

X Congreso Chileno de Antropología. Colegio de Antropólogos de Chile A. G, Temuco, 2019.

Alcances y límites del empleo de la estadística moderna en la actual Antropología Sociocultural.

Miguel Ignacio Fernández Lizana.

Cita:

Miguel Ignacio Fernández Lizana (2019). *Alcances y límites del empleo de la estadística moderna en la actual Antropología Sociocultural*. X Congreso Chileno de Antropología. Colegio de Antropólogos de Chile A. G, Temuco.

Dirección estable: <https://www.aacademica.org/x.congreso.chileno.de.antropologia/22>

ARK: <https://n2t.net/ark:/13683/edE8/omk>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Simposio 11

Antropología al cuadrado: La antropología y los Estudios Sociales de la Ciencia y la Tecnología

Coordinación:

Honald Cancino - Héctor Mora

Alcances (y límites) del empleo de la Estadística moderna en la actual Antropología Sociocultural

Miguel Ignacio Fernández Lizana⁷⁵

Resumen: En la actualidad la Estadística se manifiesta como un conjunto de teorías, métodos y técnicas diversificadas que nos permiten analizar una ingente cantidad de datos de distinta naturaleza. No obstante, múltiples son los conceptos y operaciones que no (necesariamente) han sido asimilados por la mayoría de los antropólogos. Esta “no asimilación” ha hecho que muchas herramientas estadísticas efectivamente útiles sean omitidas. De esta forma, el objetivo principal de esta ponencia es presentar los alcances (y límites) del empleo de la Estadística moderna en la actual Antropología Sociocultural. Para concretar dicho objetivo se privilegiarán tres puntos que real o potencialmente pueden ser de interés para aquellos investigadores sociales adscritos a nuestra disciplina. Éstos son: 1. Análisis de datos categóricos (o cualitativos); 2. Análisis (cuantitativo) de textos; y 3. *Software* gratuito de análisis estadístico de datos.

Palabras clave: Estadística, Análisis de datos, *Software* gratuito, Antropología Sociocultural

⁷⁵ Antropólogo. Estudiante del Magíster en Antropología del Departamento de Antropología de la Universidad Católica de Temuco, Chile. Correo electrónico: miguel.ifl@gmail.com.

Introducción

Históricamente, la estadística no ha sido una completa desconocida para los antropólogos socioculturales. Sobre todo desde que el antropólogo británico Edward Tylor (1889) la introdujera en la antropología a finales del siglo XIX, mediante su artículo *On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent*, en el cual propuso la base estadística del método comparativo. En este sentido, tanto la cuantificación como el análisis estadístico de los datos fueron tempranamente introducidos en los trabajos realizados por diversos antropólogos (Fernández, 2016). Durante la primera mitad del siglo XX el uso de la estadística en antropología sociocultural, tal como lo evidencian Linton Freeman y Alan Merriam (1956) fue del tipo relacional. Investigaciones en las que los antropólogos trataban de determinar de forma “objetiva” (véase p. ej. Clements, Schenck y Brown, 1926) el grado de relación existente entre dos o más “variables culturales”, aplicando para ello técnicas de correlación (Freeman y Merriam, 1956). Tomando como referencia la configuración de los programas de estudio en antropología cultural en las universidades estadounidenses, Thompson y Roper (1980) sostienen que, hacia finales del decenio de 1950, el entrenamiento en métodos cuantitativos de investigación y en análisis estadístico de datos era prácticamente inexistente en dicho país. Y es que, si bien es cierto, algunos antropólogos empezaron a mostrar cierto interés por algunas técnicas estadísticas más sofisticadas a partir del período posterior a la Segunda Guerra Mundial (1939-1945), hasta el decenio de 1970 el análisis estadístico multivariante era poco empleado en la antropología sociocultural (Chibnik, 1985). En la actualidad, los antropólogos socioculturales usan técnicas que parecen no alejarse mucho de los clásicos procedimientos estadísticos univariados, bivariados y multivariados, como la estadística descriptiva y sus medidas de tendencia central, de dispersión, de posición y de forma, la prueba de independencia, el coeficiente de correlación de Pearson (r), el coeficiente de correlación por rangos de Spearman (ρ), la regresión lineal (simple y múltiple), el análisis factorial exploratorio (AFE), etc. (Thompson, 2013; Handwerker y Borgatti, 2015; Bernard, 2018). Lo anterior contrasta con lo que en la actualidad sucede en otras disciplinas de las ciencias sociales: a día de hoy, algunos investigadores sociales emplean herramientas estadísticas modernas como las correlaciones policóricas y tetracóricas, el análisis factorial de ítems (IFA), los modelos de ecuaciones estructurales lineales y no lineales, el modelo contrafactual de causalidad (también conocido como “modelo de Neyman-Rubin” o de “resultados potenciales”), las regresiones discontinuas, los análisis de datos longitudinales, los modelos multinivel, la estadística bayesiana, las simulaciones de Montecarlo, etc. (Rodríguez y Asún, 2016; Hernán y Robins, 2018; Agresti, 2018, 2019). Y si a estas técnicas sumamos los nuevos desarrollos en -por ejemplo- *data mining* y *machine learning* (Foster, Ghani, Jarmin, Kreuter y Lane, 2016), nos encontramos frente a un panorama bastante complejo con el cual la antropología sociocultural pareciese estar desvinculada.

