

**COMENTARIOS A ‘TÉCNICAS DE ANÁLISIS DE DATOS
PARA INVESTIGADORES SOCIALES.
APLICACIONES PRÁCTICAS CON SPSS PARA WIN-
DOWS’ de VIDAL DÍAZ DE RADA**

María Rosa Cañada Torrecilla
Universidad Autónoma de Madrid

Nos encontramos ante un libro de *técnicas de investigación social*, en el que se plantea el desarrollo de una investigación completa, profundizando en determinadas técnicas estadísticas y en la interpretación de los resultados alcanzados. Es una obra de sólido contenido, fruto de un largo y continuado trabajo, como son la mayoría de las investigaciones que utilizan como base la recogida y el posterior tratamiento de la información a partir de encuestas.

El libro está dividido en dos partes: en la primera se plantea la justificación metodológica y la contextualización teórica del fenómeno a investigar con la formulación de las hipótesis de trabajo que guiarán la investigación; en la segunda parte se exponen las diversas técnicas de análisis de datos según el número de variables analizadas: técnicas univariantes, bivariantes y multivariantes.

El capítulo primero se inicia con el planteamiento de la investigación, lo que le da pie al autor a una discusión teórica sobre el proceso de investigación social, sobre la lógica de la ciencia en Sociología. Siguiendo a Wallace, se muestra partidario de construir teorías a partir de las observaciones para después iniciar un proceso de deducción lógica para reducir la gran abstracción de la teoría a un conjunto de proposiciones o hipótesis de trabajo, cuyo contraste permitirá aceptar, modificar o rechazar la teoría formulada. La teoría ayuda a la investigación social al orientar los hechos y realizar preguntas sobre ellos, al tiempo que hace de guía en la elección del diseño de la propia investigación. En cuanto a la elección del método de investigación, es partidario de la metodología cuantitativa, a pesar de las limitaciones que supone la reducción de la realidad social a un conjunto de variables.

Uno de los capítulos más importantes de esta obra es el relativo a la medición en Ciencias Sociales. El capítulo comienza con la definición del concepto “medir” y expone

las peculiaridades de los fenómenos sociales, peculiaridades que afectan sustancialmente al proceso de medición. A continuación recoge los tipos de escalas desarrolladas por Stevens. Enumera las propiedades matemáticas de cada escala, y como éstas posibilitan la utilización de determinados tests estadísticos. La elección de la escala condicionará la prueba estadística a utilizar. Según Stevens, estadísticos de escalas superiores (intervalo, razón) no pueden ser utilizados para escalas que no cumplan esas propiedades (nominal, ordinal). Sin embargo, algunos autores cuestionan esta proposición al argumentar que los aspectos y problemas de la medición tienen que ser analizados independientemente de las técnicas de análisis de datos a utilizar. Estamos ante un capítulo de gran interés, donde se exponen una serie de conceptos y definiciones de manera bastante clara y sencilla, válidos para todos los que nos dedicamos a la investigación, en general.

El capítulo tercero nos introduce en el mundo de la estadística mediante ordenador. En concreto, en la utilización del paquete estadístico SPSS, insertando la explicación de los procedimientos del programa dentro del análisis de datos, de modo que son las hipótesis de trabajo las que guiarán el desarrollo de la obra y no las características del programa estadístico. En este sentido, el programa tiene un carácter secundario, es un medio dentro del desarrollo de la investigación concreta. No obstante el aprendizaje del programa requiere tiempo, a pesar de la facilidad de uso de las últimas versiones para Windows. Tenemos que tener presente que el autor de la obra utiliza la versión 8, aunque actualmente está disponible en el mercado la versión 11, si bien no existen grandes diferencias de funcionamiento entre ambas. Todo se resume en operaciones con archivos de datos, ventana de sintaxis, ventana de resultados y ventana del editor de gráficos.

De todas las técnicas de análisis de datos existentes, el autor se centra en las que son propias para variables nominales y ordinales: Tablas de Contingencia, Segmentación de variables discretas y Regresión Logística.

