

EL ANÁLISIS DE DATOS EN LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS

Luis Lizasoain Hernández y Luis Joaristi Olariaga*

Universidad del País Vasco-Euskal Herriko Unibertsitatea

RESUMEN

En primer lugar, se aborda la caracterización del análisis de datos en el contexto de la evaluación de programas educativos. Para ello, se parte de la consideración de la evaluación de programas como investigación aplicada, de las etapas de la misma y de los tipos de evaluación.

En un segundo apartado se presentan algunas técnicas concretas como la regresión y discriminación, la segmentación, las técnicas factoriales, las series temporales, las ecuaciones estructurales y, sobre todo, los modelos jerárquicos lineales y el meta-análisis.

Por último, se discute lo relativo a la comunicación y difusión de datos y resultados distinguiendo dos audiencias: el resto de evaluadores y demás miembros de la comunidad científica, y los clientes, interesados o implicados.

ABSTRACT

First of all, the characterization of the data analysis in the context of the educational program evaluation is discussed. For this, program evaluation is considered as applied research, and the stages of the process and the evaluation types are examined.

The second part is devoted to present some techniques such as regression and discrimination, segmentation, the factorial techniques, the time series, the structural equations models and, above all, the hierarchical linear models and the meta - analysis.

Finally, questions related to the data and results communication, presentation and dissemination are discussed focusing the differences between two audiences: the evaluators, members of the scientific community, and the clients, users or stakeholders.

* Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación. Avda. de Tolosa 70. 20018 San Sebastián. E-mail: plplihel@sf.ehu.es

CARACTERIZACIÓN DEL ANÁLISIS DE DATOS

A nuestro entender, una manera adecuada de descubrir las características específicas del análisis de los datos es situarlo en el contexto general de la evaluación de programas y, a partir de ese punto, comenzar la búsqueda de las notas distintivas que lo caracterizan.

Y, por extraño que pueda parecer, hay una cuestión en la que todos los autores básicamente coinciden, y es en el hecho de considerar la evaluación de programas como una actividad investigadora.

En consecuencia, si asumimos lo antedicho, al análisis de datos le correspondería la tarea que habitualmente se cita cuando se habla de esta fase del proceso de investigación empírica en ciencias sociales y humanas. A saber: proporcionar evidencia suficiente o respuesta a las hipótesis y preguntas planteadas mediante técnicas analíticas (habitualmente estadísticas en el paradigma cuantitativo) transformando los datos en resultados.

Pero evidentemente la cuestión no es ni mucho menos tan sencilla. La evaluación de programas es una actividad investigadora, pero evaluación e investigación no son términos intercambiables. Un primer matiz es el de considerar la evaluación de programas como un caso típico de investigación aplicada hablándose también de investigación evaluativa.

La evaluación de programas como investigación aplicada

De la contraposición entre la investigación básica y la aplicada surgen bastantes de las notas distintivas que habitualmente se asignan a la evaluación de programas. Desde la perspectiva del análisis de datos, señalemos las siguientes:

En primer lugar, y como marco general de referencia, el hecho de que el objetivo primordial de la investigación aplicada no es la búsqueda de la *verdad* o la producción de conocimiento, sino el uso, **el empleo de dicho conocimiento para la acción, la intervención, en definitiva, la mejora**. Y esta última referencia a la mejora cobra especial significación en el caso de la evaluación de programas educativos (De la Orden, 1985; Tejedor, 1994; Pérez Juste, 1995; De Miguel, 1999).

De este carácter aplicado se derivan consecuencias importantes a la hora de plantearse la fase del análisis de datos. Así, las técnicas analíticas que se empleen deben estar claramente orientadas a proporcionar respuesta y evidencia a las preguntas y cuestiones que se hayan planteado.

Pero además, en el caso de las técnicas estadísticas, esta respuesta no basta con que sea estadísticamente significativa, ha de tener **significación sustantiva** (Botella y Barriopedro, 1995: p. 197) o *práctica* (Bickman y Rog, 1998: p. xi).

En investigación evaluativa hay que saber si los efectos tienen la magnitud suficiente como para considerarlos plenos de significado, no simplemente estadísticamente significativos.

