



Rasinger, S. M. (2019). *La investigación cuantitativa en Lingüística. Una introducción. Traducción de Ana Useros Martín y Gema Sanz Espinar*. Madrid: Ediciones Akal (2<sup>a</sup> edición). ISBN: 978-84-460-4644-8. 288 páginas.

Miguel Ángel Bargetto Fernández

UNIVERSIDAD DE LAS AMÉRICAS

CHILE

mbargetto@udla.cl

DOI: 10.4067/S0718-09342022000100199

El Dr. Sebastian M. Rasinger es profesor de Lingüística aplicada en Anglia Ruskin University, en Cambridge, Inglaterra. Su principal línea de trabajo es lengua y migración. Tiene una amplia experiencia en investigación, pues forma parte de los comités de revistas como Asian EFL Journal y Journal of Research Design and Statistics in Linguistics and Communication Science. La obra es la traducción de la segunda edición del original, publicado en 2013.

El libro se presenta como un manual específico para la investigación cuantitativa en Lingüística, dividida en tres partes y once capítulos. En la primera parte, formada por tres capítulos, se efectúa una revisión sobre aspectos fundamentales de metodología de investigación, con definiciones básicas sobre diseño y formulación de cuestionarios, fácilmente accesibles en otros manuales de metodología, pero, en este caso, con la ventaja de una aplicación práctica en Lingüística. En la segunda parte, constituida por cinco capítulos, se revisa el análisis de datos, en un nivel creciente de profundidad que va desde la frecuencia hasta los estadísticos no paramétricos; la tercera parte, de solo dos capítulos, trata sobre ANOVA, meta-análisis y software para análisis estadístico y finaliza con una serie de tablas de datos críticos para los estadísticos más empleados en investigación lingüística.

En el capítulo sobre temas básicos de investigación cuantitativa, el autor revisa los principales conceptos que un investigador inicial debe tener en cuenta al momento de formular las ideas de sus proyectos de trabajo: conceptos de datos -continuos y discretos, variables; fiabilidad y validez, hipótesis, leyes y teorías. En el caso de los diseños de investigación, también se explican los conceptos con ejemplos de trabajos propios de la línea de investigación del autor y también con ejemplos básicos de investigación lingüística, tales como evolución fonológica o de adquisición de la lengua; asimismo, también se citan trabajos clásicos. El capítulo sobre ética es breve y limitado a recomendaciones prácticas sobre qué aspectos se deben tener en cuenta al momento de planificar un trabajo de investigación, especialmente respecto de la participación de personas, así como algunos dilemas en la investigación lingüística, tales como la grabación de las intervenciones de los participantes, por el efecto que puedan producir en las respuestas. Con todo, se omite la Declaración de Singapur, del año 2010, que establece un marco de referencia general y aceptado por la comunidad científica y que debe ser de conocimiento de investigadores novatos. En esta primera parte, se revisa, también, la recopilación de datos, con especial énfasis en el cuestionario, lo que restringe la investigación en otras disciplinas, específicamente las experimentales, como la Psicolingüística. A esta objeción, se suma que tampoco hay alusión a consideraciones sobre validez y confiabilidad que considerarse en la creación de ellos. La parte finaliza con la recomendación de varias referencias específicas sobre los distintos capítulos del apartado.

La segunda parte del libro comienza con la explicación del tratamiento de los datos y conceptos elementales en matemática, tales como sumas y frecuencias. Cabe destacar el empleo de ejemplos lingüísticos y ejercicios y problemas para resolver. Seguidamente, el apartado sobre codificación de los datos entrega orientaciones claras para aspectos que pudieran resultar dificultosos para quienes no cuentan con experiencia en el uso de programas informáticos para el manejo de datos. Incluso, los ejercicios propuestos en Excel facilitan la práctica de procesos elementales, como la ordenación y la tabulación, no obstante las posibilidades de la herramienta para un desarrollo exhaustivo del tratamiento de datos no es satisfactoria. En cuanto a las recomendaciones para la construcción de gráficos, orienta al investigador novato en forma de representar visualmente los datos obtenidos; en el caso de la investigación básica, los gráficos de barra, línea y dispersión están acompañados de variados ejemplos y ejercicios, sin embargo, no se ahonda en la explicación para su obtención.

