

Introducción

Este libro es una introducción práctica al análisis de datos mediante el paquete estadístico BMDP. Está concebido como una guía práctica y didáctica donde se conjugan tres elementos:

- 1) Se presentan ejemplos de análisis reales, planteados como casos, que ayudan a identificar oportunidades de aplicación de las técnicas.
- 2) Se presentan las instrucciones que hay que utilizar en el paquete para obtener los análisis, y
- 3) Se muestran resultados (“outputs”) seleccionados, y, lo que hace particularmente práctico el manual, los resultados son comentados tanto desde el punto de vista de los estadísticos utilizados como desde el punto de vista del significado práctico de los mismos.

Una aportación didáctica interesante de este texto es su formato gráfico, que permite incluir cuadros de comentarios que expresan en cada análisis el significado concreto de las instrucciones y el significado concreto de los resultados que van introduciéndose. De este modo el texto provee una guía rápida de consulta y aprendizaje donde los “outputs” comentados ilustran directamente al lector sobre el uso y los productos del paquete estadístico.

Un manual práctico como éste adquiere su pleno significado junto al ordenador, en la práctica de los ejemplos u otros similares que el lector puede estar deseando analizar. Pero, **dado que los “outputs” de los paquetes estadísticos son muy semejantes -en muchos casos incluso en su formato- disponer de “outputs” comentados puede resultar de gran ayuda** para estudiantes en prácticas y, en general, personas que están tratando de aprender a utilizar paquetes estadísticos y a interpretar adecuadamente sus productos.

El BMDP es uno de los grandes paquetes estadísticos. Aunque los paquetes estadísticos tienen una amplia zona de solapamiento, de modo que muchos análisis pueden efectuarse indistintamente con uno u otro, también es cierto que presentan particularidades, de modo que algunos análisis que pueden efectuarse con BMDP no están fácilmente disponibles en otros paquetes.

El BMDP es fácil de aprender y de utilizar, aunque requiere un cierto entrenamiento. Sus instrucciones funcionan mediante frases que se ordenan en párrafos estructurados. Una vez que se ha adquirido la “lógica” de funcionamiento del paquete, que es común a todos los programas y problemas, es necesario contar con una guía de usuario de instrucciones al lado para poder trabajar con soltura. La documentación del programa provee esa guía y un completo manual en volúmenes independientes. Por supuesto, para un conocimiento profundo del paquete es imprescindible contar con un estudio atento del capítulo del manual relativo al tipo de análisis en que estamos interesados.

El BMDP clásico no funciona con ventanas, menús y cuadros de diálogo. El usuario principiante considera esto un inconveniente, y efectivamente lo es en el sentido de que requiere un poco más de atención para aprender su manejo. Pero es erróneo pensar que todo son ventajas en el formato de ventanas y todo desventajas en el formato de instrucciones mediante párrafos y sentencias. Cuando se efectúa análisis de datos profesionalmente, tanto en el campo aplicado como en el de

investigación, se aprecia, y mucho, la simplicidad de las instrucciones que permite resolver de un modo elegante y simple, a menudo con una sola línea, lo que en términos de ventanas supone muchas selecciones, muchos “clicks” y muchos “Ok”. Los sistemas de instrucciones mediante “ventanas” en análisis de datos hacen sencillo lo complicado, pero también muchas veces hacen complicado y tedioso lo que es sencillo. Cuando ya se ha invertido ese “un poco de tiempo más” en manejar el paquete mediante instrucciones, se ahorra mucho tiempo al disponer de conjuntos de instrucciones preparadas para los análisis que solemos efectuar más frecuentemente y se gana en control al detectar errores de análisis que pasarían inadvertidos por otro sistema de instrucciones. Obviamente lo más ventajoso es disponer de ambos sistemas a elección, pero eso no está disponible en el BMDP clásico, aunque sí en versiones parciales o diferentes del mismo (p.e. BMDP New System) que son casi autoexplicativas desde el punto de vista del manejo, y permiten conectar fácilmente con el BMDP clásico.

En cualquier caso, incluso para esas otras versiones o para los outputs de otros paquetes estadísticos de naturaleza semejante, esperamos que los ejemplos de análisis y sobre todo los “outputs” comentados sean de gran utilidad práctica. **Ver delante, de un golpe de vista, una tabla completa de “output” de un análisis de regresión, por ejemplo, y disponer de una explicación de qué es cada término creo que es el tipo de ayuda visual** que justifica particularmente este texto.

Así que este texto está orientado a:

- 1) Introducirse en el análisis de datos de un modo práctico, mediante el análisis de “casos”
- 2) Introducirse en el BMDP, aprendiendo sus aspectos básicos, generales a cualquier programa BMDP, y, en profundidad, algunos de los tipos de análisis más comunes que se suele requerir en los análisis de datos, y
- 3) Interpretar los “outputs” en sentido estadístico y en el sentido de las preguntas formuladas en los casos prácticos.