Pero es que, además de que las técnicas analíticas nos han de proporcionar resultados *sustantivamente significativos*, en evaluación de programas hay que afrontar el hecho de que dichos resultados tienen que dar clara respuesta a las cuestiones planteadas.

O dicho en otros términos: esa significatividad sustantiva, ¿garantiza, aunque sea con un cierto margen, que el programa ha producido efectos? (Bickman y Rog, 1998: p. xiii). ¿Podemos hablar de **causalidad** en la evaluación del impacto, (Apodaca, 1999), o de atribución causal de los efectos observados (Gaviria, 1999)? ¿Dónde ponemos el énfasis, en la validez externa o en la interna?, ¿Qué grado de control es deseable o simplemente viable?

Existe acuerdo en que en la investigación aplicada, los estudios cuasi-experimentales son los más frecuentes. Desde la perspectiva de la evaluación de programas educativos, el empleo de este tipo de diseños supone muchas veces el máximo grado de control alcanzable.

En el fondo, todos estos términos, todas estas preguntas, nos remiten a la **generalizabilidad** —o en término de Cronbach, **extrapolabilidad**— de dichos resultados.

Evidentemente en todas las dicotomías que aquí planteamos subyace un continuo de forma que las respuestas a estas preguntas varían en cada caso. Pero a nadie se le escapa que todas ellas tienen importantes repercusiones a la hora de plantearse el análisis de los datos.

Pasemos a otro asunto. En la introducción de la obra de Bickman y Rog que ya hemos citado, la tercera diferencia metodológica entre la investigación básica y la aplicada es formulada bajo el epígrafe «multiple vs. single level of analysis» y en la misma podemos leer que «the applied researcher, in contrast to the basic researcher, usually needs to examine a specific problem at more than one level of analysis, not only studying the individual, but often larger groups, such as organizations or even societies». (Bickman y Rog, 1998: p. xiv).

Si esta afirmación es válida en la investigación aplicada en ciencias sociales en general, en lo tocante a la evaluación de programas educativos constituye casi un axioma.

De la Orden (1999), hablando de los modelos jerárquicos lineales afirma que «Constituyen un marco de análisis natural para el tratamiento de datos con una estructura multinivel o anidada, situación muy común dentro de la investigación educativa» (p. 425).

Más adelante tendremos ocasión de detenernos con más detalle en estos modelos. Baste ahora señalar esta **estructura jerárquica o anidada** como una de las notas distintivas más importantes a la hora de caracterizar el análisis de los datos de las evaluaciones de programas educativos.

Llegados a este punto quizá alguien esté ya planteándose si cuando hablamos del análisis de datos nos referimos exclusivamente al análisis *estadístico* de datos. Avanzamos que rotundamente la respuesta es no.

Si hay una familia semántica que configura una de las notas características de la evaluación de programas, es la que se estructura alrededor de términos como *complementariedad metodológica, enfoque multimétodo, «multiplismo», debate cualitativo-cuantitativo, triangulación, superación del debate de paradigmas, etc.*

No vamos a extendernos en este punto puesto que en este mismo número hay un excelente artículo dedicado exclusivamente a este asunto. Nos limitamos a constatar aquí que si algo caracteriza a la metodología de la evaluación de programas es esta cuestión. Prácticamente todos los autores que han dedicado unas líneas a la evaluación de programas coinciden en señalar **la complementariedad como enfoque más**

acertado de forma que el análisis de los datos debe llevarse a cabo empleando de forma complementaria o simultánea **múltiples enfoques, métodos y técnicas**. Con respecto a los paradigmas, inmejorable resumen lo constituyen las normas elaboradas por el *Joint Committee* (1994) cuando al tratar del análisis de la información, ofrecen dos normas de idéntico texto con la única diferencia de que la A8 se refiere a la información cuantitativa y la A9 a la cualitativa.

Somos conscientes de que el asunto tiene en cualquier caso sus complicaciones: para algunos, el hablar de *múltiples enfoques* supone que es deseable o incluso necesaria la complementación mutua entre el enfoque cuantitativo y el cualitativo. Otros con ese término o con el de triangulación hacen referencia al uso de variedad de técnicas *dentro de un solo paradigma*. Es, por ejemplo, el sentido que Cook otorga al término «*multiplismo*» y que como tal se adopta a lo largo de la obra de Fernández Ballesteros (1995) asumiendo un enfoque cuantitativo.