En el capítulo sexto, se transita desde la estadística descriptiva hacia la inferencial. Primeramente, se expone sobre lo insuficiente que es el análisis de estadística descriptiva y releva el papel de la inferencia estadística como el conjunto de herramientas que permite al investigador efectuar generalizaciones y a establecer relaciones causales. Frente a esto, conviene destacar que la afirmación es incorrecta, puesto que la estadística descriptiva y la inferencial atienden distintos enfoques de

análisis. En el capítulo, se desarrollan temáticas estadísticas de relevantes para la investigación lingüística: la asimetría, la distribución asimétrica, los conceptos de cuartiles, desviación, distribución y error, entre otras. En el ejemplo de la organización de datos en cuartiles, quintiles y deciles, el autor señala que “cuando se revisa la bibliografía sobre estadística, no se encuentra un consenso acerca de cómo calcular los cuartiles” (Rasinger, 2019: 148). Sin embargo, la estadística descriptiva no es ambigua al momento de formular sus definiciones y conceptos, lo que convierte la afirmación del autor en una imprecisión conceptual seria y que debería ser rectificada.

En cuanto a la dispersión, el autor ocupa varias páginas para explicar los conceptos matemáticos subyacentes, tales como el rango, varianza, desviación típica, la distribución normal y el error estándar, igualmente con ejemplos y ejercicios para ejecutar en Excel, con un lenguaje especialmente claro. Estos conceptos, ampliamente utilizados en la investigación lingüística, requieren ser no solo aplicados correctamente, sino comprendidos en su utilidad matemática, lo que el autor logra ejemplificando con temáticas disciplinares limitadas a su ámbito de trabajo. Sin embargo, la limitación del cálculo a Excel vuelve insuficiente el análisis, puesto que hay elementos, como el proceso de elaboración de gráficos, que no son explicados (sustancialmente por la dificultad de obtenerlos a través de este programa) y que olvidan otras herramientas informáticas que facilitan el análisis cuantitativo.

El séptimo capítulo, dedicado a análisis de datos avanzado, aborda estadísticos que evalúan la correlación y el efecto de una variable sobre otra, los que son explicados con una adecuada claridad, pues su complejidad es mayor para quien no tenga formación matemática; de esta forma, se ahonda en varias páginas el concepto de ‘probabilidad’, con varios ejercicios, pero sin aplicarlos a programas informáticos que faciliten la tarea, sino limitados a las fórmulas para un cálculo manual. En el mismo capítulo, los estadísticos ‘chi cuadrado’, correlación de Pearson, la regresión lineal simple y múltiple son explicados, aunque no con el detalle de la probabilidad ni tampoco en aplicaciones prácticas en herramientas informáticas que permitan obtener los estadísticos y graficarlos como el mismo autor lo hace en el texto. Si bien enseña a interpretar correctamente los resultados que arrojan y a comunicar los resultados en la producción científica, vale mencionar que, en estos casos, el autor advierte del cuidado que se debe observar al momento de utilizar Excel, pues es un programa de hojas de cálculo y no un programa estadístico, por lo que recomienda el empleo de SPSS o R, aun cuando no hay explicaciones sobre cómo efectuarlos

El capítulo 8 está dedicado a la comprobación de hipótesis. Primeramente, el autor explica tres conceptos estadísticos necesarios para entender la inferencia estadística: la hipótesis, la causalidad y las colas. La causalidad se constituye en una profundización respecto de la ‘asociación’ que el texto desarrolla en capítulos previos. Consecuentemente con lo anterior, el punto de partida para establecer estas relaciones

