

TÉCNICAS DE ANÁLISIS MULTIVARIANTE PARA INVESTIGACIÓN SOCIAL Y COMERCIAL (Ejemplos prácticos utilizando SPSS versión 11)

INTRODUCCIÓN

En el verano de 1998 terminamos *Técnicas de análisis de datos para investigadores sociales*, dando cuenta de la existencia de un gran número de procedimientos de análisis de datos que no fueron explicados en ese trabajo. Nos referíamos, concretamente, a las técnicas multivariantes que no distinguen entre variables dependientes e independientes, conocidas popularmente como *técnicas de interdependencia entre variables*. Al final de la obra animamos a los lectores a profundizar en su aplicación, aunque consideramos también la explicación de éstas para algún trabajo posterior. Después de cuatro años, y tras múltiples interrupciones y cambios en su concepción, el lector tiene en sus manos un trabajo dedicado a las técnicas multivariantes de interdependencia.

Las referencias bibliográficas sobre el análisis multivariante están experimentando un rápido aumento en los últimos años, apareciendo trabajos que se ocupan tanto del desarrollo matemático como de las distintas aplicaciones de cada técnica a la investigación social y comercial. Este trabajo se acerca más a este segundo planteamiento, y tiene como objetivo explicar determinadas técnicas multivariantes centrados fundamentalmente en las condiciones de utilización de cada una, y en la interpretación de sus resultados. Con el fin de explicar las técnicas multivariantes de forma sencilla se han utilizado ejemplos de investigaciones reales, buscando problemas similares a los que el profesional se enfrenta en su trayectoria investigadora.

Se trata, en definitiva, de presentar una serie de pautas sobre cuándo utilizar una determinada técnica multivariante y de cómo interpretar sus resultados. Obviaremos los complicados planteamientos matemáticos para centrar nuestra atención en las condiciones de aplicación de la técnica y en la interpretación de los resultados obtenidos. El objetivo se centra en proporcionar una referencia sintética de las similitudes y diferencias entre las distintas técnicas multivariantes, considerando tres aspectos:

1. Presentar las diferencias entre las distintas técnicas multivariantes, dando respuesta a preguntas como: ¿es lo mismo un análisis factorial y un análisis de correspondencias?

2. Explicar qué técnica utilizar en función del objetivo de la investigación, respondiendo a la pregunta: ¿para qué sirve cada una de las técnicas?
3. Por último, prestaremos atención a cómo aplicar de forma correcta cada técnica, analizando las condiciones que deben cumplir los datos (tamaño muestral, comprobación de determinados supuestos, etc.).

En realidad el objetivo principal de esta obra es proporcionar una serie de pautas para comprender la realidad social de la mejor forma posible. Con este fin utilizaremos la técnica de análisis de datos que mejor se adapte a los objetivos de la investigación, considerando siempre que cuando dos técnicas explican lo mismo utilizaremos la más sencilla, la que sea más fácil de aplicar. Buscaremos también facilitar la comprensión de las personas que encargan la investigación, pensando siempre que éstas pueden desconocer determinados procedimientos de análisis de datos, y que generalmente demandan explicaciones muy sencillas enfocadas a la rápida toma de decisiones.

La obra que el lector tiene en sus manos, *Técnicas de análisis multivariante para investigación social y comercial*, comenzó a gestarse en los primeros meses de 1999 y sigue fiel al planteamiento de la obra anterior, que recordemos tenía como objetivo solucionar problemas de una investigación utilizando determinadas técnicas de análisis de datos. En aquel momento se presentaba una investigación completa, desde la formulación de hipótesis hasta la exposición de resultados y generalización de los mismos, "aprovechando" las preguntas de la investigación para explicar el uso e interpretación de diversas técnicas de análisis de datos. Los comentarios de muchos lectores nos han demostrado lo acertado de aquel planteamiento y, por ello, optamos por repetirlo en este trabajo.

