

PEDRO SANTANA MARTÍNEZ*
Universidad de La Rioja

Herrera Soler, H., R. Martínez Arias y M. Amengual Pizarro. 2011. *Estadística aplicada a la investigación lingüística*. Madrid: Eos Universitaria 2011. 376 pp. ISBN: 978-84-9727-430-2.

Existen no pocos manuales de estadística para lingüistas y para otros científicos o académicos que se moverían en campos habitualmente considerados más alejados de la *numeracy* o de una formación disciplinar que les haya familiarizado con números y fórmulas. Apuntan, pues, dichos manuales a lo que a la hora de la verdad se revelaría como un defecto de los programas y planes de estudios, defecto que debe ser subsanado pues las técnicas estadísticas son un componente insoslayable de prácticamente cualquier tarea investigadora en cualquier disciplina. No faltarán, con todo, quienes expresen la sospecha de que, cayendo ahora en el vicio opuesto del exceso, pueda darse en estos terrenos un efecto de *divinos números*, que vendrían a funcionar al modo de las *divinas palabras* en la obra de Valle.

Sin negar que prevenciones de este género mantengan siempre cierta pertinencia, es obvio que el arsenal de la estadística elemental forma parte de los recursos habituales del investigador. Y, si no capaz de desarrollar por sí mismo los diseños y cálculos que hagan a cada caso, deberá saber que tales cálculos y tales diseños han de hacerse y que sin ellos su trabajo puede perder todo su valor.

Otra cuestión es el buen uso que se haga de aquel arsenal y aún otra algo distinta, el modo en que se planeen y se ejerzan las relaciones entre estadística y la disciplina que se cultive. Viene esto a cuenta de un nuevo manual de estadística para lingüistas, con sus méritos y deméritos, sus puntos fuertes y sus debilidades, didácticas o técnicas, y sobre todo ello y al respecto del libro que comentamos, algo habremos de decir.

Sin embargo, quizá lo esencial en este tipo de obras (algo que por otra parte no supone un juicio crítico de la obra en consideración tomada singularmente) sea precisamente lo que hemos sugerido, que de modo más o menos explícito cada una de ellas pone de manifiesto una cierta concepción de la estadística y posiblemente también de la lingüística –o de la ciencia de que se trate– y de las relaciones entre la una y la otra, con-

cepción que mucho deberá a la experiencia docente de quienes sean sus autores, pero que también será tributaria de otras fuentes, científicas o ideológicas. En otras palabras, los autores partirán inevitablemente de una idea acerca de cómo debe enseñarse la estadística, algo a veces cargado con un melancólico e inevitable aire de eso que se conoce como *remedial teaching*; y partirán también de ciertas ideas más o menos justificadas sobre qué cosa sea la ciencia y la investigación.

Y podría pensarse que esa concepción más o menos latente se deja ver en lo que presentaremos como elecciones que los autores han tomado a la hora de escribir su libro. Ahora bien, llevar a cabo tales elecciones es cosa obligada: los autores han de seleccionar sin duda unos asuntos y dejar otros para mejor ocasión. Se argüirá que en tal selección podrán descubrirse igualmente aquellas concepciones previas de las que hablamos, pero lo cierto es que será también dicha selección una muestra del buen juicio –o del no tan bueno– de los autores. Y también, insistimos en este punto, es cierto que siempre podrán éstos fundamentar su criterio en su propia experiencia docente.

El libro que nos ocupa se divide en una *Introducción* y 11 capítulos a los que se añaden unos apéndices de los que luego hablaremos. El primer capítulo se dedica a “La investigación científica”, asunto sin duda que merece más de las aproximadamente treinta páginas que se le dedican. Podría señalarse también que la exposición no entra en detalles críticos y sobre todo que se ve marcada por cierta idea o representación recibida de ciencia y de método científico. Nos quedamos, sin embargo, con el indudable valor que tiene plantear un asunto central al que algunos lingüistas se enfrentarán en el ejercicio de su trabajo, pero rara vez de una manera sistemática y según una exposición razonada. Cuestión distinta es, otra vez, si tales asuntos debieran o no recibir mayor atención en los planes de estudios.

