiento de datos perdidos, mostraré la necesidad de uno nuevo: el Método de Imputación Múltiple mediante Análisis de Factores Múltiples Regularizado Iterativo (MIRIMFA). Luego continuaré con una simulación para comprobar sus propiedades y terminaré aplicándolo a la SCCS. Con el problema de los datos perdidos resuelto, procederé a aplicar un Análisis de Factores Múltiples para hacer selección de variables y la siguiente sección será sobre la elaboración del cuestionario mediante la traducción de estas variables a realidades prácticas.

3.1 Preparación de los datos

El primer paso consistió en preparar el conjunto de datos original de la SCCS para el análisis. Se descargaron los datos de D-Place (Kirby et al. 2016) y se organizaron para análisis.

La SCCS comprende información etnográfica de 186 sociedades diversas, que abarcan una amplia gama de aspectos culturales. Sin embargo, el conjunto de datos contiene 1781 variables, lo que plantea dificultades en cuanto a la organización y gestión de los datos.

De las variables 976 son nominales 760 son ordinales, y `rcontinuous_count`

son continuas. Dada la forma en que se recolectaron los datos, podemos decir que las variables están agrupadas por fuentes. También podríamos pensar en agruparlas por categoría, pero las categorías están razonablemente alineadas con las fuentes y dada la estructura de datos faltantes de la que hablaremos más tarde, tiene más sentido agruparlas por fuente. Así, tenemos 58 grupos.

Dada esta estructura de los datos, para la selección de variables se utilizó el análisis de factores múltiples («Multiple Factor Analysis», MFA) ya que éste está diseñado justamente para la reducción de dimensiones de bases de datos con variables mixtas agrupadas. Se hizo una selección de los factores principales y de éstos se tomaron sus contribuyentes principales.

Para que esto fuera posible, se comenzó retirando las variables categóricas que sólo tenían un valor, ya que no tenían valor contrastivo. Los códigos de éstas son: “SCCS1381, SCCS1382, SCCS1383, SCCS1385, SCCS1386, SCCS1387, SCCS1600, SCCS1606, SCCS1609, SCCS1611, SCCS1613, SCCS1622, SCCS1698, SCCS1118, SCCS1199, SCCS1200, SCCS1203, SCCS1206, SCCS1221, SCCS1222, SCCS1225, SCCS1359, SCCS1455, SCCS488, SCCS496, SCCS497, SCCS499, SCCS503 y SCCS512”.

Este problema nos señala algo más que tenemos que afrontar antes de seguir: la estructura de los datos faltantes. El análisis de éstos es vital, en especial si pretendemos que el cuestionario resultante trata de ser una respuesta a dicho problema. Por lo tanto, antes de hacer una reducción de dimensiones, tenemos que identificar en qué consisten estas ausencias.

De los 1752 [variables] * 186 [sociedades] = 325872 datos que quedan luego de retirar las variables que sólo tenían un valor, tenemos que el 45.62% son datos faltantes. Éstos se distribuyen en todos los casos, ya

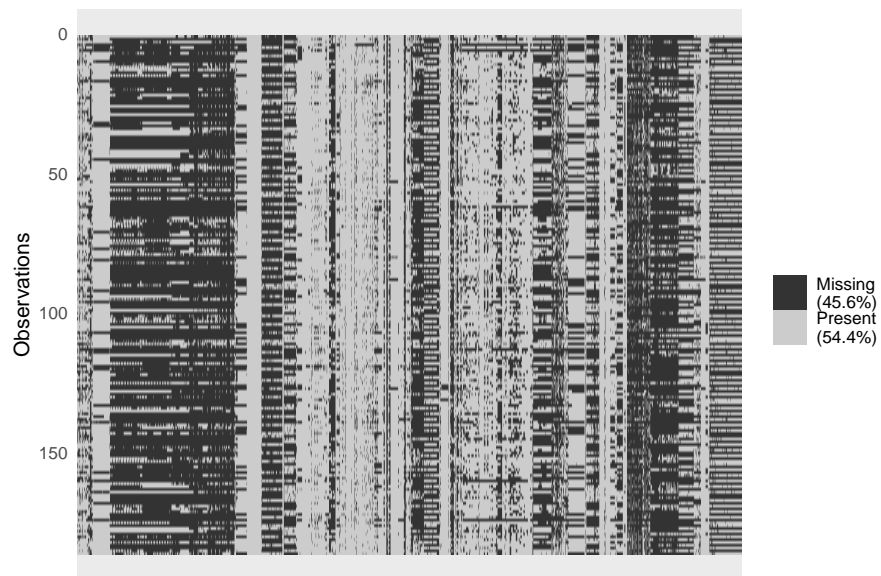


Figura 1: Patrones de pérdida de datos: Variables ordenadas en orden alfabético

que la cantidad mínima de datos faltantes por sociedad es de 18.49%, que corresponde a la sociedad SCCS147, los Comanche. Todas la demás sociedades tienen muchos más datos faltantes, como los Tobelorese, que tienen hasta un 80.31% de datos faltantes.

En cuanto a variables, tenemos que 119 no presentan datos faltantes mientras que “SCCS98, SCCS487 y SCCS500” presentan hasta 80.31% de datos faltantes.

Súmese a lo anterior la lectura de la figura 1 que muestra de izquierda a derecha las variables ordenadas alfabéticamente, y de arriba a abajo a las sociedades estudiadas y los datos faltantes para cada par variable-sociedad. Podemos ver que no pareciera haber un patrón general de datos faltantes entre sociedades. Excepto por unos cuantos hilos, la gráfica pare-

ciera ser homogénea de arriba a abajo. En cambio, de izquierda a derecha, en la dirección de las variables, vemos claros haces de variables con datos faltantes. Por lo tanto, podemos decir que el mecanismo de pérdida de los datos no es completamente al azar («Missing completely at random», MCAR). Esto, sumado a que no tenemos casos completos y la proporción de datos faltante, me obliga a concluir que el único tratamiento adecuado es la imputación múltiple.

Los datos faltantes incluyen un nivel adicional de incertidumbre a una base de datos. Si los datos faltan debido a un mecanismo completamente ajeno a ellos y que afecta la probabilidad de que no se registren de forma homogénea, decimos que se perdieron completamente al azar («Missing completely at random», MCAR) y podemos ignorarlos, ya sea mediante descartar todos los casos incompletos, o hacer los cálculos con cada uno de los datos existentes pertinentes a ese caso. Sin embargo, si el mecanismo que hizo que los datos se perdiesen responde al valor de otros datos sí observados («Missing at random», MAR), o al verdadero valor del dato que nos falta («Missing not at random», MNAR), ignorar los datos faltantes suele introducir un sesgo. Hay muchos métodos de tratar la situación, pero la más responsable epistemológicamente es la imputación múltiple.

La imputación múltiple consiste en sustituir los datos faltantes por valores posibles hasta obtener una base de datos completa que posteriormente se analiza con métodos ordinarios. A diferencia de la imputación simple, que sólo hace esto una vez y que uno podría criticar como inventarse los datos, esta operación se hace múltiples veces para obtener múltiples bases de datos y análisis. Estos análisis se comparan para estimar el error e incertidumbre adicional que nos dieron los datos faltantes y

se resumen en un análisis final global. Este procedimiento se esquematiza en la figura 2.

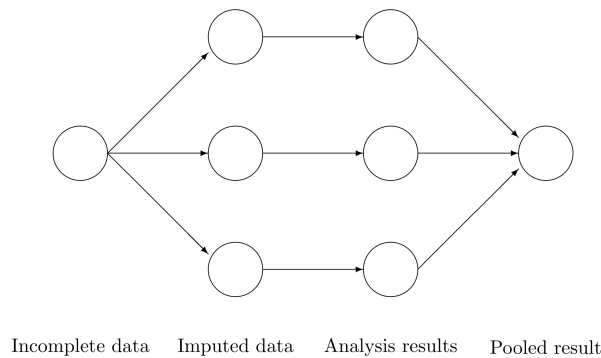


Figura 2: Diagrama de flujo de la imputación múltiple (Van Buuren, 2018). Los datos incompletos («incomplete data») se usan para crear datos imputados («imputed data»), que se usan para analizar los resultados («analysis results») y finalmente agregarlos («pooled results»).

La forma de determinar qué valores son posibles para cada dato faltante es compleja, ya que depende del mecanismo de pérdida de los datos, lo que se pueda deducir de la distribución original a través de los datos observados, la incertidumbre inferida por expertise de área, el tipo de variables, el poder de cómputo disponible, el rigor epistemológico que se pretenda, entre otros. Dado que muchas veces no es posible afirmar muchas de estas propiedades, se sugiere el uso de métodos robustos a la ruptura de sus supuestos. Una complicación particular a este análisis es que la mayoría de los métodos de imputación conocidos están diseñados con el fin de preservar las propiedades de las distribuciones de probabilidad condicional ya que el análisis típico incluye algún tipo de regresión de interés. En cambio, modelos que traten de conservar las propiedades de la distribución conjunta son raros, y más los métodos de agregación

que no exijan una regresión lineal de antemano, sino que se satisfagan con la aplicación de un método de reducción de dimensiones.