Sin embargo, la diversidad de técnicas que se explican en esta obra impiden utilizar la misma estrategia, puesto que es muy difícil aplicar todos estos procedimientos en una misma investigación y, más importante, es una concepción que nada tiene que ver con la situación real de la investigación. Aunque es posible utilizar en una misma investigación el análisis factorial, el análisis de correspondencias y el análisis cluster, por hablar de algunas de las técnicas que aquí se explican, lo normal es utilizar sólo alguna de éstas, en función del objetivo de la investigación y el tipo de variables disponibles en la investigación.

Este hecho, unido a los buenos resultados del planteamiento utilizado en el libro anterior, nos animaron a buscar una solución intermedia, un desarrollo expositivo donde los problemas de investigación sean los que *determinen* la utilización de una técnica de análisis. No se trata de explicar el análisis exploratorio en el segundo capítulo, el análisis factorial en el tercero,... sino de proponer *situaciones de investigación* que

precisen el empleo de una *herramienta* para lograr los objetivos fijados; es decir, que sean las *necesidades de la investigación* las que indican las técnicas a utilizar.

Esto se consigue eligiendo situaciones reales de investigación, de modo que hemos optado por una exposición similar al libro anterior planteando varias investigaciones que requieran la utilización de técnicas determinadas. Así en vez de utilizar una misma investigación desde el principio hasta el final, como hicimos en aquel trabajo, aquí se utilizan investigaciones concretas que precisan de la aplicación de una o varias técnicas de análisis de datos. Conviene dejar claro que se han utilizado datos de investigaciones reales, con las consecuencias que ello conlleva. Trabajar con datos reales plantea una situación de *simulación real* muy similar a lo que es la práctica investigadora diaria, pero requiere también enfrentarse a decisiones complejas sobre cómo tratar los casos perdidos, la solución a tomar cuando una variable no cumple los supuestos paramétricos, qué hacer con los casos extremos, etc. La solución a estos problemas no suele aparecer en los libros al uso, puesto que al utilizar datos preparados no surgen estas situaciones "anormales". Consideramos que ésta es una de las novedades del presente libro, trabajar con los datos *brutos* tal y como son recogidos por el investigador, presentando todas las manipulaciones y transformaciones que faciliten el tratamiento más adecuado. Se trata, en definitiva de reproducir la práctica cotidiana de cualquier instituto de investigación.

En línea con las ideas presentadas en el párrafo anterior, la novedad de este texto está en la ausencia de obras publicadas que expliquen *con profundidad* el proceso de una investigación mediante encuesta. Pese a que en los últimos años se ha producido una proliferación de publicaciones sobre el funcionamiento de programas estadísticos, así como otros textos donde se exponen ejemplos de investigaciones reales (entre otros Soler y Perdíguer 1992, González Lobo 2000, Alonso Rivas y Ruiz de Maya 2001, etc.), el texto que aquí presentamos unifica ambos planteamientos al exponer problemas concretos del diseño de la investigación, junto a las dificultades del análisis de datos. Creemos que hay una importante carencia en el mercado español de un texto que considere conjuntamente estos planteamientos: por un lado los criterios metodológicos de la investigación mediante encuesta, por otro el análisis de datos reales de una investigación y, por último, la capacidad de generalizar conclusiones utilizando las propiedades de los paquetes estadísticos.

Las obras que explican la dinámica de la investigación basadas en ejemplos de investigaciones reales nos parecen excelentes, pero plantean algún problema puesto que en ocasiones se producen grandes "saltos" entre el planteamiento de la investigación y los resultados proporcionados; no quedando muy claro -por ejemplo- las transformaciones realizadas antes de aplicar el análisis factorial, los diversos análisis exploratorios realizados antes de presentar el modelo definitivo, etc. Este trabajo sigue una filosofía muy similar a estos textos, aunque aquí se solventan estos problemas

presentando todos los cálculos, hipótesis alternativas, dudas del investigador, etc. ocurridas en cada uno de los ejemplos. Se trata, en definitiva, de presentar menos ejemplos de investigaciones pero ser sumamente exhaustivos en el desarrollo de los ejemplos elegidos.