causales se inicia en la formulación de hipótesis, específicamente la ‘hipótesis nula’. El autor hace especial hincapié en la diferencia estadística entre la comprobación de la hipótesis y las mediciones del capítulo anterior referido al diseño experimental, poniendo como caso el trabajo de Sadighi y Zare (2006) para argumentar sobre la manipulación de variables en experimentos reales. También ocupa un espacio importante en el capítulo la ‘prueba t’ y la ‘prueba F’ por sus implicancias estadísticas. La primera, utilizada para determinar si la media aritmética entre dos grupos presenta diferencias reales considerando la variabilidad (manifestada en la desviación típica o varianza) en datos con distribución normal y con un mínimo de 10 datos por grupo. En el caso de la prueba F, la explicación de su empleo está en la utilidad de definir si las varianzas de los grupos son iguales o no, puesto que el cálculo de la prueba t en Excel requiere de tal definición. El mismo autor adapta algunos datos del experimento de Sadighi y Zare (2006) para describir los pasos del cálculo en Excel y explicar los resultados que devuelve el programa computacional, especialmente el valor  $p$  que permitirá decidir el tipo de prueba t que se aplicará: la prueba t dependiente y la prueba t independiente. En ambos estadísticos, el autor sigue su didáctica, explicando los conceptos matemáticos y ejercitando con datos ficticios ya sea preparando los cálculos o graficando los resultados de los mismos. Al final del capítulo octavo, el autor dedica varias páginas al análisis de varianza (ANOVA) de uno y de dos factores, partiendo por la condición fundamental: que los conjuntos de datos relacionados sean por lo menos tres y que cuenten con normalidad en su distribución. Junto con explicar las ventajas de efectuar este análisis por sobre sucesivas pruebas t (sustancialmente para no cometer un error estadístico de tipo 1), el autor enseña a ejecutar el análisis de un factor en Excel y a interpretar adecuadamente las tablas de resultados que arroja el cálculo, de modo tal que para que exista diferencia estadística entre los grupos, el valor  $p$  debe ser mayor a 0,05 y el valor F debe ser superior al valor crítico para F. En el caso del ANOVA de dos factores, el punto de partida es la incidencia de dos variables independientes. Para este caso, toma el trabajo de Doughty y Pica (1986), y adapta ciertos datos para que pueda ser calculado en Excel, al mismo tiempo que advierte la dificultad que conlleva aquello; en cuanto a la interpretación de los resultados, el valor de mayor relevancia es la ‘interacción’, pues su valor  $p$  y su valor crítico para F entregarán el juicio estadístico para verificar si las dos variables independientes juntas producen efectos sobre la variable dependiente.

El capítulo noveno se aboca al análisis de datos no paramétricos. La primera prueba descrita por el autor es la correlación de Spearman de la cual el autor ofrece un ejercicio para calcular una  $r$  con datos no normales y de esa manera, el lector -el investigador lingüista-, podrá verificar que la curva no responde a lo esperable para datos normales. En el caso del estadístico rho de Spearman, por otra parte, el empleo de planillas de cálculo es mucho más complejo, básicamente por la inclusión de fórmulas, que lo vuelve innecesariamente extenso. La siguiente prueba es la *tau de Kendall*, utilizada para la descripción de relaciones en datos no paramétricos o no

normales y que emplea grados en vez de valores reales; la descripción de este estadístico es más breve y no tan profunda como en los otros casos. Lo mismo sucede, seguidamente, con la prueba de los rangos de Wilcoxon, la equivalente a la prueba  $t$  para los datos no normales; al igual que la *tau de Kendall*, esta prueba no se puede ejecutar directamente en Excel, por lo que el autor expone paso a paso el desarrollo de la prueba a partir de la adaptación de los datos del trabajo de Keim (1984), para que, una vez finalizados los cálculos del ejercicio, el lector emplee las tablas estadísticas del apéndice para verificar el valor  $p$ . El último estadístico del capítulo es la Prueba U de Mann-Whitney, equivalente a la prueba  $t$  de dos factores que es explicada de forma similar a la anterior. Al final del capítulo, el autor recomienda cinco obras sobre estadística aplicada a la lingüística, algunas incluso de la década del 80 del siglo pasado. Sin embargo, podría haberse citado otras obras que incorporan el empleo de R, tal como Gries (2009) o bien, en futuras ediciones, incorporar obras recientes que incorporan la estadística y las herramientas computacionales, como la de Eddington (2016).

En la tercera parte, el autor avanza hacia análisis avanzados y deja de lado las herramientas estadísticas de Excel. El primer apartado es la interpretación de MANOVA, a partir de la explicación funcional del estadístico: su empleo cuando se está en presencia de varias variables que inciden en el problema estudiado. Seguidamente, el autor coloca ejemplos para comprender la dimensión del análisis, sin ejercitarlo, pues señala que la explicación profunda del estadístico va más allá del objetivo del libro. La segunda parte de capítulo está destinada al metaanálisis de datos, aplicado específicamente al trabajo de los artículos de revisión; en ese sentido, la referencia a la investigación de Hunter y Schmidt (2004) aclara bastante la metodología que se debe aplicar al momento de abordar este tipo de trabajo, especialmente con la secuencia de fórmulas matemáticas para verificar si las fuentes consultadas son realmente consistentes entre sí. Por lo mismo, una de las recomendaciones del autor es contar con dobles revisores, tanto para la bibliografía como para la valoración de los datos.