Ante este panorama, los métodos propuestos por Josse y Husson (2016) podrían parecer una respuesta, ya que ellos han diseñado varios métodos de agregado para procedimientos de reducción de dimensiones, tales como el análisis de componentes principales («Principal Components Analysis», PCA), el análisis de correspondencias («Correspondence Analysis», CA), o el de correspondencias múltiples («Multiple Correspondence Analysis», MCA), entre otros. Incluso tienen una propuesta para el MFA mediante regularización iterativa (RIMFA), pero su propuesta es de imputación simple, no múltiple, lo que conlleva una incapacidad de tomar en cuenta la incertidumbre de los datos, como antes se mencionó.

Felizmente, Voillet et al. (2016) diseñaron un método de imputación múltiple con análisis de factores múltiples («Multiple Imputation with Multiple Factor Analysis», MIMFA). Sin embargo, este método no me sirve por dos razones: 1) remueve la posibilidad de trabajar datos mixtos y sólo opera datos numéricos, y 2) sólo puede operar con datos perdidos en forma de individuos perdidos por fuente, como se puede ver en 3.

Ante tal vacío en la literatura, no queda más que proponer una solución propia. Por ello, propongo un nuevo método: la imputación múltiple mediante análisis de factores múltiples regularizado iterativo, que por continuar con la nomenclatura hasta la fecha, llamo MIRIMFA: «Multiple Imputation by Regularised Iterative Multiple Factor Analysis».

En resumen, MIRIMFA consta de aplicar el método de análisis de factores múltiples regularizado iterativo («Regularised Iterative Multiple Factor Analysis imputation», RIMFA) (Josse y Husson 2016) en múltiples ocasio-

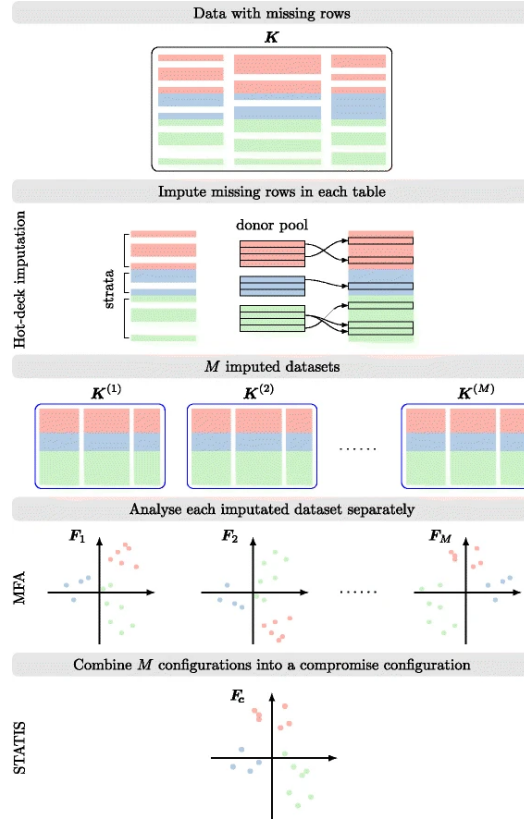


Figura 3: Resumen visual del método MIMFA (Voillet et al. 2016)

nes, ya que éste puede manejar cualquier forma de los datos faltantes, así como datos mixtos, y luego aplicar un Análisis Conjunto de Cuadros - Estructuración de Cuadros en Tres Índices del Estadístico («Analyse Conjointe de tableaux - Structuration des Tableaux à Trois Indices de la Statistique», ACT-STATIS) (Lavit et al. 1994) a las tablas disyuntivas resultantes para encontrar un espacio compromiso que permita hacerlas comparables. Finalmente, transforma la tabla disyuntiva de vuelta a una tabla de datos originales ya imputados. Así, MIRIMFA es un hí-

brido entre las propuestas de Josse & Husson y Voillet et al.: RIMFA y MIMFA. Por lo anterior, refiero al lector a los artículos correspondientes de estar interesado en las bases matemáticas de los pasos, así mismo, anexo un paquete de R (R Core Team 2018) que permita correr MIRIMFA. Por mi parte, me satisfago en mostrar la validez de mi método mediante una simulación que muestre empíricamente sus propiedades para luego proceder a aplicarlo en el problema principal: la SCCS.

3.2 Aplicación de MIRIMFA con la base de datos «Poison»

Para demostrar las propiedades de MIRIMFA, se corrió sobre la base de datos «Poison», que se puede obtener del paquete de R FactoMineR (Lê, Josse, y Husson 2008). «Poison» es una base de datos obtenida de una encuesta a alumnos y personal de una escuela primaria que experimentaron intoxicación por alimentos. Se les preguntaron sus síntomas y qué comieron. «Poison» tiene 55 individuos registrados y 15 variables agrupadas en cuatro grupos: dos variables continuas de tiempo y edad, dos categóricas de sexo y condición de enfermedad, cinco de síntomas y seis de alimentos ingeridos.

La ventaja de usar MIRIMFA en «Poison» es que tiene una estructura similar a la SCCS, aunque en menor escala, además de que tenemos los datos completos, podemos amputarlos, aplicar MIRIMFA a los datos amputados y comparar el análisis del resultado con el de los datos completos. La estructura de datos perdidos obtenida de la imputación puede verse en la figura ??.

Al aplicársele tanto MIRIMFA ($M=20$) como RIMFA a «Poison», vemos que el uso de múltiples imputaciones sí logra estabilizar el modelo mucho más cerca de los datos originales, tal como se ve en las gráficas 6 a 7.

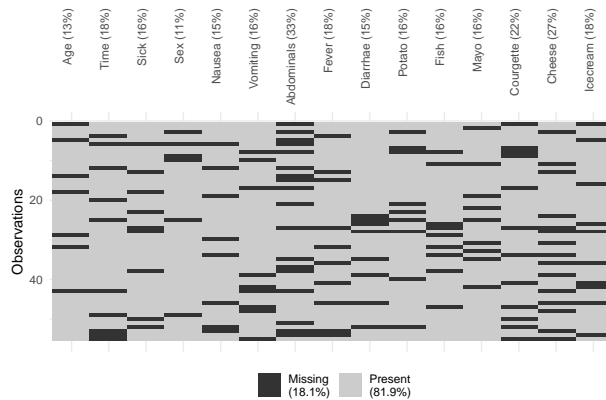


Figura 4: Estructura de datos perdidos en amputación de «Poison»

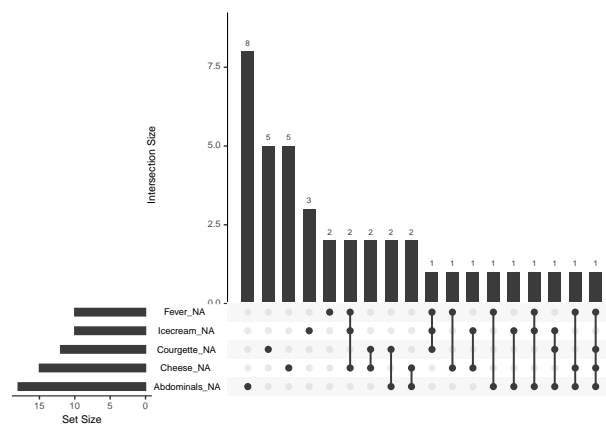


Figura 5: Estructura de datos perdidos en amputación de «Poison»

Así mismo, vemos en la figura 8 que MIRIMFA suele ajustar las coordenadas de las variables compromiso bastante cerca de las de las imputaciones individuales, aún cuando haya dispersión entre estas. Si la dispersión fuera demasiado grande, diríamos que la imputación no es confiable. Otra forma de identificar si la imputación es razonable, es mediante la gráfica de proyección de los factores principales (figura 9). Si los factores de las imputaciones se alínean con los del espacio compromiso, podemos decir

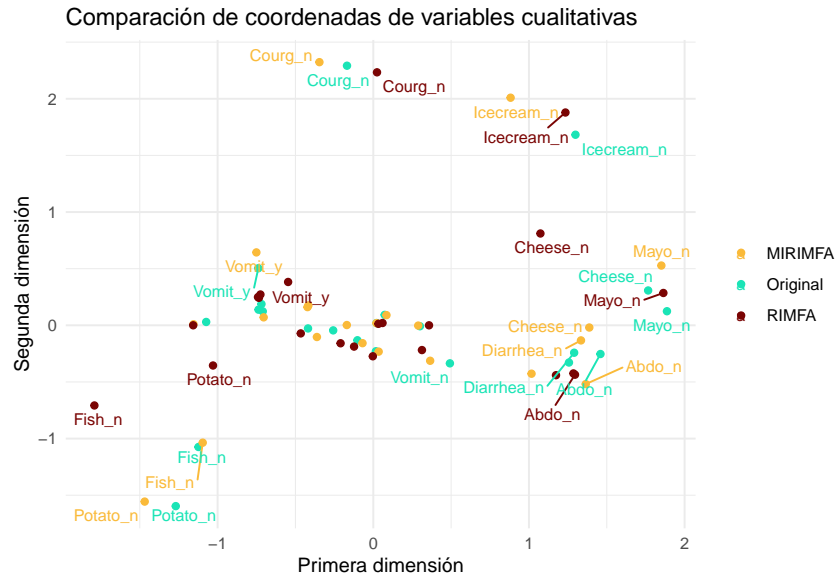


Figura 6: Comparación de coordenadas de variables cualitativas en las primeras dimensiones de un MFA aplicado a los datos originales, los datos agregados de MIRIMFA, y los datos imputados mediante una sola aplicación de RIMFA.

que la imputación es apropiada. Visualmente, esto se manifiesta como dos haces de flechas yendo en dirección de los ejes horizontal y vertical. Si los haces de flechas van en todas direcciones y se revuelven, entonces la imputación es incierta. Vemos en este caso que ambas gráficas muestran una cercanía considerable, por lo que podemos decir que la imputación es de calidad y la incertidumbre está bajo control.