La complejidad en el cálculo de algunas técnicas de análisis multivariante, unida a la amplia difusión de ordenadores personales y paquetes estadísticos, nos ha llevado a realizar todos los cálculos con un programa estadístico, reproduciendo al máximo la investigación realizada en cualquier instituto de investigación. En este trabajo se ha elegido uno de los programas estadísticos más utilizados en la investigación social y comercial, el SPSS para Windows. Las razones que han motivado esta elección son, entre otras, la facilidad de su uso, la explicación de determinados procesos estadísticos en sus menús de ayuda, y la existencia de una versión para estudiantes que posibilita - a un precio muy asequible- que éstos puedan adquirirlo y familiarizarse con él. Conviene dejar claro que no es la explicación del programa estadístico la que guiará la exposición, sino que son las necesidades de la investigación las que aconsejan la utilización -y explicación- de determinados procesos de análisis de datos.

Todos los cuadros de diálogo del programa SPSS se presentarán en *figuras*, mientras que las *ilustraciones* serán utilizadas para realizar síntesis de explicaciones, presentar esquemas, explicar gráficos, etc. Los resultados obtenidos por cada procedimiento serán presentados en *tablas* y *gráficos* cuando se trate de los resultados directos proporcionados por el programa, utilizando también *cuadros* para exponer sintéticamente y ordenadamente los datos de una tabla. En el texto, todas las menciones a cuadros de diálogo del SPSS, recuadros, etc. aparecerán en letra cursiva.

En cuanto a su estructura, el libro comienza con un capítulo dedicado a la definición del análisis multivariante, realizando una clasificación de las técnicas multivariantes y una presentación breve de las técnicas más utilizadas en la actualidad. La utilización del análisis multivariable en una investigación debe comenzar con un estudio detallado de la información recogida, aspecto que es analizado en el segundo capítulo. El objetivo de este capítulo es proporcionar una visión general de las técnicas de exploración de datos, desde el simple proceso de inspección visual de las representaciones gráficas hasta el proceso de comprobación de la evaluación de los supuestos del análisis multivariante, pasando por el tratamiento de los casos atípicos y ausentes.

Los capítulos siguientes están dedicados a la explicación de diversas técnicas de análisis multivariante, concretamente las técnicas de interdependencia: el análisis factorial en el capítulo tres, el análisis de correspondencias en el cuatro, y el análisis cluster en el quinto. Todos estos capítulos siguen una estructura muy similar, que comienza con una presentación donde se exponen los objetivos de la técnica, y

continúa con la elección del ejemplo utilizado para explicar cada técnica. El trabajo termina con un anexo donde se presenta el lenguaje de sintaxis de SPSS.

Respecto a los destinatarios de este libro, y aunque en un primer momento fue realizado pensado para utilizarlo como manual en un curso avanzado de análisis de datos, las sucesivas modificaciones realizadas en los últimos años -tratando de hacer más fácil la comprensión del mismo- lo hacen muy apropiado para todo tipo de personas interesadas en el análisis de encuestas: sociólogos, trabajadores sociales, investigadores de mercado, psicólogos, economistas, departamentos de marketing, etc.; aunque los ejemplos utilizados son más propios de la sociología y la investigación comercial.

El libro está escrito en un lenguaje claro y sencillo, y considera que el lector tiene conocimientos sobre investigación social. Pese a que nuestro interés es realizar una explicación "intuitiva" de cada una de las técnicas multivariantes, la comprensión de éstas precisa del lector unos conocimientos previos de estadística, digamos a un nivel intermedio. Se supone también que el lector ha trabajado con técnicas uni y bivariantes, y que conoce la lógica y la interpretación de las pruebas de significación. De la misma forma, solucionar los ejemplos con el programa estadístico SPSS requiere que el lector haya trabajado con él, o que conozca -al menos- la lógica de funcionamiento y esté familiarizado con los resultados proporcionados por éste o por cualquier otro paquete estadístico. Este libro se plantea como una segunda parte de *técnicas de análisis de datos para investigadores sociales*, de modo que aconsejamos su lectura a los lectores que no todavía no tienen este conocimiento previo.