Como muestra del uso de técnicas tanto cualitativas como cuantitativas a la hora de analizar los datos valgan los siguientes ejemplos:

Freixa y otros (1996) emplean técnicas estadísticas para datos textuales al realizar el análisis de contenido de las respuestas a entrevistas semi-estructuradas. Sogunro (1997) combina técnicas estadísticas clásicas con el análisis de contenido de los datos cualitativos. Yin (1998) al tratar del estudio de casos expone la posibilidad de emplear unidades de análisis anidadas como muestra de la factibilidad de simultanear un estudio cualitativo y cuantitativo. Fetterman (1998) al hablar de los análisis propios de los estudios etnográficos (pp. 494-498) cita entre otras la triangulación, elaboración y comparación de patrones, los sucesos-clave, el uso de mapas y gráficos, el análisis de contenido, la cristalización y la *estadística* especialmente las técnicas no paramétricas. Otro ejemplo es el artículo de Knight y Kuleck (1999) titulado *The value of multimethod qualitative/quantitative research methodology in an educational program evaluation: a case study*. Hemos transcrito el título completo porque pensamos es lo suficientemente claro como para requerir comentario adicional alguno. Por último, González Ramírez (2000) realiza análisis estadísticos (descripciones univariadas y análisis de varianza) sobre los resultados de los alumnos y análisis de contenido para evaluar los resultados desde la perspectiva de los profesores.

Sin negar la complejidad que el asunto tiene, con estos pocos ejemplos hemos querido simplemente mostrar que el mestizaje (usando un término muy en boga) metodológico es una de las características más específicas del análisis de los datos en investigaciones evaluativas.

La última nota distintiva que nace de la contraposición entre investigación básica/investigación aplicada, se refiere a los agentes que intervienen en ésta.

Simplificando un tanto las cosas, en un esquema investigador tradicional habitualmente se distingue sólo entre un agente (el investigador o equipo investigador) y un paciente (los sujetos o unidades investigadas). (Ciertamente no hay que olvidarse de los que evalúan y financian —o no— las investigaciones, pero éste es otro asunto). Aquí el proceso se desarrolla bajo el control del investigador y los resultados se comunican a otros posibles agentes mediante la difusión por los canales habituales de comunicación de la comunidad científica formada por dichos agentes.

Por el contrario, en la investigación aplicada, y más en la evaluación de programas, es extremadamente frecuente que a este esquema haya que añadir un elemento más. Los términos varían mucho, pero los más frecuentes son: *clientes*, *audiencias*, *interesados*, *implicados*, *participantes*, etc.

La distinción terminológica es importante porque trasluce muy diferentes enfoques en la forma en que se aborda el proceso evaluativo y el papel asignado a los diferentes participantes en el mismo.

No es lo mismo el típico «encargo» que un evaluador profesional realiza como evaluador externo, que un proceso evaluador participativo en el que el hipotético papel del evaluador sea el de facilitador, animador o coordinador de un grupo de implicados que son parte activa y los verdaderos agentes de la, en este caso, autoevaluación.

Entre estos dos extremos se dan muchas situaciones intermedias en las que es frecuente que los interesados participen en algunas fases del proceso de evaluación, y en el caso de las evaluaciones educativas, el énfasis puesto en la mejora recomienda optar por evaluaciones internas realizadas por los mismos profesores (Pérez Juste, 1995: p. 112). Pero no es menos cierto que frecuentemente se realizan evaluaciones externas y que, sobre todo, las condiciones muchas veces aconsejan un enfoque mixto.

De una manera o de otra, independientemente de lo participativo que se pueda considerar el enfoque adoptado, lo importante es hacer notar aquí que tanto alguna de las fases del proceso evaluador como los resultados del mismo han de ser comunicados y compartidos con el conjunto de **agentes implicados**.

En el caso concreto de los resultados, esto supone que han de ser transmitidos a dos audiencias muy distintas: al resto de la comunidad científica, pues como investigación que es debe ser difundida siguiendo los canales y normas al uso; y a los *clientes*, *implicados* o *interesados*, y aquí el asunto cambia porque este segundo tipo de audiencia no tiene por qué compartir (de hecho normalmente no lo hace) ni el lenguaje ni el soporte técnico que subyace al proceso evaluativo.