Con lo anterior, podemos decir que MIRIMFA trabaja apropiadamente y que nos permite visualizar la incertidumbre causada por los datos perdidos. Por ello procederé a aplicarlo sobre la SCCS.

Análisis de factores múltiples – Individuos

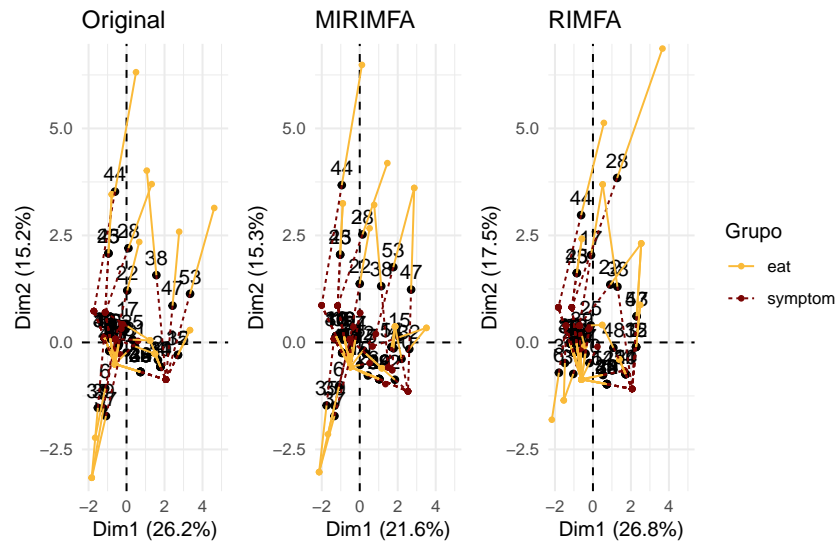


Figura 7: Gráfica de individuos parciales según grupo para cada base de datos: original, imputada con MIRIMFA, imputada sólo con RIMFA.

3.3 Aplicación de MIRIMFA a la SCCS

Para aplicar MIRIMFA en la SCCS, tenemos que solucionar un problema antes: el mecanismo de pérdida de los datos. Dado que los datos perdidos se deben a la falta de reporte, y los autores tienden a no reportar lo que consideran típico u ordinario para ellos, es de esperar que los datos faltantes muestren un sesgo a favor de los valores típicos en las sociedades industrializadas, que son el lugar de origen de la mayoría de los etnógrafos. Este comportamiento corresponde al del mecanismo de pérdida MNAR, para el cual no se suelen recomendar ni RIMFA, ni MIMFA, y por lo tanto MIRIMFA hereda este problema de ellos. MIRIMFA, entonces, es recomendable para bases con datos perdidos mediante MCAR o MAR.

Afortunadamente, el mecanismo de pérdida de la SCCS está sesgado ha-

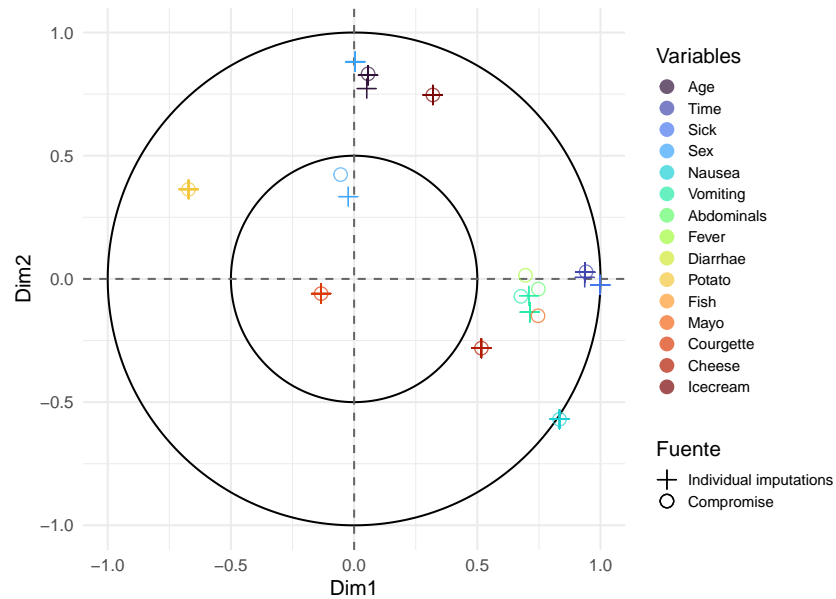


Figura 8: Coordenadas de variables en las imputaciones con referencia al espacio compromiso.

cia los valores típicos de ciertos casos o sociedades, por lo que el sesgo introducido por esta pérdida es modelable mediante mayores pesos dados a ciertas sociedades. Y es justo este caso el único de MNAR para el cual RIMFA está bien equipado, ya que acepta la introducción deliberada de pesos a los renglones de la matriz de observaciones. Es por eso que, en este trabajo, he decidido darle un peso doble a las sociedades urbanizadas con dinero y escritura verdaderos ($SCCS149=5$, $SCCS155 \geq 4$, $SCCS152 \geq 4$). Con el criterio anterior, tenemos que las sociedades que se usarán como base para introducir este peso son las etiquetadas como “Egyptians, Riffians, Ancient Romans, Spanish Basques, Irish, Russians, Turks, Hebrews, Babylonians, Koreans, Chekiang, Japanese, Uttar Pradesh, Burmese, Negri Sembilan, Thai, Javanese, Balinese, Aztec y Haitians”.

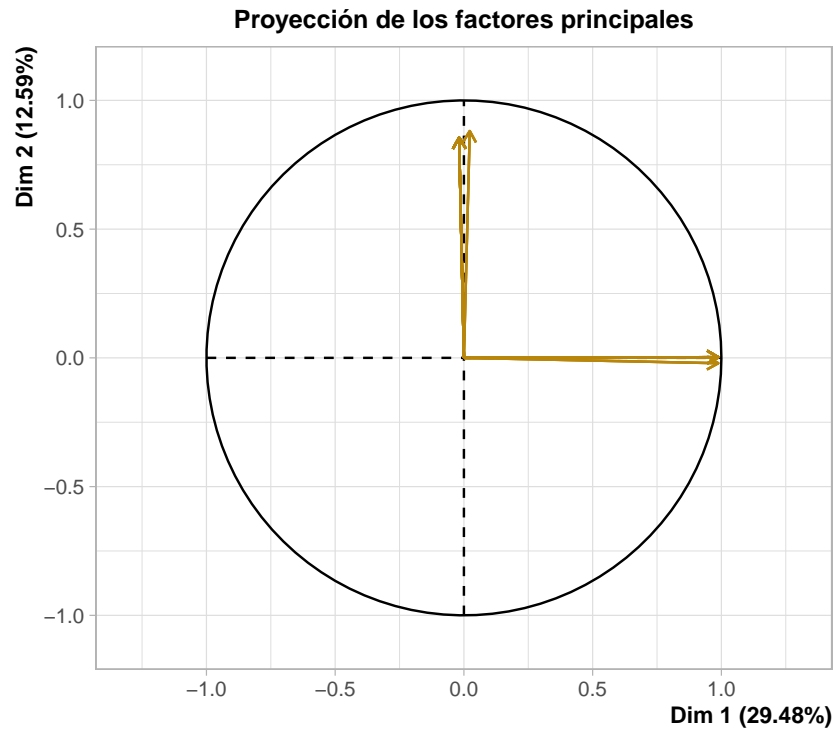


Figura 9: Proyección de los factores de las imputaciones con referencia al espacio comprometido.