Así, son múltiples los conceptos y operaciones que no (necesariamente) han sido asimilados por la mayoría de los antropólogos. Esta “no asimilación” ha hecho que muchas herramientas estadísticas efectivamente útiles sean omitidas. De esta forma, el objetivo principal de este trabajo es presentar los alcances (y límites) del empleo de la Estadística moderna en la actual Antropología Sociocultural. Para operacionalizar lo anterior aquí se presentan tres elementos que real o potencialmente pueden ser de interés para aquellos investigadores sociales adscritos a nuestra disciplina. Éstos son: 1. Análisis de datos categóricos (o cualitativos); 2. Análisis (cuantitativo) de textos; y 3. *Software* gratuito de análisis estadístico de datos.

Sobre categorías, textos y *software*

1. Análisis de datos categóricos (o cualitativos):

Una de las deficiencias de la clásica estadística paramétrica es que ésta, parte de supuestos teóricos que, en la práctica, difícilmente se adaptan a los datos que comúnmente son analizados por los antropólogos (normalidad, homocedasticidad, etc.). En este sentido, para nadie es un secreto que en el campo de las ciencias sociales y, particularmente, en el ámbito de la antropología sociocultural, se analizan casi exclusivamente datos categóricos (o cualitativos). Comúnmente, las variables categóricas son comprendidas como todas aquellas variables no métricas que expresan un conjunto de cualidades o características no directamente medibles (Agresti, 2019). Un ejemplo de variable cualitativa sería “preferencia musical” (cuyas categorías podrían ser: rock, jazz, blues, música clásica, etc.). Por otro lado, y mediante un proceso de codificación numérica, las variables categóricas efectivamente pueden ser analizadas estadísticamente: así, se pueden realizar desde simples recuentos y gráficos de barras, pasando por modelos de respuesta discreta y análisis logarítmico-lineales, hasta complejos análisis bayesianos para datos categóricos multivariados (Rudas, 2018; Upton, 2017; Agresti, 2019). Una gran ventaja de las técnicas de análisis de datos categóricos es que varias de éstas parten del análisis de tablas de contingencia (bidimensionales o multidimensionales), como el caso de los análisis de correspondencias y los análisis logarítmico-lineales. Como es bien sabido, las tablas de contingencia, dadas sus ventajas, son ampliamente empleadas en el análisis de datos sociales, debido a que mediante éstas se pueden analizar conjuntamente dos o más variables categóricas vertidas en filas y columnas.

Cuadro N°1.