Tras la definición del programa de análisis a utilizar, el capítulo cuarto comienza con una exposición de cómo deben prepararse los datos para posibilitar un correcto procesamiento de éstos mediante un programa estadístico. Se trata de convertir las “opiniones sociales” en un archivo de datos en formato magnético, para lo cual es necesario cumplir una serie de etapas: definición de variables, entrada y modificación de datos.

En el proceso de preparación de los datos se exponen de manera pormenorizada los requisitos que éstos deben cumplir antes de ser introducidos en el ordenador para que sean procesados por el programa estadístico. Se contemplan con detalle todas las posibles respuestas-variables que generan el número de preguntas del cuestionario, así como las características de las mismas. A continuación, se pasa al apartado de codificación, grabación, y edición de los datos. Una vez que las variables han sido definidas, el siguiente paso es la introducción de las respuestas de los cuestionarios en el archivo definido anteriormente.

Esta tarea de introducción de los datos es aburrida y tediosa pero debe realizarse con mucho cuidado para no cometer errores, aunque existan un conjunto de procesos diseñados para detectar de forma rápida y eficaz los fallos cometidos en el proceso de recogida y de grabación de la información. Nos referimos a las estrategias de depuración de la información recogida a fin de evaluar y aumentar la fiabilidad y validez de nuestros datos. Coincidimos con Díaz de la Rada y otros autores, en que este aspecto recibe cada día menos atención, tal vez por la necesidad de agilizar la publicación de resultados y el

abuso de ordenadores, lo cual ha producido un descuido en los trabajos de depuración, a pesar del importante impacto que puedan tener en la calidad de los datos recogidos. En realidad las estrategias de depuración son múltiples: buscar inconsistencias entre ciertas preguntas, verificar si hay valores que no tienen lugar en determinadas preguntas, analizar las respuestas de las preguntas filtro, etc. Una vez que se han detectado errores en el cuestionario, una solución es la imputación de valores, dar valores nuevos a esos datos. El archivo de datos resultante completa los datos que faltaban, pero trae consigo nuevos problemas derivados del método de imputación, hasta tal punto que algunos autores piensan que se está alterando y cambiando la opinión de los entrevistados.

La parte segunda del libro se titula "de la metodología al análisis de datos", y comprende los capítulos del quinto al décimo. Se inicia con la exposición de una serie de procedimientos de análisis de datos originados por las necesidades surgidas para explicar correctamente las hipótesis de la investigación. Se comienza visualizando las respuestas que los entrevistados dan a cada una de las preguntas, visualización que se complica en la medida en que son solicitados estadísticos y gráficos de los datos recogidos. La peculiaridad en las respuestas de algunas preguntas requiere, en ocasiones, realizar determinadas transformaciones como la unión de diferentes categorías de una variable, reduciendo así el número de opciones de respuesta. En ocasiones la información expresada por toda la muestra se puede comparar con la expresada por determinados colectivos. Incluso al final del capítulo quinto se construyen unos índices mediante la técnica de transformación de variables, utilizando procedimientos de cálculo y ponderación específica de ciertos valores. Para todo ello resulta muy útil el SPSS.

He leído con mucho interés, los capítulos sexto y séptimo, dedicados al Análisis Bivariante. Me ha interesado más que los resultados de la investigación concreta, la metodología empleada, la utilización de determinados estadísticos, no tan frecuentes para los que nos dedicamos al tratamiento de información cuantitativa más que cualitativa, como es el caso de la persona que escribe estas páginas; el estudio en concreto de las relaciones entre variables nominales y ordinales, a partir de las Tablas de Contingencia. Previamente el autor ha dedicado unas páginas a la idea de inferencia y significación estadística, conceptos que al lector seguramente le vendrá muy bien recordar y que son necesarios para comprender el sentido de la relación entre variables.

La interpretación de las Tablas de Contingencia se realiza mediante una exposición muy extensa de diferentes estrategias: utilización de residuos, Odds ratio, diferencia de proporciones, coeficiente "d" y Riesgo relativo.

Para medir la relación entre variables nominales utiliza diversos estadísticos: Ji – Cuadrado, Coeficiente de Contingencia, Phi, V de Cramer y Lambda. Es de valorar la consideración que realiza el autor sobre la utilización de estos indicadores, exponiendo los requisitos que debe cumplir una buena medida de asociación.