Y el que esta comunicación sea eficaz es crucial para la evaluación de programas, pues no se trata de una cuestión accesoria nacida de un deseo o prurito de difundir o comunicar al máximo los resultados. Por el contrario, saber comunicar a los interesados los resultados del análisis es objetivo fundamental si queremos que la evaluación produzca efectos, sea de utilidad para la acción y los procesos de mejora.

Pensamos que las hasta aquí vistas son —desde la perspectiva del análisis de datos— las notas distintivas que emanan de la consideración de la evaluación de programas como investigación aplicada. Ahora, para completar este intento de caracterización, centremos nuestra atención en el proceso evaluativo en sí: en sus etapas y fases y en los tipos de evaluación que de las mismas se derivan.

Las etapas del proceso y los tipos de evaluación

Dado que la evaluación de programas es una actividad investigadora, la consecuencia que de ello se deriva es que las fases de la misma no han de diferir excesivamente de las propias del proceso investigador.

Y de hecho así ocurre. Si examinamos en la bibliografía esta cuestión, o sin ir muy lejos, si observamos el índice de este número monográfico, nos daremos cuenta que casi siempre, después de las decisiones sobre el diseño y de la etapa de recogida de información, aparece la fase del análisis de los datos, a la que suele seguir la elaboración del informe y la puesta en práctica de las conclusiones.

Y, por descontado, nada hay que objetar a esta secuencia en la medida en que refleja el orden lógico del proceso investigador. Los matices suelen surgir en el número de etapas que se distinguen, o en las diferencias según los enfoques de la evaluación.

Pero desde nuestra particular perspectiva centrada en la fase del análisis de los datos, nos interesa señalar que, en primer lugar, las decisiones sobre qué técnicas emplear para analizar los datos van a estar a su vez, lógicamente en función de las decisiones que en las fases previas se hayan ido tomando: qué enfoque se ha dado a la evaluación (cuantitativo/cualitativo/mixto), qué preguntas o cuestiones se plantean, qué diseños se han empleado, qué unidades se han definido, cómo han sido seleccionadas y —si ha sido el caso— asignadas, qué tipo de información y mediante qué instrumentos se ha recogido.

Segundo, si bien esta secuencia es lógicamente consistente, no debe ser tomada al pie de la letra en el sentido de que después de recoger la información, ésta se organiza y se analiza dando respuesta a las preguntas planteadas. Evidentemente el proceso no es ni mucho menos tan lineal porque el análisis de los datos es un **proceso abierto, cíclico e interactivo**: los resultados de los análisis preliminares condicionan los siguientes y así sucesivamente, a menudo modificando las previsiones iniciales.

Este carácter abierto de la fase del análisis es válido y aplicable a cualquier proceso de investigación, aunque lógicamente en mayor grado cuanto más exploratorio sea el enfoque adoptado. Pero en el caso de la evaluación de programas cobra especial relevancia habida cuenta del papel que los agentes implicados o interesados pueden jugar en esta fase. Cuanto más abierto y participativo sea el enfoque, más interactivo y cíclico es el análisis de los datos.

Pero además de esta ordenación lógica basada en las fases del proceso investigador, también es posible distinguir etapas en función del desarrollo, de la secuencia del propio programa a evaluar. Aquí, en función de la fase del proceso en que se centre la actividad evaluadora, nos podemos encontrar con distintos tipos de evaluación: evaluación de necesidades; evaluación del diseño; evaluación de la evaluabilidad; evaluación de la implementación; evaluación de la cobertura; evaluación del proceso; evaluación de resultados, impactos o producto; y, evaluación económica.

Desde nuestra habitual perspectiva centrada en el análisis de los datos, en primer lugar señalemos que el análisis de datos encuentra su lugar, digamos, natural en la fase o contexto de la **evaluación de resultados** o de impactos.

Segundo, que al igual que antes, las técnicas analíticas a emplear estarán también en función de los resultados y de las consiguientes decisiones que se hayan ido tomando en evaluaciones previamente realizadas (del proceso, por ejemplo).