Ejemplos de técnicas estadísticas multivariantes para datos categóricos

Tipos de análisis	Tipos de variables	Técnicas estadísticas multivariantes para datos categóricos
Análisis de relaciones de interdependencia V - V	Variables categóricas	Análisis de tablas de contingencia multidimensionales
		Análisis log-lineal
		Análisis de correspondencias
		Análisis de clasificación
		Análisis de escalamiento multidimensional no métrico
Análisis de relaciones de dependencia. VI ? VD	Variables independientes categóricas y variables dependientes categóricas (dicotómicas, politómicas u ordinales)	Regresión logística binaria
		Regresión logística multinomial
		Regresión logística ordinal

Fuente: Elaboración propia con base en López-Roldán y Fachelli (2015).

Nota: VI = Variable(s) independiente(s). VD = Variable(s) dependiente(s).

2. Análisis (cuantitativo) de textos:

En sus múltiples variaciones y, a pesar de lo que a priori se pudiera pensar, el análisis de textos se erige como un campo abierto e intensamente interdisciplinario (Santander, 2011; Sayago, 2014; Levshina, 2015; Mizumoto y Plonsky, 2015; Hernández, 2016; Welbers, Van Atteveldt y Benoit, 2017), en el cual se encuentran disciplinas tan diversas como la lingüística, las comunicaciones, el periodismo, la antropología, la psicología social, la sociología, las ciencias políticas, la estadística, las ciencias de la computación, etc., y en el que convergen aproximaciones cualitativas y cuantitativas junto con sus correspondientes y misceláneas bases epistemológicas y teóricas (Andréu, 2002). Afortunadamente para los investigadores en el campo de la antropología sociocultural en la actualidad existen variados *software* y paquetes especializados en el análisis (cuantitativo) de textos (véase p. ej. Heiden, Magué y Pincemin, 2010; Ratinaud, 2014; Welbers, Van Atteveldt y Benoit, 2017; Bouchet-Valat y Bastin, 2018; Feinerer y Hornik, 2018), los cuales se constituyen en potentes plataformas que permiten de forma holgada superar la añosa e incómoda dicotomía cualitativo-cuantitativo, la cual, a pesar de las amplias discusiones que se han dado a lo largo de las décadas, sigue siendo reproducida por algunos autores o, peor aún, sigue presentándose bajo el pretérito supuesto que indica que las investigaciones de base cuantitativa son científicamente más robustas en comparación con las investigaciones de tipo cualitativo (véase p. ej. el libro de: Llanos, 2009)⁷⁶. En este sentido, el análisis de textos permite fomentar el diálogo cualitativo-cuantitativo en una disciplina internamente diversa como la antropología sociocultural (Restrepo, 2016), disciplina que, por lo demás, históricamente ha promovido la interdisciplinariedad, lo multitécnico y la reflexividad en un ambiente ligado con la vigilancia de los procesos de sus investigaciones empíricas y de sus correspondientes construcciones teóricas (Apud, 2013).

3. Software gratuito de análisis estadístico de datos:

En la actualidad resulta difícil concebir que un investigador del área de las ciencias sociales realice análisis estadísticos de datos sin la ayuda de un *software* especializado. En efecto, el uso de este tipo de programas tiene considerables ventajas con respecto al cálculo manual. En términos generales, un *software* de análisis estadístico hace alusión a todos aquellos programas informáticos que disponen de módulos orientados a la tabulación, gestión, modificación, análisis y representación gráfica de datos (entre otras funciones). Por otro lado, a menudo se hace la distinción entre los *software* de pago y los *software* gratuitos de análisis estadístico de datos. Entre los variados *software* de pago actualmente existentes en el mercado nos podemos encontrar, entre otros, con IBM SPSS Statistics, Stata, SAS & JMP, Minitab, Statgraphics, STATISTICA, SYSTAT, SPAD, XLStat, etc. Sin embargo, también existe la opción de acceder a *software* gratuito de análisis estadístico de datos, los cuales, en muchos casos, resultan ser muy superiores a los *software* de pago en prácticamente todos los ámbitos.