Nótese que esto puede considerarse como meter mano en el modelo y lo más apropiado sería correr un análisis de sensibilidad para ver cuánto afectan distintos pesos inducidos (Van Buuren 2018). Desafortunadamente, ya como queda el modelo, el poder de cómputo necesario es inmenso y hacer el análisis de sensibilidad queda fuera del alcance de este trabajo. No tengo más remedio, pues, que asumir que la elección de pesos fue apropiada y seguir con el análisis.

Con esto fuera del camino, procedo a aplicar MIRIMFA a la SCCS. Dado que el porcentaje de datos perdidos es 45.62%, corrí 45.62 imputaciones,

según la recomendación estándar (Van Buuren 2018).

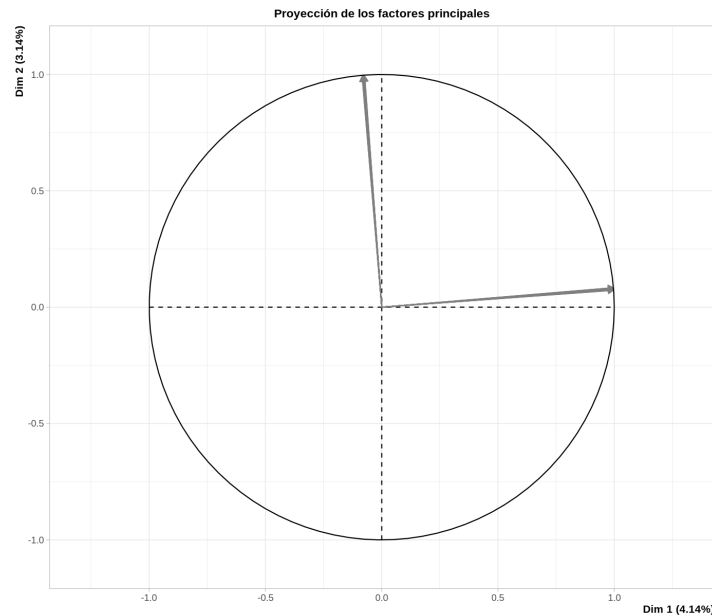


Figura 10: Proyección de los primeros factores de las imputaciones de la SCCS con referencia al espacio compromiso.

Primero, diagnosticaré qué tan apropiadas son las imputaciones. Dado el poder de cómputo requerido y la gran cantidad de variables que sobresaturarían el gráfico, no muestro la gráfica de las coordenadas de las variables de cada imputación. En cambio, me satisfaré con la gráfica de los primeros dos factores principales, que vemos en la figura 10. Vemos en ésta que los dos primeros factores de las imputaciones están tan bien alineados que es prácticamente imposible distinguirlos, aunque vemos también un sesgo del espacio compromiso con respecto a ellos (el ángulo de los factores con respecto a los ejes horizontal y vertical). Presumiblemente este sesgo se deba a que el espacio compromiso tenga que mantener la perpendicularidad con respecto a los otros factores no grafi-

cados aquí. A pesar de este sesgo, doy por buena la imputación y procedo a correr el MFA final para obtener el ordenamiento de las variables.

Cuadro 1: Contribuciones de las dimensiones obtenidas del Análisis de Factores Múltiples en la SCCS.

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.1	35.907877	4.1398339	4.139834
Dim.2	27.218622	3.1380461	7.277880
Dim.3	12.401747	1.4298025	8.707683
Dim.4	10.788529	1.2438140	9.951496
Dim.5	10.465710	1.2065960	11.158093
Dim.6	9.719642	1.1205815	12.278674
Dim.7	9.568074	1.1031072	13.381781
Dim.8	9.232142	1.0643775	14.446159
Dim.9	8.477768	0.9774053	15.423564
Dim.10	8.247442	0.9508510	16.374415
Dim.11	8.078304	0.9313510	17.305766
Dim.12	7.911526	0.9121231	18.217889
Dim.13	7.695267	0.8871906	19.105080
Dim.14	7.521945	0.8672082	19.972288
Dim.15	7.423910	0.8559057	20.828194
Dim.16	7.321249	0.8440699	21.672263
Dim.17	7.212471	0.8315288	22.503792
Dim.18	7.112478	0.8200005	23.323793
Dim.19	7.020981	0.8094518	24.133244
Dim.20	6.946642	0.8008813	24.934126
Dim.21	6.857059	0.7905532	25.724679

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.22	6.740640	0.7771311	26.501810
Dim.23	6.710813	0.7736924	27.275502
Dim.24	6.667870	0.7687415	28.044244
Dim.25	6.609508	0.7620129	28.806257
Dim.26	6.524660	0.7522307	29.558488
Dim.27	6.441433	0.7426355	30.301123
Dim.28	6.350710	0.7321759	31.033299
Dim.29	6.316455	0.7282267	31.761526
Dim.30	6.236268	0.7189819	32.480508
Dim.31	6.179351	0.7124199	33.192928
Dim.32	6.098198	0.7030638	33.895991
Dim.33	6.062284	0.6989232	34.594915
Dim.34	6.024673	0.6945870	35.289502
Dim.35	5.935938	0.6843568	35.973858
Dim.36	5.918222	0.6823143	36.656172
Dim.37	5.830206	0.6721669	37.328339
Dim.38	5.813623	0.6702549	37.998594
Dim.39	5.785430	0.6670046	38.665599
Dim.40	5.763340	0.6644579	39.330057
Dim.41	5.703671	0.6575785	39.987635
Dim.42	5.652893	0.6517244	40.639360
Dim.43	5.632007	0.6493164	41.288676
Dim.44	5.600660	0.6457024	41.934379
Dim.45	5.542973	0.6390516	42.573430
Dim.46	5.541256	0.6388537	43.212284
Dim.47	5.463808	0.6299247	43.842208

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.48	5.436683	0.6267975	44.469006
Dim.49	5.365519	0.6185929	45.087599
Dim.50	5.336846	0.6152872	45.702886
Dim.51	5.302515	0.6113292	46.314215
Dim.52	5.269656	0.6075408	46.921756
Dim.53	5.260828	0.6065231	47.528279
Dim.54	5.205185	0.6001079	48.128387
Dim.55	5.187279	0.5980435	48.726430
Dim.56	5.107503	0.5888461	49.315277
Dim.57	5.082960	0.5860166	49.901293
Dim.58	5.034319	0.5804086	50.481702
Dim.59	5.009232	0.5775165	51.059218
Dim.60	4.983538	0.5745541	51.633772
Dim.61	4.963897	0.5722897	52.206062
Dim.62	4.916717	0.5668504	52.772912
Dim.63	4.900761	0.5650107	53.337923
Dim.64	4.848297	0.5589621	53.896885
Dim.65	4.829662	0.5568137	54.453699
Dim.66	4.782399	0.5513648	55.005064
Dim.67	4.765624	0.5494307	55.554495
Dim.68	4.740568	0.5465421	56.101037
Dim.69	4.698756	0.5417215	56.642758
Dim.70	4.694047	0.5411787	57.183937
Dim.71	4.662449	0.5375356	57.721473
Dim.72	4.617508	0.5323544	58.253827
Dim.73	4.589370	0.5291103	58.782937

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.74	4.545199	0.5240178	59.306955
Dim.75	4.488993	0.5175379	59.824493
Dim.76	4.466259	0.5149168	60.339410
Dim.77	4.443409	0.5122825	60.851692
Dim.78	4.438083	0.5116684	61.363361
Dim.79	4.406628	0.5080420	61.871403
Dim.80	4.391349	0.5062805	62.377683
Dim.81	4.377672	0.5047036	62.882387
Dim.82	4.357435	0.5023705	63.384757
Dim.83	4.317038	0.4977131	63.882470
Dim.84	4.293458	0.4949946	64.377465
Dim.85	4.238772	0.4886897	64.866155
Dim.86	4.222192	0.4867782	65.352933
Dim.87	4.203802	0.4846580	65.837591
Dim.88	4.164322	0.4801064	66.317697
Dim.89	4.133587	0.4765630	66.794260
Dim.90	4.097527	0.4724056	67.266666
Dim.91	4.087412	0.4712394	67.737905
Dim.92	4.059212	0.4679882	68.205894
Dim.93	4.037376	0.4654707	68.671364
Dim.94	4.008130	0.4620990	69.133463
Dim.95	3.978991	0.4587394	69.592203
Dim.96	3.946930	0.4550431	70.047246
Dim.97	3.910647	0.4508601	70.498106
Dim.98	3.886992	0.4481329	70.946239
Dim.99	3.878576	0.4471626	71.393401