Tercero, que las técnicas analíticas pueden emplearse —y de hecho se emplean— en muchas fases o tipos de evaluación, *y no sólo en la evaluación de resultados o de impactos* a pesar de que algunos consideren ésta la fase analítica por excelencia. Lógicamente

te también, las técnicas analíticas a emplear estarán en función del tipo de evaluación en cuyo contexto se lleven a cabo. Por ejemplo, Sánchez y Anguera (1993) analizan las posibilidades que la técnica del PERT brinda en la fase de la implantación y planificación de la evaluación de programas.

En el ámbito de la evaluación económica, ejemplos recientes de lo dicho encontramos en Francis y otros (1999) que llevan a cabo un análisis de costo-eficacia de un programa de tratamiento de drogadicciones, o en Plotnick y Deppman (1999) que abordan los análisis de costo-beneficio en evaluación de programas en general y en programas de prevención del maltrato infantil en particular.

En lo relativo al análisis de datos en la evaluación de necesidades, destaca aquí lo referente al manejo y análisis de datos secundarios. A saber: indicadores, estadísticas oficiales, etc.

Se trata de instrumentos fundamentales para la evaluación y tanto si son empleados en el análisis de necesidades como en la estimación de los impactos de un programa, el evaluador puede optar por elaborar un sistema de indicadores *ad hoc* o por emplear algunos sistemas de indicadores ya elaborados por diversidad de organismos e instituciones. Lázaro (1992) analiza las condiciones que los mismos deben reunir, compara los de diferentes organismos y examina el caso específico de los indicadores en educación. Cornejo y Martínez (1992) estudian las posibilidades que brinda el manejo sistemático de fuentes secundarias de información en evaluación de programas de intervención social y ofrecen una muy interesante aplicación del análisis de correspondencias a los datos del censo de población.

Por último, con respecto a la evaluación de resultados, no podemos por menos que asumir que éste sigue siendo el tipo de evaluación o la fase del proceso evaluativo en que se cumple el objetivo clásico del análisis de los datos: verificar, con un grado de confianza aceptable, que el programa permite obtener los resultados planteados.

Todas las cuestiones que anteriormente hemos apuntado relativas a la causalidad, la validez interna, el control, los diseños, la integración metodológica, etc. encuentran aquí perfecto acomodo.

Botella y Barriopedro (1995) apuntan cuatro funciones principales del análisis de datos en la valoración del impacto: descripción comprensiva de datos brutos; valorar si las relaciones detectadas son o no significativas; estimación de magnitudes y diferencias; y, estimar la sustantividad de los efectos.

Desde esta perspectiva, muchas son las técnicas analíticas que aquí se pueden emplear y, centrándonos en las técnicas estadísticas, más adelante trataremos algunas de ellas. Pero no queremos dejar pasar este punto sin justificar el por qué en dicha exposición no vamos a tratar de los análisis descriptivos básicos y de las técnicas inferenciales —paramétricas y no paramétricas— clásicas.

La razón es evidente: creemos que son cuestiones suficientemente conocidas y perfectamente desarrolladas por prestigiosos autores en obras y manuales ampliamente difundidos.

Y ésta es la única justificación. Queremos dejar claro que su ausencia no implica que estén en desuso (muy al contrario como ahora veremos) por ser inadecuadas, excesivamente sencillas o simplemente pasadas de moda. Asumimos además que

estas dos últimas pretendidas *razones* no son tales, pero, lamentablemente, pensamos que hay que incluirlas pues todavía a estas alturas hay quien se rige por criterios de moda o actualidad o quien piensa que la sencillez es un defecto.

Como muestra de su uso y adecuación, en las siguientes referencias encontramos estudios evaluativos con diseños pretest-postest en que el análisis de los datos se efectúa mediante pruebas paramétricas y no paramétricas de contraste de medias: Hernández, Aciego y Domínguez (1994), Echeburua y otros (1996), Herráiz (1998), Pérez y Salvador (1998), y Filella (2000).

ALGUNAS TÉCNICAS ESTADÍSTICAS PARA EL ANÁLISIS DE DATOS

Hecha esta relación de notas distintivas y de características deseables del análisis de datos en la evaluación de programas educativos, a continuación vamos a comentar someramente algunas técnicas y procedimientos de análisis que pensamos reciben o deben recibir un tratamiento especial.