Para finalizar el libro, el autor incorpora una serie de tablas con los valores críticos para cada uno de los estadísticos revisados a lo largo del libro, con diferentes niveles de confianza y graduados en diferentes grados de libertad. También se incluyen las soluciones a los ejercicios propuestos a lo largo de los capítulos del libro.

En general, el trabajo del Dr. Rasinger es, como el subtítulo de su obra dice, una introducción que puede ser de gran utilidad para estudiantes de lingüística que se inician en la investigación. En el caso de los profesores, el libro se constituye en un manual práctico para orientar la planificación de la enseñanza de métodos y herramientas de investigación y análisis, pero insuficiente incluso en trabajos iniciales.

En cuanto a las debilidades del manual, hay importantes omisiones respecto de la tradición investigativa en lingüística cuantitativa, tales como Zipf, con sus trabajos estadísticos para establecer rigurosamente los límites de las frecuencias léxicas; Markov y los modelos aplicados a la gramática generativa; Mandelbrot y los fractales aplicados al cálculo de error en los corpus; Herdan y la lingüística cuantitativa, Menzerath y la formulación de la ley de aumento de constructos. Por lo mismo, la principal debilidad del libro es que está circunscrito a los estudios cuantitativos en sociolingüística, hecho sobre el cual el mismo autor repara. De ahí que un título como el del libro pudiera ser engañoso, ya que no se da cuenta de todas las ramas de la lingüística que podrían valerse de los métodos expuestos, ya sea porque son básicos y no contribuyen al desarrollo disciplinar o bien porque no pueden ser aplicados pues se requiere de estadísticos y análisis que dan cuenta de variables más complejas.

Por otra parte, el empleo casi exclusivo de Excel como software estadístico es otra debilidad seria del libro: el autor supone que hay un grupo importante de investigadores que prefiere el cálculo estadístico realizado de forma manual, que es facilitado por este recurso, pero no repara en los estadísticos de mayor complejidad, para los que hay programas más poderosos y especialmente elaborados. Sin duda que dichos empleos exigen la conformación de equipos interdisciplinarios que trabajen colaborativamente en el análisis de datos. En la misma línea, las recomendaciones bibliográficas señaladas para el trabajo con SPSS son escasas y limitadas a un campo de estudio. En cuanto al empleo de R, la explicación está circunscrita a lo obvio, entrega un par de ejemplos con cálculos básicos, tales como sumar o agrupar variables y omite claves de uso de la herramienta, como *Rcommander*, que simplifica el cálculo para quienes no dominan la programación. Respecto de la presentación de los resultados, también debe ser considerado como una debilidad importante la construcción de representación gráfica de los resultados, pues no basta con señalar los valores estadísticos en la investigación, sino que la construcción de tablas y gráficos cobra vital importancia. En Excel, dicha representación es más bien limitada, mientras que en R las opciones son mucho más variadas, dada su especificidad para la estadística.

Finalmente, lo valorable del libro consiste en explicaciones claras, didácticas y metódicas de conceptos investigativos y estadísticos que, acompañados de ejemplos – aunque limitados a la Sociolingüística– y de ejercicios para Excel, son de utilidad al principiante en la investigación lingüística, pues permite una fácil comprensión de la naturaleza y la utilidad de los cálculos. Por lo tanto, un volumen como este es recomendable para cursos introductorios de ciencias del lenguaje, tal como reza su subtítulo.

## REFERENCIAS BIBLIOGRÁFICAS

- Doughty, C. & Pica, T. (1986). Information gap tasks: Do they facilitate second language acquisition? *TESOL Quarterly*, 20(2), 305-325.
- Eddington, D. (2016). *Statistics for linguists: A step-by-step guide for novices*. Cambridge: Cambridge Scholars Publishing.
- Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction* (Vol. 208). Londres: Walter de Gruyter.
- Hunter, J. & Schmidt, L. (2004) *Methods of meta-analysis: Correcting errors and bias in research findings*. Londres: Sage.
- Keim, I. (1984). *Untersuchungen zum Deutsch türkischer Arbeiter*. Tübingen: Gunten Narr.
- Sadighi, F. & Zare, S. (2006). Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners. *The Linguistics Journal*, 1, 110-126.