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.100	3.845870	0.4433919	71.836793
Dim.101	3.840762	0.4428030	72.279596
Dim.102	3.831900	0.4417813	72.721378
Dim.103	3.822657	0.4407157	73.162093
Dim.104	3.810784	0.4393469	73.601440
Dim.105	3.773533	0.4350522	74.036492
Dim.106	3.732895	0.4303669	74.466859
Dim.107	3.726855	0.4296707	74.896530
Dim.108	3.698035	0.4263480	75.322878
Dim.109	3.676446	0.4238590	75.746737
Dim.110	3.655385	0.4214308	76.168168
Dim.111	3.629057	0.4183955	76.586563
Dim.112	3.613594	0.4166127	77.003176
Dim.113	3.590838	0.4139892	77.417165
Dim.114	3.580572	0.4128056	77.829971
Dim.115	3.534201	0.4074595	78.237430
Dim.116	3.524621	0.4063550	78.643785
Dim.117	3.509364	0.4045960	79.048381
Dim.118	3.495281	0.4029724	79.451354
Dim.119	3.478178	0.4010006	79.852354
Dim.120	3.453132	0.3981130	80.250467
Dim.121	3.409341	0.3930644	80.643532
Dim.122	3.403174	0.3923533	81.035885
Dim.123	3.388844	0.3907012	81.426586
Dim.124	3.356089	0.3869248	81.813511
Dim.125	3.339220	0.3849801	82.198491

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.126	3.315042	0.3821926	82.580684
Dim.127	3.290220	0.3793308	82.960015
Dim.128	3.263045	0.3761979	83.336212
Dim.129	3.248309	0.3744989	83.710711
Dim.130	3.238927	0.3734173	84.084129
Dim.131	3.211969	0.3703093	84.454438
Dim.132	3.180532	0.3666849	84.821123
Dim.133	3.157004	0.3639723	85.185095
Dim.134	3.140790	0.3621030	85.547198
Dim.135	3.118578	0.3595421	85.906740
Dim.136	3.103989	0.3578602	86.264600
Dim.137	3.067913	0.3537009	86.618301
Dim.138	3.052709	0.3519481	86.970249
Dim.139	3.033702	0.3497567	87.320006
Dim.140	3.013142	0.3473864	87.667392
Dim.141	3.001370	0.3460292	88.013422
Dim.142	2.981022	0.3436832	88.357105
Dim.143	2.954571	0.3406337	88.697739
Dim.144	2.927939	0.3375633	89.035302
Dim.145	2.911555	0.3356744	89.370976
Dim.146	2.895360	0.3338073	89.704784
Dim.147	2.867292	0.3305713	90.035355
Dim.148	2.847279	0.3282640	90.363619
Dim.149	2.809649	0.3239256	90.687544
Dim.150	2.801892	0.3230313	91.010576
Dim.151	2.791069	0.3217835	91.332359

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.152	2.738590	0.3157332	91.648092
Dim.153	2.726685	0.3143606	91.962453
Dim.154	2.709214	0.3123464	92.274799
Dim.155	2.676606	0.3085870	92.583386
Dim.156	2.674183	0.3083077	92.891694
Dim.157	2.653939	0.3059737	93.197668
Dim.158	2.613972	0.3013659	93.499034
Dim.159	2.597027	0.2994123	93.798446
Dim.160	2.592535	0.2988945	94.097340
Dim.161	2.554276	0.2944836	94.391824
Dim.162	2.527096	0.2913499	94.683174
Dim.163	2.494671	0.2876117	94.970786
Dim.164	2.460387	0.2836590	95.254445
Dim.165	2.438443	0.2811291	95.535574
Dim.166	2.405849	0.2773714	95.812945
Dim.167	2.373631	0.2736569	96.086602
Dim.168	2.344635	0.2703139	96.356916
Dim.169	2.277495	0.2625733	96.619489
Dim.170	2.241804	0.2584585	96.877948
Dim.171	2.225854	0.2566196	97.134567
Dim.172	2.182792	0.2516550	97.386222
Dim.173	2.129946	0.2455624	97.631785
Dim.174	2.092346	0.2412274	97.873012
Dim.175	2.039347	0.2351171	98.108129
Dim.176	2.018311	0.2326919	98.340821
Dim.177	1.905030	0.2196317	98.560453

	Eigenvalor	Varianza (%)	Varianza acumulada (%)
Dim.178	1.869237	0.2155051	98.775958
Dim.179	1.719384	0.1982285	98.974187
Dim.180	1.680560	0.1937525	99.167939
Dim.181	1.606515	0.1852157	99.353155
Dim.182	1.548581	0.1785365	99.531691
Dim.183	1.510692	0.1741683	99.705860
Dim.184	1.388477	0.1600781	99.865938
Dim.185	1.162824	0.1340624	100.000000

3.4 Análisis de Factores Múltiples

El Análisis de Factores Múltiples (MFA) redujo las 1752 que nos quedaban a 185 dimensiones. De éstas, se necesitan 3 para alcanzar el codo, pero eso sólo cubre el 8.71%. Por lo tanto, para que realmente las dimensiones representen a la SCCS, utilicé 96 dimensiones, que cubren el 70.05.

Para ver que estas dimensiones tengan sentido, exploraremos las primeras dos. Como se ve en la figura 12, las variables fueron ordenadas en forma de “V” con respecto a las primeras dos dimensiones.

Revisando el significado de los códigos de cada una, se puede ver que el eje horizontal contrasta variables sobre complejidad social y egalitarismo de género contra patrilinearidad. A la izquierda tenemos variables como el alto involucramiento de los padres en la crianza de las hijas, falta de juicios morales ante el aborto espontáneo, continuidad entre paredes y techo (típico de métodos de construcción simple), y falta de arado, irrigación e impuestos. En la otra mano, a la derecha tenemos presencia de

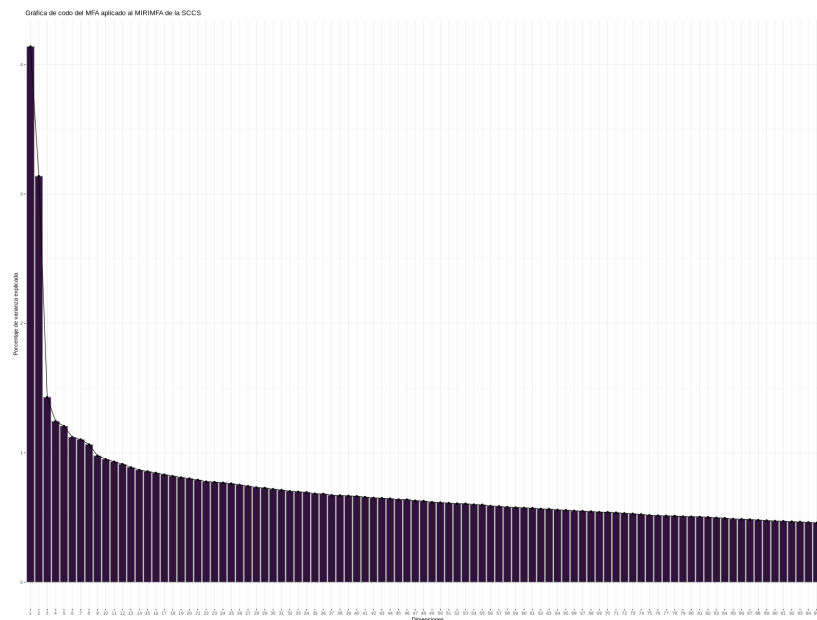


Figura 11: Gráfica de codo del MFA aplicado al MIRIMFA de la SCCS.

estos tres últimos, burocracia, estados y estratos sociales así como tabúes menstruales afectando a la mujer, falta de contracepción, evasión de parientes del sexo opuesto, entre otros. Lo anterior muestra una dicotomía en el trato de género y en complejidad social.

En cambio, el eje vertical muestra categorías relacionadas con la laxitud, en la parte superior, en contraste con una alta punitividad, en la parte inferior. Por ejemplo, arriba vemos variables como falta de castigo al incesto, poco control a niñas, falta de infanticidio y permisividad a relaciones con parientes políticos. Abajo, en cambio, vemos un alto uso del sistema judicial, burocracia, mutilación de mujeres, y alto control de los hijos. Es por esto que concluyo que esta dimensión es de la punitividad.

Podemos comprobar esto con la figura 13, donde se clasifican las socie-

dades según estas mismas dimensiones y vemos que se replica la misma forma. Vemos también que las sociedades más complejas y patrilineales, como los rusos (SCCS54), los romanos (SCCS49) o los vascos (SCCS50) están en el cuadrante superior derecho, tal como se esperaría de la lectura de la gráfica de variables. En la parte inferior, notoriamente tenemos a los Ashanti (SCCS19), que son famosos por su sistema legal altamente sofisticado y punitivo. Finalmente, en el cuadrante superior izquierdo vemos comunidades como la Yukaghir (SCCS120), una sociedad cazadora-recolectora con poca distinción entre géneros y poca punitividad.

Vemos también en la figura 13 que las sociedades complejas y patriarcales son las mejores representadas por estos ejes, ya que tienen un mayor valor de coseno cuadrado, que es una medida de la calidad de representación. Aún así, esta calidad es pequeña ($<30\%$), y vemos también que la varianza explicada por estas primeras dimensiones es apenas del 7.28%. En términos simples, esto nos muestra que el MFA considera que las dimensiones más importantes a la hora de clasificar sociedades son el patriarcado y complejidad social, y el punitivismo, pero aún así, ésto sólo explica un 7.28% de la, por llamarla así, “diversidad” de las sociedades, y para dar cuenta de al menos un 70%, requerimos de las 96 dimensiones mencionadas anteriormente.