Es preciso volver a insistir que el objetivo de este trabajo no se centra tanto en una exposición de distintas técnicas de análisis de datos, sino que se explica cómo solucionar problemas de una investigación utilizando determinadas técnicas de análisis de datos. Este planteamiento lleva a centrar nuestro interés en razonar por qué se utiliza una determinada técnica, y cómo interpretar los resultados que proporciona. No nos preocupa la comprensión de fórmulas complejas, que además trataremos de evitar en la medida de lo posible, sino que deseamos profundizar en la utilización de determinadas técnicas y en la interpretación de los resultados que proporcionan. Estos aspectos diferencian este trabajo de otras obras destinadas a la descripción de un programa de ordenador, o a la exposición de diversas técnicas de análisis estadístico. En estas últimas cuando se analizan datos suelen ser pocos y muy simples, puesto que lo prioritario es la técnica de análisis. Cuando se analiza el funcionamiento de un programa de ordenador los ejemplos se utilizan como "un medio" para explicar el programa, y nunca como un fin.

Un breve apunte para recordar algunas de las personas que me han ayudado en la elaboración de esta obra. La mayor parte de las ideas de este trabajo se originaron durante una estancia de investigación del autor en la universidad de Essex, que fue posible gracias a una ayuda concedida por el vicerrectorado de profesorado de la Universidad Pública de Navarra. Me gustaría agradecer la ayuda proporcionada por Tony Coxon, Nick Buck y el resto del personal del *ESRC Research Centre on Micro-Social Change* de la citada universidad. Prácticamente todos los contenidos aquí presentados fueron expuestos en un curso sobre *análisis multivariante en ciencias de la salud*, impartido dentro del programa de formación continua del personal del Departamento de Salud del Gobierno de Navarra. Deseo agradecer a M^a Jesús Guembe Suescun e Isabel Ezpeleta la oportunidad de impartir el citado curso. Gracias también a María Pellejero, Leyre Iribarren y Maite Ruffo por la cuidada revisión de las primeras versiones del trabajo. Recordar, y agradecer también, a Encarna y Leo que han estado conmigo en los momentos más importantes de mi vida académica. Para terminar, me gustaría dedicar este libro a Demetrio, un amigo por encima de todo.

Vidal Díaz de Rada
Pamplona, 2002

INTRODUCCIÓN

CAPÍTULO 1: ANÁLISIS MULTIVARIANTE: DEFINICIÓN Y CLASIFICACIÓN

1. Definición de Análisis Multivariante
2. Clasificación de las técnicas multivariantes
3. Presentación y descripción breve de las técnicas multivariantes más utilizadas
 - 3.1. Técnicas de dependencia
 - 3.2. Técnicas de interdependencia
4. Una precisión sobre el contenido de la obra

CAPÍTULO 2: PRIMER ACERCAMIENTO A LOS DATOS: ANÁLISIS EXPLORATORIO

1. Introducción: objetivos de la técnica
2. Explicación mediante un ejemplo: exploración univariante
 - 2.1. Definición del problema a investigar
 - 2.2. Proceso de realización del análisis exploratorio utilizando el programa estadístico SPSS
 - 2.3. Visualización de la distribución: examen gráfico de los datos
 - 2.4. Comprensión de la distribución utilizando estadísticos univariantes
 - 2.5. Transformaciones de la distribución
 - 2.6. Tratamiento de casos atípicos y ausentes
3. Explicación utilizando un ejemplo con dos variables
 - 3.1. Selección de los procedimientos: proceso de realización del análisis exploratorio utilizando el programa estadístico SPSS
 - 3.2. Interpretación de resultados
4. Evaluación de los supuestos del análisis multivariante

5. Otras técnicas de análisis para conocer la distribución de los datos

CAPÍTULO 3: ANÁLISIS FACTORIAL

1. Introducción: objetivos de la técnica
2. Tipos de Análisis Factorial
 - 2.1. Análisis de Componentes Principales
 - 2.2. Análisis de Factores Principales
 - 2.3. Otros tipos
3. Análisis Factorial de Componentes Principales: explicación mediante un ejemplo
 - 3.1. Definición del problema a investigar
 - 3.2. Primer análisis de la información
 - 3.3. Proceso de realización del análisis factorial con el programa estadístico SPSS
 - 3.4. Primeros resultados: comprobación del ajuste del modelo
 - 3.5. Interpretación de resultados
 - 3.6. Puntuaciones factoriales
 - 3.7. Ayudas a la interpretación: representaciones gráficas