Expone, a continuación, una serie de coeficientes adecuados para medir la relación entre variables ordinales, cuyo objetivo es analizar la relación entre la clasificación de cada individuo en cada una de las variables utilizadas. En el caso concreto de esta investigación, relaciona, por un lado, la posición social de los encuestados y nivel de equipamientos y por otro, el nivel de equipamientos con el nivel de ingresos.

Tras la exposición se llega a la conclusión de que la elección de una determinada medida es un criterio personal, aconsejando el autor que cada investigador utilice la

medida que mejor comprenda, aunque también se pueden solicitar todos los estadísticos y estudiar las diferencias entre ellos, para escoger el más interesante según el objetivo del análisis.

En los capítulos octavo y noveno el autor explica algunas técnicas propias del Análisis Multivariante. En primer lugar analiza Tablas de Contingencia formadas por tres o más variables. Se trata de introducir una variable de control (Z) cuya función es controlar la relación entre las variables X e Y.

A continuación el autor expone la técnica del CHAID (*Chi-Square Automatic Interaction Detector*), cuyo objetivo es descubrir el efecto de la interacción entre variables, cómo el efecto de una variable independiente varía según el valor de otra variable independiente. Algunos autores denominan a esta técnica "Análisis de Segmentación", en el sentido de "polarizar" las frecuencias de las categorías de una variable dependiente cualitativa seleccionando los sujetos que cumplan determinadas características. La utilización de esta técnica ha permitido conocer las categorías de las variables independientes con un comportamiento homogéneo respecto a la variable dependiente. Aunque para utilizar esta técnica de análisis de datos, basta hacer clic sobre "SPSS chaid" dentro de la opción "Estadísticos" del menú principal, me parece bastante complejo el proceso que hay que seguir hasta obtener los resultados definitivos.

Sorprende la utilización de la técnica de la regresión logística, muy utilizada por los investigadores de ciencias sociales de todo el mundo, y poco empleada en nuestro país. La regresión logística permite conocer como un número de variables tanto cuantitativas como cualitativas influyen en la predicción de un determinado fenómeno. Frente a otras técnicas de regresión, el modelo logístico tiene como variable dependiente una variable dicotómica y no requiere que los datos cumplan unos supuestos tan restrictivos.

El autor nos comenta que después de construir el modelo hay que comprobar el ajuste de éste mediante métodos gráficos y otros basados en la razón de verosimilitud. Después hay que interpretar los coeficientes obtenidos para ver la magnitud y el sentido de la relación entre cada variable independiente y la variable dependiente, viendo como la primera influye en la segunda.

La obra acaba con unas conclusiones donde el autor hace una síntesis de los temas abordados a lo largo de todo el libro. Hay una primera parte referida a la investigación sobre la emergencia de un "nuevo consumidor" en la sociedad navarra que queda demostrado que no se produce tras analizar las hipótesis planteadas. Y una segunda parte donde realiza una visión de conjunto sobre las técnicas de análisis de datos empleadas teniendo como base los datos de una investigación real sobre los hábitos de consumo de la sociedad navarra.

Es interesante la propuesta que realiza para que el lector se anime a realizar un pequeño trabajo de campo con el cuestionario incluido en el anexo I y así comparar los hábitos de consumo de su círculo social inmediato con los mostrados en este trabajo sobre la población navarra.

Por todo lo expuesto consideramos que el libro de Díaz de Rada es muy útil, en general, para todos los que nos dedicamos a la investigación y, en particular, para los que orientamos nuestra labor a la docencia de técnicas de cuantificación aplicadas a las Ciencias Sociales. Constituye un buen manual de Estadística Aplicada de alto contenido didáctico, donde se incluyen junto con la explicación teórica de la técnica una demostra-

ción de cómo utilizarla con el SPSS y una interpretación de los resultados. Este es el camino más apropiado para comprender las características y condiciones de utilización de las técnicas de análisis de datos.

Referencias

Díaz de Rada, V. (1999) *Técnicas de análisis de datos para investigadores sociales. Aplicaciones prácticas con SPSS para Windows*. Madrid: Ra-Ma.