3.5 Selección de variables

Con estas 96, se sumaron las calidades de representación (cosenos cuadrados) ponderadas por la varianza explicada por sus respectivas dimensiones. Las sumas resultantes se tomaron como la contribución de la variable al MFA total y se ordenaron de mayor a menor. Las primeras 20 variables seleccionadas de esta manera se pueden ver en 2. La lista

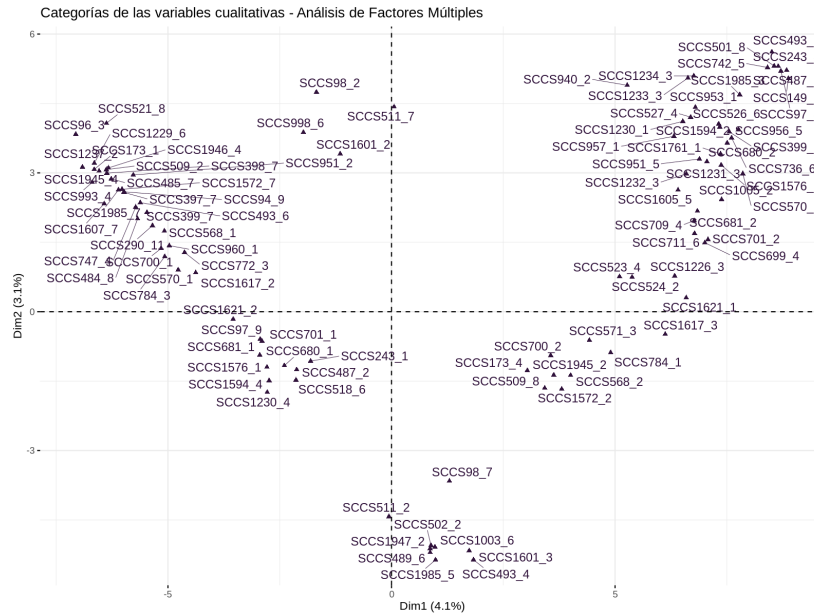


Figura 12: Gráfica de variables en las primeras dimensiones del MFA sobre la SCCS

completa se puede ver en el repositorio de GitHub

Cuadro 2: Primeras variables de la SCCS seleccionadas por el MFA.

Código	Título	Categoría
SCCS1756	Size of Local Community	Subsistence, Kinship, Politics
SCCS1123	Major Agricultural Staple	Subsistence
SCCS1862	Highest Number Counted	Economy, Subsistence

SCCS917	Historical Frame for Pinpointing Date Where Slaveholding Present	Economy, Labour, Politics
SCCS65	Types of Dwelling	Settlement, Community
SCCS1753	Depth of Unilineal Descent	Subsistence, Kinship, Politics
SCCS644	Patterns for Siblings-in-law	Kinship
SCCS1445	Secondary Crop Name	Household, Labour, Gender
SCCS94	Political Power - Second Most Important Source	Politics
SCCS204	Subsistence economy: hunting [Note, identical to EA002]	Subsistence, Economy
SCCS1122	Log10 of Total Population	Population
SCCS833	Subsistence Economy: Dominant Mode	Subsistence, Economy
SCCS86	Selection of Executive	Politics
SCCS206	Subsistence economy: animal husbandry [Note, identical to EA004]	Subsistence, Economy
SCCS285	House construction: wall material [Note, identical to EA081]	Community organization, Housing, Ecology
SCCS1349	Primary Crop Name	Household, Labour, Gender

SCCS1649	Frequency of Internal Warfare (Resolved Rating)	Warfare, Politics
SCCS290	House construction (secondary house type): wall material [Note, identical to EA086]	Community organization, Housing, Ecology
SCCS642	Patterns for Siblings	Kinship
SCCS93	Political Power - Most Important Source	Politics

En esta pequeña selección vemos que varias de las variables mejor ubicadas tienen significados parecidos (por ejemplo: SCCS1756, SCCS1122). Estas variables suelen ser distintas operacionalizaciones del mismo constructo y no tiene sentido conservarlas en el cuestionario final. Además, algunas de estas preguntas son fáciles de preguntar directamente a un investigador, como

3.6 Elaboración del cuestionario

Se comenzó con el análisis

Para demostrar las propiedades de MIRIMFA, se corrió sobre la base de datos «Poison», que se puede obtener del paquete de R FactoMineR. «Poison» es una base de datos obtenida de una encuesta a alumnos y personal de una escuela primaria que experimentaron intoxicación por alimentos. Se les preguntaron sus síntomas y qué comieron. «Poison» tiene 55 individuos registrados y 15 variables agrupadas en cuatro grupos: dos variables continuas de tiempo y edad, dos categóricas de sexo y condición de enfermedad, cinco de síntomas y seis de alimentos ingeridos. La ventaja de usar MIRIMFA en «Poison» es que tenemos los datos com-

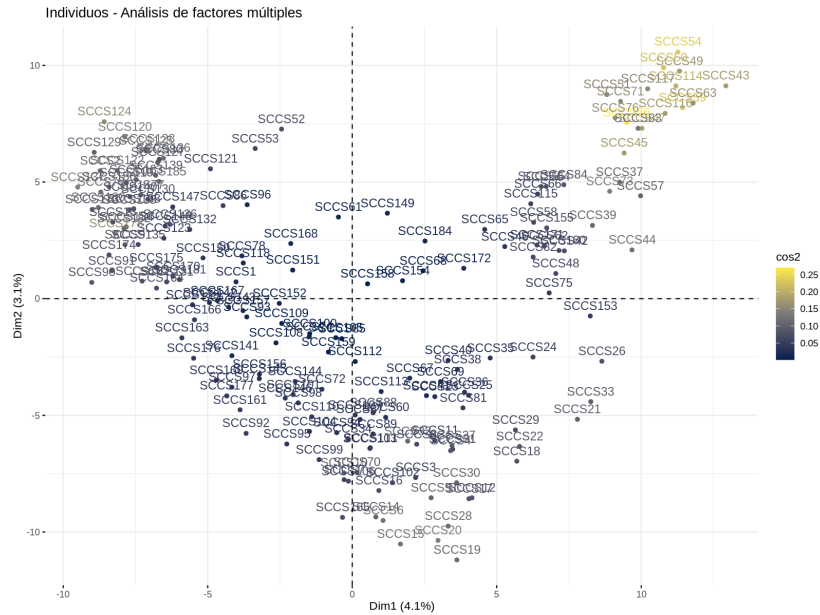


Figura 13: Gráfica de individuos en las primeras dimensiones del MFA sobre la SCCS

pletos, podemos amputarlos, aplicar MIRIMFA a los datos amputados y comparar el análisis del resultado con el de los datos completos.

Por lo anterior, elegí la especificación completamente condicional («Fully Conditional Specification», FCS) implementada mediante el algoritmo de imputación múltiple mediante cadenas de ecuaciones («Multiple Imputation by Chained Equations», MICE). Dicho esto, también tendré que hacer varios ajustes manuales ya que yo considero que el mecanismo de pérdida es MNAR. Anteriormente ya descartamos la posibilidad de que la pérdida fuese por MCAR, lo que nos dejaba con las opciones de MAR y MNAR. Como consecuencia de la definición de MNAR, es imposible diferenciar entre este y MAR. Para determinarlo, siempre se recurre a

experiencia de área. Dadas entrevistas con etnólogos que he tenido, he llegado a la conclusión de que gran parte del mecanismo de pérdida de los datos se debe al mismo valor de los datos faltantes. Esto es así porque, en la época en la que se hicieron las etnografías, diarios de misioneros y de locales en las que se recarga el SCCS, el comportamiento típico de los escritores era no reportar aquello que no consideraban digno de anotar. ¿Qué no es digno de anotar? Lo trivial y ordinario a los ojos del observador. Esto significa que es muy probable que si un dato no se reporta, es porque ese dato tenía un valor típico para las sociedades industrializadas de las que venía el investigador. Por ello, he decidido separar a las comunidades urbanas, con dinero y alfabetizadas del resto de los datos, obtener la medida de tendencia central de cada variable en estas condiciones e implementar un desplazamiento δ en dirección al valor típico de las comunidades industriales sobre el cual se aplica la imputación múltiple y concluir con un análisis de sensibilidad sobre el agregado de resultados final con el propósito de verificar que mi suposición sea razonable. Esto es vital ya que tenemos un caso de más del 20% de los datos faltantes, por lo que el mecanismo de pérdida no es ignorable.

4 Reducción de variables culturales mediante análisis de componentes principales

Para la selección de los factores culturales, encontré que la SCCS ya incluye varias variables, pero como mencioné el cuerpo principal del texto, éstas están desactualizadas. Así que hay que recolectar información sobre ellas.