Se trata, como no podía ser de otra manera, de una selección personal responsabilidad de quienes esto firmamos, y, por tanto, sesgada. En algunos casos hemos incluido técnicas o procedimientos que consideramos de presencia inexcusable como lo relativo al meta-análisis o los modelos jerárquicos lineales. Muy probablemente todo el mundo podrá coincidir en ello. Quizá no ocurra lo mismo en otros casos en que su aparición obedece a razones más parciales: nuestra personal opinión o nuestra experiencia investigadora.

Como ha quedado dicho, no vamos a ser exhaustivos; lo que a continuación vamos a hacer es exponer aquellas particularidades de las técnicas más generales en el sentido de su aplicabilidad a problemas de muy diversos tipos.

No se puede decir que hay técnicas específicas de análisis de datos para la evaluación de programas (Tejedor, 1994). Lo que sí es cierto es que en función del tipo de evaluación algunas son más apropiadas que otras, de manera que en algunos son más convenientes técnicas asociadas a diseños transversales y en otros, longitudinales.

Así, si en algún caso se requiere de «jueces» son útiles técnicas como el escalamiento multidimensional; o el análisis de datos textuales para el tratamiento estadístico de cuestiones abiertas, suponiendo que se pretende superar lo simple con un contenido exploratorio considerable. Si, por otra parte, algún aspecto de la evaluación se desarrolla a lo largo del tiempo, la diversidad de técnicas cuantitativas es amplia, empezando desde las técnicas asociadas a los diseños experimentales y cuasi-experimentales, con sus pruebas paramétricas y no paramétricas de comparación de dos o más grupos, pasando por modelos de medidas repetidas, series temporales, tablas múltiples, etc.

A la postre, el diseño de la evaluación también depende de las herramientas de que se disponga para el análisis. Es cierto que quien lo desconoce todo sobre la metodología cuantitativa suele, por lo común, plantear diseños cuyos datos no hay quien analice. Pero también ocurre lo contrario: a veces se recurre a diseños muy simples en los que se pierde voluntariamente la complejidad de la realidad.

Nuestra idea no es desarrollar los distintos tipos de técnicas sino, con el fin de ser útiles, centrarnos en los que son más generales. Otros criterios que tenemos presentes

son la sencillez de interpretación, así como que tengan condiciones de aplicación poco restrictivas, pues ya sabemos lo frustrante y cansado que es el análisis de los residuos.

Análisis discriminante, regresión logística. Regresión y discriminación no paramétricas (técnicas de segmentación)

Desde una perspectiva paramétrica, si la variable dependiente es dicotómica y las independientes son cuantitativas, podremos aplicar las dos primeras técnicas citadas. Si se plantean dudas acerca de si aplicar una o la otra, no está de más recordar que siempre es más útil disponer de diversos puntos de vista por lo que lo recomendable es realizar ambos y comparar sus resultados. Pero en tal caso conviene tener presente que el análisis discriminante precisa de condiciones de aplicación ante las que la regresión logística es más robusta (en ésta sólo hay que tener precaución con la colinealidad). Un estudio detallado de ambas técnicas se puede encontrar en Hair, Anderson y otros (1999). Hannan y Murray (1996) llevan a cabo una simulación con métodos Monte Carlo comparando el rendimiento de los análisis lineales y logísticos.

Se trate de regresión o de análisis discriminante, lo habitual es utilizar sus versiones paramétricas, tanto si se trata de explicar y predecir como si el fin es explorar. Cuando los datos de una evaluación son no experimentales es preciso explorar. Y en este caso se dispone de un conjunto de técnicas, libres de restricciones y, aunque con un proceso de computación tedioso, de una gran sencillez de interpretación. Nos estamos refiriendo a la segmentación. La variable dependiente (única, pues no hay generalización multivariante) puede ser de cualquier naturaleza así como las independientes o predictoras. Consiste, en pocas palabras, en dividir progresivamente la muestra en clases, pero está más relacionada con la regresión y la discriminación que con las técnicas de clasificación, pues se trata de encontrar los grupos de casos que mejor expliquen las modalidades de una variable cualitativa o los valores de una cuantitativa.