Se ha evitado explicar primero el funcionamiento del paquete o sus instrucciones para después aplicarlo. En lugar de ello, desde el principio se comienza realizando análisis prácticos para un problema concreto, integrado estadística y aprendizaje del paquete. Poco a poco, mediante ejemplos, se van introduciendo y desvelando los principales aspectos de procedimiento relativos al funcionamiento del programa.

El material está estructurado en 6 partes:

- 1) Análisis descriptivos.
- 2) Descripción detallada y tabulación de variables.
- 3) Análisis de regresión.
- 4) Análisis de regresión curvilínea.
- 5) Contrastes estadísticos de 1 y 2 muestras.
- 6) Contrastes estadísticos de 2 o más muestras.

Una visión más detallada de los tópicos que cubren estas partes puede obtenerse consultando el índice que se presenta a continuación. El material de este texto recorre de modo exhaustivo o casi exhaustivo 5 programas básicos (1D, 2D, 3D, 7D, 1R) y se hace mención o se utilizan parcialmente sólo unos pocos más (5D, 6D, 8D, 4F, 2R, 3R, 5R, 6R, y 9R). Este manual puede resultar adecuado para iniciarse en las necesidades básicas en descripción, pruebas t, análisis de la varianza y análisis de regresión lineal simple, múltiple y curvilínea, y, para abrir las puertas al uso de los programas del resto del paquete cuya estructura de funcionamiento es siempre constante.

A pesar de esta limitación de espectro, el tratamiento de estos tópicos es más sofisticado del que suele encontrarse en textos introductorios, aprovechando las posibilidades del paquete para ofrecer diversas alternativas de análisis cuya utilidad depende de los propósitos del investigador y de la

naturaleza de los datos de los que dispone. Así por ejemplo se dedica atención a la obtención de medias ponderadas, a las transformaciones con operadores lógicos, a los estadísticos de tendencia central robustos (medias recortadas, estimador de Hampel y estimador Biweight), a la media geométrica, armónica, y en general a las medias de Foster, al test de normalidad W de Shapiro y Wilks, a las transformaciones no lineales, a los gráficos de tallo y hojas, a la regresión stepwise, a la determinación de los mejores conjuntos de predictores lineales, a la T^2 multivariada de Hotelling y a la D^2 de Mahalanobis, a las pruebas t recortadas o a los gráficos QQ, por poner algunos ejemplos. Y lo interesante es que se ofrecen ejemplos de elección e interpretación de los estadísticos de un modo muy gráfico, como anotaciones sobre el output.

Sin embargo, el BMDP es una librería de programas mucho más extensa. Contiene alrededor de unos 44 programas integrados que funcionan del mismo modo y abarcan amplias zonas de las necesidades estadísticas del análisis de datos. Alguno de estos programas se ocupa de cuestiones relativamente sencillas, como el cálculo de estadísticos descriptivos, pero otros muchos son poderosas herramientas capaces efectuar muchos tipos de análisis distintos dentro de una misma área de técnicas. Permite análisis sofisticados en análisis de varianza, análisis discriminante, regresión, supervivencia, log-lineal, cluster, series, factorial, análisis de correspondencias, análisis canónico, regresión logística, estimación de máxima verosimilitud, etc. Los programas se clasifican en series simbolizadas por una letra: descripción (D), tablas de frecuencias (F), tablas de vida y análisis de supervivencia (L), análisis multivariados (M), regresión (R), especial (S), series temporales (T) y análisis de varianza (V).

El paquete o programa estadístico que utilizamos no es la estadística, ni siquiera debiera confundirse con la estadística aplicada al análisis de datos. Pero, de hecho, los investigadores y los profesionales ven frecuentemente la estadística aplicada y el análisis de datos a través del paquete estadístico que utilizan. En algunos casos la identificación puede llegar a ser tal que no se hace lo que el paquete no puede hacer y se hace aquello que el paquete que se utiliza hace. Evidentemente hemos de evitar esta suplantación y sus efectos perversos sobre el modo en que se diseña y se realiza el análisis de los datos.

Esta es una de las razones por las que, en mi opinión, es importante que el profesional y el investigador que han de enfrentarse frecuentemente con análisis de datos conozcan más de un paquete estadístico, para evitar una visión demasiado limitada y para poder disponer de una abundancia de recursos que permita la suficiente flexibilidad y capacidad de dar respuesta a las necesidades de análisis.

En ese sentido el BMDP es una excelente alternativa y la necesidad de estudiar y preparar los análisis estimula un abordaje menos mecánico y más comprensivo del análisis de datos.

Los ejemplos de este manual están pensados como parte del análisis de casos en cursos prácticos de análisis de datos o en cursos para postgraduados vinculados al área de las organizaciones, la empresa y los recursos humanos, pero espero que resulte fácil a otros profesionales y estudiantes de ciencias sociales y de la salud pensar en como adaptarlos a sus propias variables y necesidades.

Una cuestión que ha de merecer atención práctica inmediatamente es cómo “llamar” al programa desde el sistema operativo, cómo ponerlo en marcha. Esto depende de la versión del sistema operativo y de la versión del programa, no obstante, en la sección “modos de ejecutar un programa BMDP” (pagina 19) se explican algunas de las situaciones más comunes.