Dado que no sabemos qué variables culturales podrían tener un impacto

en el efecto Lombard y el costo de un levantamiento llega a ser prohibitivo, la encuesta que utilicemos para levantar los valores debe de ser simultáneamente breve, para no abrumar a los investigadores, y representativa de la distribución conjunta de los datos. Por ello, propongo la reducción de las 1751 variables sobre economía, rituales, roles sociales, género, entre otros que contempla la SCCS original. Claramente, recolectar tal información sería tedioso para los documentalistas. Para ello, se utilizó un análisis de componentes principales de los datos originales de la SCCS y se la utilizó para seleccionar variables a levantar.

Es importante considerar que al realizar un análisis de componentes principales directamente sobre el SCCS, se busca identificar relaciones entre las variables y resaltar aquellas que mejor representan a las demás. Sin embargo, estas variables destacadas no necesariamente coinciden con las que definen mejor a una comunidad, sino más bien con las que destacan en la selección inicial de variables. En nuestro caso, al aplicar el ACP, se observó que de las 574 variables originales del SCCS relacionadas con el género, más de la mitad estaban estrechamente vinculadas con otras variables como la infancia, el matrimonio y el liderazgo.

Es complicado argumentar que el género es la característica definitoria de una comunidad, especialmente cuando estoy desarrollando un cuestionario para investigaciones en campos como la fonética, donde se espera que las preocupaciones estén más centradas en aspectos como la subsistencia, la economía, la geografía y la política, en lugar de en ritos específicos como los de iniciación. A pesar de que las variables relacionadas con ritos de iniciación se destacaron en la selección, sería difícil justificar el llenar un cuestionario de amplia cobertura con preguntas sobre estos ritos en lugar de otros aspectos más ampliamente relevantes

en el tejido social de la mayoría de las comunidades, como los ritos de matrimonio o funerales.

Por esta razón, se optó por dividir la lista de variables del SCCS original en categorías establecidas previamente y luego aplicar el ACP a cada una de ellas. De esta manera, se pudieron seleccionar las variables más destacadas de cada categoría y agruparlas para crear una lista final de preguntas. Este enfoque permite que cada categoría tenga igual importancia en el cuestionario final, en lugar de ser dominada por categorías más pesadas como el género. Con este método, se espera obtener una serie de variables que reflejen de manera más precisa las características de las comunidades, sin la necesidad de utilizar el cuestionario completo que incluye las 1781 variables del SCCS original.

Primero se efectuó un análisis de datos perdidos ya que el método de recolección en la SCCS causó que hubiera variables que no estaban recolectadas excepto para unas cuantas observaciones. Se decidió arbitrariamente remover toda variable cuya tasa de datos faltantes fuese mayor a 50%. Esto permitió que el porcentaje de datos faltantes se redujese de 45.62% a 17.35% con el costo de reducir el número de variables de 1751 a 866.

Para determinar qué variables eran más significativas en cada categoría de la SCCS, utilicé un algoritmo basado en el Análisis de Correspondencia Múltiple (MCA). Este método evalúa la contribución de cada variable a las dimensiones principales del análisis, lo que se refleja en su “carga” dentro del MCA. El algoritmo itera sobre cada categoría, realiza un MCA para cada una y calcula la suma ponderada de las cargas de las variables en las dimensiones principales, teniendo en cuenta la proporción de varianza

explicada por cada dimensión.

Para determinar el número óptimo de dimensiones, seguí un criterio de revisión manual. Observé la gráfica de codo y, si era plana, seleccioné el número de dimensiones que explicara al menos el 80% de la varianza. Si la gráfica mostraba codos cercanos al 80%, tomé ese punto como referencia. En casos donde la gráfica era completamente plana, utilicé todas las dimensiones, ya que no había reducción de dimensionalidad.

Después, el algoritmo ordenaba las variables según estas sumas ponderadas en orden descendente, lo que nos permitía identificar las variables más relevantes en cada categoría.

Luego de este proceso automatizado, llevé a cabo una revisión manual de las variables en comparación con el glosario de la SCCS. Esto incluyó la eliminación de todas las categorías “proxy” para los datos faltantes, generando así un nuevo cuestionario.

El objetivo del cuestionario era recoger aproximadamente 100 variables, asegurando que cada categoría estuviera equitativamente representada por sus variables más importantes. Si una variable ya estaba incluida en otra categoría, no se repetía, sino que se saltaba para incluir la siguiente más importante dentro de la categoría en evaluación. Utilicé estas variables para formular las preguntas, basándonos en las variables mismas. A veces, fue necesario descomponer una sola variable en varias preguntas o formular una pregunta que cubriera varias variables. En otras ocasiones, la variable original de la SCCS era difícil de medir en la actualidad, por lo que hice un esfuerzo por encontrar el mejor equilibrio entre la facilidad de respuesta para el encuestado y el espíritu o intención detrás de la variable original.

A continuación se enumeran los números de pregunta (según la versión PDF del cuestionario) y sus variables correspondientes en la SCCS.

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
1	Email de contacto	-	-
2	Género del encuestados	-	-
3	Edad del encuestados	-	-
4	Nacionalidad del encuestados	-	-
5	Pertenencia a la comunidad	-	-
6	Nombre de la comunidad	-	-
7	Ubicación: coordenadas	-	-
8	Población	Settlement, Population	64, 156, 1122
9	Tiempo focal	-	-
10	Área cubierta	Modernisation	64, 156
11	Proporción de sexos	Politics	1689
12	Endónimos de vecinos pasados	Politics	1877
13	Endónimos de nuevos vecinos	Politics	1878
14	Estabilidad de provisión de alimentos	Subsistence, Economy	19
15	Conserva de alimentos (provisión estable)	Subsistence, Economy	19

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
16	Conserva de alimentos (variación diaria)	Subsistence, Economy	19
17	Conserva de alimentos (variación estacional)	Subsistence, Economy	19
18	Conserva de alimentos (variación anual)	Subsistence, Economy	19
19	Resiliencia de redes comerciales	Subsistence, Economy	19
20	Actividad de producción alimentaria	Subsistence, Gender	833, 820
21	Contribución de agricultura	Modernisation, Gender	151, 814
22	Cultivo principal	Subsistence, Ecology	233, 1123
23	Contribución de pesca	Subsistence, Economy	205
24	Medios de carga en tierra	Economy	13, 154
25	Presencia de esclavitud	Class	275
26	Esclavitud hereditaria	Class	274
27	Frecuencia de manumisión	Economy, Labour	918
28	Población esclava	Economy, Labour	920
29	Contribución de géneros por tarea	Economy, Labour	104, 109, 110, 112, 113, 136, 148
30	Comienzo de destete	Infancy	44

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
31	Final de destete	Infancy	45
32	Distancia entre padres y recién nacidos	Infancy	23
33	Acompañantes de sueño primera infancia	Household	1710
34	Acompañantes de juego y actividades primera infancia	Infancy	56
35	Educación indulgente o intransigente	Childhood, Life Cycle	467
36	Uso de discursos en la enseñanza	Childhood, Life Cycle	439, 440
37	Inculcación de autosuficiencia	Childhood, Life Cycle	308, 309
38	Inculcación de represión sexual	Life Cycle	333
39	Inculcación de agresividad	Life Cycle	300, 301
40	Ubicación de sueño de adolescentes	Household	1711
41	Patrón de asentamiento	Community Organisation, Housing	234

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
42	Tamaño medio de comunidades hermanas	Community Organisation, Population	235
43	Autoridad política extra-territorial	Community Organisation	237
44	Organización familiar	Settlement, Community	68
45	Forma de casa-habitación	Settlement, Community	65
46	Forma de techo (casa rectangular)	Community, Community Organisation, Housing	286
47	Forma de techo (casa semiabierta)	Community, Community Organisation, Housing	286
48	Forma de techo (tienda)	Community, Community Organisation, Housing	286
49	Forma de techo (otros)	Community, Community Organisation, Housing	286
50	Material de paredes y muros	Community, Community Organisation, Ecology, Housing	285
51	Material de techos	Community, Community Organisation, Ecology, Housing	287
52	Tipo de liderazgo	Settlement, Community	76

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
53	Sucesión de liderazgo	Settlement, Community, Leadership	77, 276, 277
54	Homicidio individual	Politics, Warfare	1665
55	Asalto individual	Warfare	1666
56	Homicidio grupal	Warfare	1675
57	Guerra interna	Politics, Warfare	1648, 1649
58	Guerra externa	Politics, Warfare	1648, 1650
59	Estratificación social	Modernisation	158
60	Modernización	Modernisation	1810, 1820
61	Existencia de matrimonio	-	-
62	Habitación de recién casados	Settlement, Marriage, Kinship	69, 213
63	Habitación de parejas establecidas	Marriage, Kinship	215
64	Sistema marital	Settlement, Marriage	68, 210
65	Relación entre esposas múltiples	Marriage	878
66	Habitación de esposas múltiples	Marriage	878
67	Matrimonio entre primos	Marriage	227
68	Permisividad de matrimonio entre primos en primer grado	Marriage	227, 229