⁷⁶ En su libro *Epistemología de las Ciencias Sociales* el filósofo peruano Marino Llanos (2009) básicamente asume que sólo las investigaciones de base cuantitativa son científicamente válidas y que éstas se encuentran en las antípodas de lo subjetivo, lo poco o anti-científico y lo no-estandarizado, características que según este autor ostentarían las investigaciones de tipo cualitativo. Por supuesto, el autor del presente trabajo rechaza tales afirmaciones al considerar que Llanos desconoce el desarrollo histórico de las ciencias sociales y las variabilidades presentes en estas ciencias y, más aún, dentro de cada una de aquellas ciencias (antropología, psicología social, sociología, etc.).

Cuadro N°2.**Ejemplos de software gratuito de análisis estadístico de datos**

Software	Página web
Lenguaje de Programación R	https://www.r-project.org/
PSPP	https://www.gnu.org/software/pspp/
JASP	https://jasp-stats.org/
SAS University Edition	https://www.sas.com/en_us/software/university-edition/download-software.html
Past	https://folk.uio.no/ohammer/past/
InfoStat Versión estudiantil	http://www.infostat.com.ar/index.php?mod=page&id=37
Gretl	http://gretl.sourceforge.net/
MaxStat Lite Version	https://maxstat.de/en/home-en/
OpenStat	http://openstat.info/

Fuente: Elaboración propia.

Ahora bien, de entre los variados *software* gratuitos actualmente disponibles destaca uno que, por su cada vez mayor aceptación y promoción en los ámbitos académico e investigativo, se ha convertido en un potente referente en lo que respecta a la computación estadística de alto nivel como apoyo a las más diversas disciplinas científicas; nos referimos a R. Como tal R es un lenguaje de programación empleado primordialmente para efectuar análisis estadístico de datos y construcción de gráficos (R Core Team, 2018). Actualmente R es considerado la *lengua franca* de la estadística (Mizumoto y Plonsky, 2015) debido a algunas de sus características que lo sitúan muy por encima de prácticamente todos sus “competidores”: R es gratuito y libre, es muy versátil, permite realizar una cantidad insospechable de procedimientos estadísticos y gráficos, permite construir gráficos de calidad inmejorable, etc. Cabe destacar que la potencia de R reside en su calidad de lenguaje. En este sentido, con R se pueden realizar cosas que van mucho más allá del análisis estadístico de datos y construcción de gráficos, como por ejemplo: (1) Analizar y/o editar imágenes. (2) Realizar análisis de sonido. (3) Analizar textos (cuantitativa y cualitativamente). (4) Generar mapas. (5) Realizar análisis de redes sociales (ARS). (6) Extraer, analizar y visualizar datos de páginas web (como Facebook, Twitter, Instagram, etc.). (7) Escribir textos. (8) Crear gráficos interactivos y animaciones. (9) ¡Enviar correos electrónicos! (10) Inclusive, con R se pueden crear piezas de arte abstracto o, por lo menos, simularlas⁷⁷.

⁷⁷ En torno a esta temática se pueden consultar las siguientes páginas: 1. Simulating Abstract Art with R (<https://www.r-bloggers.com/simulating-abstract-art-with-r/>) 2. Data Art: The R graph Gallery (<http://www.r-graph-gallery.com/portfolio/data-art/>) 3. Reproducible art with R (<https://www.r-bloggers.com/reproducible-art-with-r/>).