Siguen, por otro lado, en esto los autores una tradición bien establecida de manuales en que la estadística es precedida o acompañada de nociones sobre la metodología de la investigación, nociones que en este contexto desde luego no pueden pretender sustituir a una discusión en forma de mayor profundidad crítica, filosófica o histórica.

Tras ese primer capítulo, se entra propiamente en el campo de la estadística. Los siguientes tres se dedican a lo que *grosso modo* cabría clasificar bajo la etiqueta de Estadística descriptiva. En el quinto capítulo, “La inferencia estadística”, se encuentra el lector con los problemas centrales de lo que indica su título. Los capítulos seis y siete se dedican a los tests paramétricos y el ocho a los no paramétricos. En el nueve, se introduce el análisis de la varianza; en el número diez, la correlación y en el undécimo y último, la regresión lineal. Los capítulos incluyen a su final una sección de “Actividades”

El libro se cierra con unas “Respuestas a las actividades”, un apéndice con las tablas estadística más habituales, lo que nos parece un acierto por más que los paquetes estadísticos las hayan hecho innecesarias, una lista de “Websites de interés”, un Glosario y una Bibliografía, todo en el orden dicho.

Puede observarse en la enumeración de los capítulos que se dejan fuera algunos temas que otros autores consideran relevantes en un curso de introducción a la estadística como el enfoque bayesiano o el análisis factorial. Ya dijimos, sin embargo, que hay

que reconocer a los autores el derecho de seleccionar aquellas partes de la materia que les parezcan más adecuadas para un libro de las características del que comentamos. Sin duda, en el juicio que merezca dicha selección se comprometen principios que se compartirán o no con los autores, pero que habría que explicitar en cada caso.

De cualquier manera, y como propusimos antes, a continuación vamos a intentar sistematizar nuestras observaciones sobre esas opciones que los autores han tomado, no tanto porque unas nos parezcan “virtuosas” y otras erróneas, sino simplemente porque eso nos permitirá dar una visión más completa de la obra, de la concepción de la estadística que propugna y también de la concepción de la lingüística que le subyace.

Y es que existen distintas opciones a la hora de escribir un libro de iniciación a la estadística, opciones sobradamente ejemplificadas por la literatura existente en el mercado, en todas las lenguas de las llamadas de cultura y también en cualquier formato o soporte.

En primer lugar, se puede acompañar o no al manual de estadística de capítulos sobre otras disciplinas del tipo de la metodología científica. Ya hemos hablado de este asunto más arriba.

Cabe también la posibilidad de acogerse o no a una “especialización”: Hay manuales de estadística y hay manuales de estadística para lingüistas, historiadores, médicos, etc. En el caso del libro que nos ocupa, se trata de la lingüística, pero parece claro que los autores piensan –incluso se habla en el texto de “lingüística aplicada” aquí y allá– en esta disciplina en especial. Se diría, en cambio, que no se piensa en las aplicaciones propias de campos más o menos cercanos como la lingüística cuantitativa o el llamado procesamiento de lenguaje natural.

Una opción que toca más al núcleo del asunto es la de exponer o no, al nivel que se considere más apropiado, elementos de teoría de la probabilidad. Este es un asunto central desde el punto de vista interno. Si no se hace, es porque se piensa que se puede aprender la suficiente estadística sin aventurarse en terrenos que se esperan más complicados o más temibles. En este libro, la palabra “probabilidad” aparece poco y nos tememos que eso hace que toda la exposición quede marcada por esa ausencia.

En la exposición de las materias propias de la estadística, teóricas y prácticas, puede avanzarse al mismo tiempo que se presenta y explica gradualmente un programa o paquete estadístico, o bien puede dejarse ese importante aspecto del aprendizaje para otro lugar. Los autores se apoyan en el SPSS. Esto –puede argumentarse– sirve para disminuir la distancia entre el aprendizaje y la utilización práctica de las técnicas estadísticas. Es muy posible igualmente que entre el aprendizaje de la estadística y el de un programa haya cierto refuerzo mutuo.