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
69	Preferencia de matrimonio entre primos	Marriage	228, 230
70	Intercambio de pareja	Sexual practices	172
71	Actitud ante sexo premarital	Sexual practices	165, 166
72	Frecuencia de sexo premarital	Sexual practices	167
73	Actitud ante sexo extramarital	Sexual practices	169
74	Castigo por sexo extramarital: mujeres	Gender	663
75	Iniciaciones	Ritual	559, 545
76	Tema de iniciación	Ritual	559
77	Componentes cognitivos de iniciación	Ritual	545
78	Miedo al mal de ojo	Ritual, Religion	1188
79	Creencias sobre trance y posesión	Ritual	200
80	Clasificación de hermanos	Kinship	642
81	Clasificación de cuñados	Kinship	644
82	Pertenencia familiar	Kinship	836

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
83	Herederos de bienes muebles	Wealth transactions, Property	279
84	División entre herederos de bienes muebles	Wealth transactions, Property	281
85	Herederos de bienes inmuebles	Wealth transactions, Property	278
86	División entre herederos de bienes inmuebles	Property	280
87	Transacciones de matrimonio	Wealth transactions	208
88	Prevalencia de enfermedades infecciosas	Ecology, Health	1260, 1257
89	Animales domésticos	Ecology	244
90	Hambrunas e inseguridad alimentaria	Health	1263
91	Frecuencia de hambrunas	Health	1266
92	Tipos de juego	Games	239
93	Prevalencia de chisme	Politics, Kinship, Gossip	1805
94	Temas de chisme	Gossip	1790, 1801, 1782, 1794, 1786, 1795, 1781, 1792, 1789
95	Permiso de divorcio	Gender	663

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
96	Producción no doméstica por mujeres	Gender	663
97	Demanda de producción no doméstica por mujeres	Gender	663
98	Independencia financiera de mujeres	Gender	663
99	Participación política de mujeres	Gender	663
100	Grupos de solidaridad femenina	Gender	663
101	Mutilación genital de hombres	Gender	241
102	Escarificación	Ecology, Health	1692
103	División de clases	Class	270
104	División en castas	Class	272, 273
105	Presencia de religiones mayores	Religion, Death, Mourning	2002, 2001
106	Desarticulación de cuerpos	Religion, Death, Mourning	1851, 1853
107	Funerales secundarios	Religion, Death, Mourning	1850, 1852, 1856
108	Lengua nativa		

Continúa en la siguiente página

Cuadro 3: Variables Culturales

Pregunta	Tema	Categoría que influyó en su elección	Variable correspondiente
109	Lengua colonial		
110	Lengua de señas		
111	Presencia de lengua nativa vs. colonial		
112	Viaje tradicional a comunidad cercana		
113	Viaje motorizado a comunidad cercana		
114	Frecuencia de interacción con comunidad cercana		
115	Frecuencia de interacción con extranjeros		
116	Acceso a información		

Así, la lista final de variables culturales de la SCCS consideradas fue:

SCCS??877 SCCS??890

4.1 Análisis de Correspondencias Múltiples (ACM)

Para reducir la dimensionalidad del conjunto de datos e identificar las variables más significativas, empleamos el Análisis de Correspondencias Múltiples (ACM). El ACM es un método de reducción de la dimensionalidad que se adapta bien a los datos categóricos, por lo que resulta

adecuado para analizar el conjunto de datos de la SCCS, que está formado predominantemente por variables categóricas.

4.2 Selección de variables

Mediante el ACM, identificamos las variables más importantes que contribúan a la varianza del conjunto de datos. El proceso de selección consistió en clasificar las variables en función de sus cargas en las dimensiones principales extraídas mediante ACM. Las variables con mayores cargas se consideraron más influyentes a la hora de captar la diversidad y la relevancia culturales.

4.3 Desarrollo del cuestionario de la encuesta

A partir de los resultados del proceso de selección de variables, elaboramos un cuestionario de encuesta destinado a recopilar datos etnográficos actualizados de las sociedades de la SCCS. El cuestionario se diseñó para abarcar una amplia gama de aspectos culturales, incluidos temas tradicionales de la SCCS y temas emergentes de la antropología contemporánea.

4.4 Aislamiento de temas y refinamiento de preguntas

Para garantizar la exhaustividad y pertinencia del cuestionario de la encuesta, aislamos temas dentro del conjunto de datos de la SCCS y refinamos las preguntas de la encuesta en consecuencia. En este proceso se tuvo en cuenta el solapamiento entre categorías y se garantizó una representación equilibrada de las distintas dimensiones culturales.

4.5 Incorporación de temas antropológicos contemporáneos

Además de los temas tradicionales tratados en la SCCS, como la subsistencia, la economía y la organización social, el cuestionario de la encuesta incorporó temas antropológicos contemporáneos, como los géneros no binarios, la dinámica lingüística y el aislamiento geográfico. Estas adiciones reflejan los recientes avances en el discurso antropológico y proporcionan una comprensión más completa de la diversidad cultural.

4.6 Administración de la encuesta

El cuestionario final de la encuesta se administró a investigadores y antropólogos familiarizados con el conjunto de datos de la SCCS. Se solicitaron comentarios para garantizar la claridad, la no repetición y la pertinencia de las preguntas de la encuesta.

4.7 Recopilación y análisis de datos

Los datos recogidos mediante el cuestionario de la encuesta se analizaron utilizando métodos estadísticos para identificar patrones y tendencias en las distintas dimensiones culturales. Los resultados se utilizarán para actualizar y mejorar el conjunto de datos de la SCCS, lo que permitirá realizar análisis longitudinales y estudios comparativos.

4.8 Consideraciones éticas

Las consideraciones éticas fueron primordiales durante todo el proceso de investigación. El cuestionario de la encuesta se diseñó respetando las sensibilidades y tradiciones culturales de las sociedades de la SCCS, y se obtuvo el consentimiento informado de los participantes.

5 Bibliografía

- Andreski, Stanislav. 1972. «Social sciences as sorcery».
- Arnett, Jeffrey J. 2016. «The neglected 95%: why American psychology needs to become less American.»
- Dutton, Edward. s. f. «The Philosophy of Anthropology». *The Internet Encyclopedia of Philosophy*. Accedido 20 de abril de 2024. <https://iep.utm.edu/philosophy-of-anthropology/>.
- Ember, Carol R., y Jacob Kalodner. 2024. «Status and Role of the Elderly». *HRAF Explaining Human Culture*. <https://hraf.yale.edu/ehc/summaries/status-and-role-of-the-elderly>.
- Henrich, Joseph, Steven J Heine, y Ara Norenzayan. 2010. «The weirdest people in the world?» *Behavioral and brain sciences* 33 (2-3): 61-83.
- Josse, Julie, y François Husson. 2016. «missMDA: A Package for Handling Missing Values in Multivariate Data Analysis». *Journal of Statistical Software* 70 (1): 1-31. <https://doi.org/10.18637/jss.v070.i01>.
- Kirby, Kathryn R, Russell D Gray, Simon J Greenhill, Fiona M Jordan, Stephanie Gomes-Ng, Hans-Jörg Bibiko, Damián E Blasi, et al. 2016. «D-PLACE: A global database of cultural, linguistic and environmental diversity». *PloS one* 11 (7): e0158391.
- Latour, Bruno. 2004. «Why has critique run out of steam? From matters of fact to matters of concern». *Critical inquiry* 30 (2): 225-48.
- Lavit, Christine, Yves Escoufier, Robert Sabatier, y Pierre Traissac. 1994. «The act (statis method)». *Computational Statistics & Data Analysis* 18 (1): 97-119.
- Lê, Sébastien, Julie Josse, y François Husson. 2008. «FactoMineR: A Package for Multivariate Analysis». *Journal of Statistical Software* 25 (1): 1-18. <https://doi.org/10.18637/jss.v025.i01>.

- Mace, Ruth, Mark Pagel, John R Bowen, Biman Kumar Das Gupta, Keith F Otterbein, Mark Ridley, Thomas Schweizer, y Eckart Volland. 1994. «The comparative method in anthropology [and comments and reply]». *Current anthropology* 35 (5): 549-64.
- Murdock, George Peter. 1967. «Ethnographic atlas: a summary». *Ethnology* 6 (2): 109-236.
- Murdock, George P, y Douglas R White. 1969. «Standard cross-cultural sample». *Ethnology* 8 (4): 329-69.
- R Core Team. 2018. «R: A language and environment for statistical computing; 2018».
- Van Buuren, Stef. 2018. *Flexible imputation of missing data*. CRC press.
- Voillet, Valentin, Philippe Besse, Laurence Liaubet, Magali San Cristobal, y Ignacio González. 2016. «Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework». *BMC bioinformatics* 17: 1-16.