Conclusiones

La estadística moderna ostenta una multiplicidad de conceptos y operaciones que no (necesariamente) han sido asimilados por la mayoría de los antropólogos. Esta “no asimilación” ha hecho que muchas herramientas estadísticas efectivamente útiles sean omitidas y/o de plano rechazadas⁷⁸. Según este trabajo aquellas herramientas que provienen de la estadística y que son real o potencialmente útiles en el campo de la antropología sociocultural guardan correspondencia principalmente con las técnicas de análisis para datos categóricos y el análisis (cuantitativo) de textos. Por otro lado, y si bien es cierto en este trabajo no se abordaron los malos entendidos en torno a la estadística y al empleo poco razonado de algunos conceptos y operaciones provenientes de aquella disciplina (Wasserstein y Lazar, 2016), es necesario recalcar aquí que dicha ciencia no debe ser tomada nunca “a la ligera” (Silva, 1997; Senn, 2008). Y es que, querámoslo reconocer o no, una de las grandes problemáticas en el marco de la enseñanza de la estadística en la antropología (y en general en las ciencias sociales) es que ésta -la estadística- es enseñada muchas veces como una “caja negra” en nuestra disciplina, es decir, se les dice a los estudiantes cosas como las siguientes: *introduzcan las variables X, Y y Z en SPSS y esperen que éste les arroje los resultados esperados y si es que dicho software les indica un p valor significativo ($p < 0.05$) dense por satisfechos*. Con la analogía de la caja negra nos estamos refiriendo a que efectivamente se les plantea a los estudiantes qué procedimientos emplear al estar manipulando determinadas variables, pero no necesariamente se les enseña cómo funcionan en realidad dichos procedimientos, y esta caja negra también se ve potenciada al hacer un uso indiscriminado de software estadístico como, precisamente, SPSS, el cual efectivamente opera como una caja negra en la cual se van introduciendo variables y se espera que aquel software arroje por otro lado los resultados que más nos acomoden, pero no tenemos acceso al conocimiento teórico-algorítmico que subyace a los procedimientos estadísticos realizados por dicho software. En este sentido, no está de más considerar que para lograr un uso adecuado -y por ende efectivo- de los múltiples *software* de análisis estadístico de datos, se requiere del despliegue de conocimientos estadísticos que permitan efectuar de manera correcta los análisis e interpretaciones que potencialmente deban ser llevados a cabo. Igualmente, es importante reconocer que la estadística no es una *panacea* que “mágicamente” mejore argumentos débiles y/o articulaciones teóricas pobres. La producción acrítica de números y, peor aún, el mal análisis de éstos, jamás podrán sustituir razonamientos fundados en la coherencia, la rigurosidad y la sistematicidad que todo antropólogo sociocultural debe ser capaz de desplegar en el campo científico. Asimismo, debemos recordar siempre que “significancia estadística” no es sinónimo de “relevancia científica” (no debemos nunca confundir ambos elementos).

78 En otro trabajo ya se han presentado los motivos de esta “no asimilación” o “relegación” (Fernández, 2016).

Referencias bibliográficas

- Agresti, A. (2018). *Statistical Methods for the Social Sciences*. 5th Ed. Pearson.
- Agresti, A. (2019). *An Introduction to Categorical Data Analysis*, 3rd Ed. Hoboken, NJ: John Wiley & Sons.
- Andréu, J. (2002). Las técnicas de análisis de contenido: Una revisión actualizada. Sevilla: Fundación Centro de Estudios Andaluces.
- Apud, I. (2013). Repensar el método etnográfico: Hacia una etnografía multitécnica, reflexiva y abierta al diálogo interdisciplinario. *Antípoda. Revista de Antropología y Arqueología*, (16), 213-235.
- Bernard, H.R. (2018). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, 6th Ed. Lanham, MD: Rowman & Littlefield.
- Bouchet-Valat, M. & Bastin, G. (2018). *RcmdrPlugin.temis: Graphical Integrated Text Mining Solution*. R package version 0.7.10.
- Chibnik, M. (1985). The Use of Statistics in Sociocultural Anthropology. *Annual Review of Anthropology*, 14, 135-57.
- Clements, F., Schenck, S. & Brown, T.K. (1926). A New Objective Method for Showing Special Relationships. *American Anthropologist*, 28(4), 585-604.
- Feinerer, I. & Hornik, K. (2018). *tm: Text Mining Package*. R package version 0.7-5.
- Fernández, M. (2016). Caracterización del rol de la Estadística en la construcción histórica de los métodos cuantitativos de investigación en Antropología Sociocultural. En J, Gibert. (Coord.). *Metodología y Epistemología en las Ciencias Sociales*. Simposio llevado a cabo en el 9º Congreso Chileno de Sociología, Universidad Católica del Maule, Talca, Chile.
- Freeman, L. & Merriam, A. (1956). Statistical Classification in Anthropology: An Application to Ethnomusicology. *American Anthropologist*, 58, 464-472.
- Foster, I., Ghani, R., Jarmin, R., Kreuter, F. & Lane, J. (2016). *Big Data and Social Science: A Practical Guide to Methods and Tools*, 1st Ed. Boca Raton, FL: Chapman & Hall/CRC Press.
- Handwerker, W. & Borgatti, S. (2015). "Reasoning with Numbers". En H.R, Bernard. & C, Gravlee. (Eds.). *Handbook of Methods in Cultural Anthropology* (pp. 519-531). Lanham, MD: Rowman & Littlefield.
- Heiden, S., Magué, J-P. & Pincemin, B. (2010). "TXM: Une plateforme logicielle open-source pour la textométrie - conception et développement". En S, Bolasco, I, Chiari. & L, Giuliano. (Ed.). *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010, Vol. 2* (pp. 1021-1032). Roma: Edizioni Universitarie di Lettere Economia Diritto.
- Hernández, D. (2016). *Análisis estadístico de corpus cronológicos: Aplicación al estudio de bases bibliográficas y textos retóricos* (Tesis para optar por el grado de Doctor en Estadística e Investigación Operativa). Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, España.
- Hernán, M.A. & Robins, J.M. (2018). *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC, forthcoming.
- Llanos, M. (2009). *Epistemología de las Ciencias Sociales*. Lima: Fondo Editorial de la Universidad Nacional Mayor de San Marcos.

- Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.
- López-Roldán, P. & Fachelli, S. (2015). Capítulo III.5: "Clasificación de las técnicas de análisis de datos". En P, López-Roldán. & S, Fachelli. *Metodología de la Investigación Social Cuantitativa, 1ª edición*. Bellaterra: Dipòsit Digital de Documents, Universitat Autònoma de Barcelona.
- Mizumoto, A. & Plonsky, L. (2015). R as a Lingua Franca: Advantages of Using R for Quantitative Research in Applied Linguistics. *Applied Linguistics*, 37(2), 284-291.
- Ratinaud, P. (2014). *Iramuteq. Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres*. Version 0.7 alpha 2.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Restrepo, E. (2016). Antropologías disidentes y sentido común antropológico. *Boletín del Colegio de Etnólogos y Antropólogos Sociales*, 51-57.
- Rodríguez, K. & Asún, R. (2016). Desarrollos recientes en Estadística: Aportes teórico-metodológicos a la investigación sociológica. *Sociología y Tecnociencia/Sociology and Technoscience*, 1(6), 1-13.
- Rudas, T. (2018). *Lectures on Categorical Data Analysis*. eBook: Springer.
- Santander, P. (2011). Por qué y cómo hacer análisis de discurso. *Cinta de Moebio*, (41), 207-224.
- Sayago, S. (2014). The construction of qualitative and quantitative data using discourse analysis as a research technique. *Quality & Quantity*, 49(2), 727-737.
- Senn, S. (2008). *Statistical Issues in Drug Development*, 2nd Ed. West Sussex: Wiley.
- Silva, L. (1997). *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Madrid: Díaz de Santos.
- Thompson, R. & Roper, R. (1980). Methods in Social Anthropology: New Directions and Old Problems. *American Behavioral Scientist*, 23(6), 905-924.
- Thompson, B. (2013). Overview of Traditional/Classical Statistical Approaches. *The Oxford Handbook of Quantitative Methods. Volume 2: Statistical Analysis* (pp. 7-25). New York, NY: Oxford University Press.
- Tylor, E. (1889). On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 18, 245-272.
- Upton, G. (2017). *Categorical Data Analysis by Example*. Hoboken, NJ: John Wiley & Sons.
- Wasserstein, R. & Lazar, N. (2016). The ASA's Statement on *p*-Values: Context, Process, and Purpose. *The American Statistician*, 70, 129-133.
- Welbers, K., Van Atteveldt, W. & Benoit, K. (2017). Text analysis in R. *Communications Methods and Measures*, 11(4), 245-265.