Se puede establecer otra diferencia según que la exposición de la estadística incluya la explicación de aquellos conceptos teóricos básicos indispensables para una cabal comprensión de la materia o según se vaya más directamente a los instrumentos prácticos más habituales de la estadística inferencial. Este libro tiene una vocación quizá demasiado práctica, se centra en los tests de estimación puntual y de contraste de hipótesis más habituales y deja de lado los temas más comprometidos teóricamente. Aunque

el capítulo quinto, como tuvimos ya ocasión de advertir, es fundamental por cuanto se exponen en él los fundamentos de la estadística inferencial, conceptos como el de intervalo de confianza, verosimilitud, etc. se hallan virtualmente ausentes de esta obra. Incluso podría darse la desafortunada situación de que el lector apresurado no prestase la debida atención a ese citado capítulo quinto, con el resultado de que no aprovecharía lo que es, sin duda, el núcleo teórico de la exposición.

Otra opción es la de hacer mayor o menor énfasis en los problemas con que se encuentra la aplicación de los tests habituales y en los límites de las técnicas estadísticas. A este respecto, algo se dice en el libro, lo que es particularmente importante. Se trata de evitar que el lingüista se lance alegremente a comprobar que ha obtenido un *p-value* de menos de 0,05 para luego olvidarse de lo que tal cosa puede significar y también de todo lo demás. Al menos, ese lingüista no podrá decir que nada se le ha dicho acerca de las precauciones que deben tomarse.

Como la posición de los autores es sobre todo práctica, intentan transmitir un ejercicio razonable de la estadística elemental. Sin embargo, desde una posición más rigorista, en las opciones tomadas puede haberse incurrido en cierto extremismo. Hemos visto que los autores no hablan de teoría de la probabilidad y, si dejamos de lado los conceptos definidos exiguamente en el *Glosario* que cierra el libro, dejan en la ignorancia al lector sobre lo que sea, por ejemplo una *variable aleatoria* o una *distribución de probabilidad*: Se habla en el libro de “distribuciones de frecuencias” (página 87, sobre todo) y se evita aquel término y el concepto correspondiente.

Consecuencia de ello es, entre otras, que se resientan las explicaciones de algunos temas centrales, como puede ser, por ejemplo, la distinción entre pruebas paramétricas y no paramétricas. Así, en la página 121, al definir los “supuestos distribucionales” que se requieren para aplicar pruebas paramétricas, se incluye el de la “Distribución normal” sin mayores especificaciones.

Distribución que, por cierto, es prácticamente la única de la que se habla en el libro y que se presenta, según acabamos de ver, como “distribución de frecuencias”: no se explicita su fórmula y no se ve cómo sus parámetros, media y desviación típica, intervienen en la misma.

Que sea (con la excepción de la *t* de Student, aunque véase la cita que incluimos más abajo) la única de la que se habla en el libro nos hace sospechar el tipo de investigación lingüística a que va enfocada toda la obra. A saber, tests de competencia lingüística a sujetos dispuestos en un grupo de control y en otro sometido a otras condiciones, tests que suelen procurar datos que se ajustan a una distribución normal. Piénsese en la perplejidad de quien se haya iniciado de este modo en la estadística cuando se enfrente a una distribución hiperbólica, queremos decir a alguna distribución próxima a las frecuencias con que aparecen los distintos lemas en un corpus. O cuando se encuentre en situaciones en que aparezca cualquier otra distribución. Por otro lado, la habitual relación que se establece entre la distribución binomial, la de Poisson y la normal no es asunto tratado en este libro.