CAPÍTULO 4: ANÁLISIS DE CORRESPONDENCIAS

1. Introducción: objetivos de la técnica
2. La lógica del Análisis de Correspondencias
3. Análisis de Correspondencias Simples: explicación mediante un ejemplo
 - 3.1. Definición del problema a investigar
 - 3.2. Primer análisis de la información
 - 3.3. Proceso de realización del análisis de correspondencias simples con el SPSS
 - 3.4. Explicación de los resultados obtenidos
 - 3.5. Análisis de los componentes gráficos
 - 3.6. Interpretación de resultados
 - 3.7. Análisis de correspondencias con modalidades suplementarias

4. Análisis Factorial de Correspondencias Múltiples: explicación mediante un ejemplo
 - 4.1. Definición del problema a investigar
 - 4.2. Primer análisis de la información
 - 4.3. Proceso de realización del Análisis de Correspondencias Múltiples con el programa estadístico SPSS
 - 4.4. Interpretación de resultados
5. Anexo: proceso de elaboración de una tabla de datos para el Análisis de Correspondencias Simples

CAPÍTULO 5: ANÁLISIS DE CONGLOMERADOS (CLUSTER)

1. Introducción: objetivos de la técnica
2. Tipos de Análisis de Conglomerados
3. Análisis de Conglomerados Jerárquico: explicación mediante un ejemplo
 - 3.1. Definición del problema a investigar
 - 3.2. Primer análisis de la información
 - 3.3. Proceso de realización del Análisis de Conglomerados Jerárquico con el programa estadístico SPSS
 - 3.4. Explicación de los resultados obtenidos
 - 3.5. Interpretación de la clasificación resultante: características de los conglomerados
4. Análisis de Conglomerados no Jerárquico aplicado a los resultados del Análisis de Componentes Principales
 - 4.1. Definición del problema a investigar
 - 4.2. Primer análisis de la información
 - 4.3. Proceso de realización del Análisis de Conglomerados no Jerárquico con el programa estadístico SPSS
 - 4.4. Análisis de resultados
 - 4.5. Validación de la clasificación efectuada
 - 4.6. Una tipología del consumidor navarro
5. Anexo: medidas de distancia y proximidad

ANEXO: LENGUAJE DE SINTAXIS DE SPSS

1. Introducción: organización de las instrucciones de sintaxis
2. Exploración de los datos
 - 2.1. Análisis exploratorio de datos
 - 2.2. Comandos de sintaxis utilizados en los ejemplos del capítulo dos
3. Análisis Factorial
 - 3.1. Análisis Factorial de Componentes Principales
 - 3.2. Comandos de sintaxis utilizados en el ejemplo del capítulo tres
4. Análisis de Correspondencias
 - 4.1. Análisis de Correspondencias Simples
 - 4.2. Comandos de sintaxis utilizados en el primer ejemplo del capítulo cuatro
 - 4.3. Análisis de Correspondencias Múltiples
 - 4.4. Comandos de sintaxis utilizados en el segundo ejemplo del capítulo cuatro
 - 4.5. Proceso de elaboración de tablas de datos para el Análisis de Correspondencias Simples
 - 4.6. Comandos de sintaxis utilizados en la elaboración de tablas de contingencia para el Análisis de Correspondencias Simples
5. Análisis de Conglomerados (Cluster)
 - 5.1. Conglomerados Jerárquicos
 - 5.2. Comandos de sintaxis utilizados en el primer ejemplo del capítulo cinco
 - 5.3. Conglomerados no Jerárquicos (método k-medias)
 - 5.4. Comandos de sintaxis utilizados en el segundo ejemplo del capítulo cinco
 - 5.5. Medidas de distancias y proximidad
 - 5.6. Comandos de sintaxis utilizados en el quinto apartado del capítulo cinco