La edición del texto se resiente de una notable dosis de descuido. El aparente apresuramiento de la redacción hace que se incurran en deslices teóricos que quizá, dado el alcance del libro, no tengan demasiada importancia. Es de temer, sin embargo, que no quepa demasiado optimismo a este respecto. No nos resistimos a citar algunas de las expresiones infelices que jalonan las páginas del libro. Así, por ejemplo, en la página 104, un apartado encabezado con el título de “Distribución normal” comienza del siguiente modo: “Este tipo de distribución no es sólo una distribución normal sino una familia de distribuciones, cada una de las cuales está determinada por su media y su desviación típica”, enunciado en que, si se señalan correctamente los parámetros de esta distribución, no se hace sin haber previamente evacuado una innecesaria y torpe aclaración.

Más adelante, en un apartado dedicado a “Las distribuciones t de Student para muestras independientes”, leemos que aquéllas “Son una serie de distribuciones normales que, como la distribución normal, son simétricas y su curva de campana se define con más precisión a medida que el tamaño de la muestra se incrementa.” Se trata, creemos en este caso, de modos de hablar que son inadmisibles en la expresión escrita. El descuido, se aprecia también en la existencia de “frases viudas”, que parecen haber quedado fosilizadas de algún borrador anterior como sucede en la página 126 con un misterioso “Así debe evitarse la frase “aceptar la hipótesis nula” como sinónimo de “estadísticamente no significativo” porque implica una afirmación de la validez de la hipótesis nula. Errores en la comprobación de la hipótesis,” período con que acaba la sección 2.2.2.

Que en alguna ocasión este modo de hablar tan ligero vulnere no sólo el rigor, sino también la verdad será opinión a la que no faltarán adeptos. Es una lástima, no obstante, que se hallen no pocos a lo largo de esta obra, porque se adivina fácilmente que tales deslices se deben a la voluntad pedagógica de los autores.

Con todo, el resultado final es que la exposición de la estadística queda suspendida de hilos no demasiado fiables a falta de un fundamento teórico de la suficiente entidad, cuando la bibliografía existente prueba que una mayor dosis de rigor es posible. Se da la paradoja, incluso, de que al no haber definido ciertos conceptos básicos, se produce una introducción un tanto abrupta de algunos otros. Acuda el lector a las páginas en que se introducen las “distribuciones de frecuencias”, que no de probabilidades, o a aquéllas en que se habla de estadística paramétrica sin haber hablado de parámetros (excepción hecha del Glosario).

Nuestra reflexión final tiene que ver con una cuestión de fondo. Si el lingüista realiza un trabajo aceptable desde el punto de vista estadístico, eso no quiere decir que tenga una buena comprensión de lo que está en juego. Puede saber si el *p-value* obtenido en un test concreto está o no por debajo de cierto número estipulado previamente, pero puede que ignore para qué y por qué ha de realizarse tal cosa. Por eso, es tan importante que sepa qué se trae entre manos, qué es lo que está haciendo y qué le aporta la estadística. La obra que comentamos es consciente de que ésa es la esencia del asunto y no olvida nunca señalar la subsidiariedad de los resultados del cálculo a lo que podríamos

llamar su sentido, pero aun así nos atreveríamos a mantener que la misma tiene en su mayor virtud su principal flaqueza.

Si, como señalamos más arriba, el enfoque del curso hace que el aprendiz pueda poner manos a la obra rápidamente y aplicar la estadística en estudios reales, eso se hace al precio de la imprecisión teórica y de una fundamentación escasa. Reconozcamos que hay seguramente que mantener un difícil equilibrio entre lo posible y lo riguroso, y es en ese terreno donde los autores se han tenido que mover. Es igualmente posible que ese equilibrio no se establezca igual en una exposición escrita que en el discurso hablado

Esperemos que nuevas ediciones de esta obra limen los aspectos un tanto descuidados de la composición y la redacción del libro, una obra en que se ha optado por una estrategia a la que no pueden negarse ni sus ventajas ni sus inconvenientes, una estrategia sobre la que se puede y debe discutir, pero que es perfectamente respetable.

NOTA

* Correspondencia a: Pedro Santana Martínez. Universidad de La Rioja. Edificio de Filología. c/ San José de Calasanz s/n. 26004-Logroño (La Rioja). E-mail: pedro.santana@unirioja.es.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.