Las técnicas de segmentación son muy usadas en estudios de mercado. En nuestro campo, a pesar de las ventajas que hemos apuntado, no lo son tanto. Una excepción es el trabajo de Repetto y otros (1994) donde emplearon el modelo CHAID en el análisis de los datos de la evaluación de un programa de mejora de la comprensión lectora.

El análisis factorial y las tablas múltiples

La segmentación es una forma de explorar considerando las variables dependientes de forma aislada. De alguna manera se pierde la multidimensionalidad de la realidad. Por ello convendría recurrir a los modelos exploratorios por excelencia. Esto es, el análisis factorial (análisis factorial en factores comunes y específicos así como componentes principales y análisis factorial de correspondencias).

Dada la naturaleza de las variables a considerar en todo fenómeno social o educativo y siguiendo a los ya citados Cornejo y Martínez (1992), el análisis de correspondencias es una eficaz herramienta para la descripción y evolución en el tiempo de los fenómenos macrosociales implicados en la evaluación de programas. Para ilustrar dicha afirmación, aplican el análisis de correspondencias a los datos del censo de pobla-

ción operando con once índices de riesgo de desadaptación social en la infancia extraídos de dichas fuentes con el fin de caracterizar las situaciones de riesgo potencial.

Otros estudios evaluativos en que se emplean estas técnicas son los realizados por Carballo (1990), y por los también previamente citados Freixa y otros (1996); y Kreft (1998) que usa la terminología sajona refiriéndose al análisis de correspondencias como análisis de homogeneidad.

Pero además de la perspectiva apuntada, hay que tener en cuenta que en alguna etapa de una evaluación nos vamos a encontrar con varias tablas de datos, dado que va a haber un antes y un después; eso cuando menos, pues nos podemos encontrar con una serie temporal. Cada tabla o grupo de variables puede corresponder a instantes o contextos diferentes (Lebart, Morineau, Piron, 1995).

Según los análisis factoriales clásicos, en que todas las tablas se analizarían conjuntamente o de forma yuxtapuesta, no se podrían estudiar globalmente las estructuras, tanto comunes como específicas de estos grupos de variables, ni construir tipologías de los individuos ni de las variables. Bajo este enfoque multivariante asociado al análisis factorial, sugerimos que la compleja realidad evolucionando a través del tiempo será más adecuadamente analizada por medio del análisis de tablas múltiples. Se analiza un conjunto de individuos caracterizado por varios grupos de variables (por ejemplo el pretest y el postest tratados multidimensionalmente; bajo el análisis factorial de las tablas yuxtapuestas puede ocurrir que las tipologías de los individuos estén construidas sólo según unas cuantas variables del pretest y, sin embargo, nos interesa analizar cómo evolucionan los individuos a través del tiempo, caracterizándolos por variables de ambos grupos). Al no tratar las tablas de forma exclusivamente conjunta, permite representar los individuos según cada grupo de variables, descubrir qué tablas (o grupos de variables) se asemejan, proporcionar una tabla que sintetice el conjunto y describir la dispersión entre tablas con relación a la sintética y determinar el origen de tales desviaciones. Para profundizar en estas técnicas cabe mencionar la obra de Escofier y Pagès (1990).

Pero además de estas técnicas, hay modelos creados específicamente para estudiar la evolución de un fenómeno a lo largo del tiempo.

Diseños de medidas repetidas y series temporales

Se plantean como las alternativas más intuitivas, dado que un programa se desarrolla en el tiempo. En general el desarrollo en el tiempo se basa en un número reducido de momentos.

En los diseños de medidas repetidas, analizados por medio del análisis de la varianza multivariante (Ximénez y San Martín 2000), debido a que a cada sujeto se le aplican todos los tratamientos, no se precisa del control ni de la aleatorización completa del análisis de la varianza, pero al estar inmersos en el modelo lineal general, están sometidos a las típicas condiciones de aplicación, además de que resultan «penosos» si se quiere considerar el contexto y los condicionamientos prácticos.

Por su parte una serie temporal consiste en un conjunto de observaciones de una variable en intervalos regulares de tiempo (Uriel 1992). La problemática de las series

temporales empieza porque se precisa una medición a intervalos regulares de tiempo, que como poco debe ser de 50; así, en los modelos ARIMA se precisa de numerosos instantes de observación, pues en caso contrario hay dificultades en la identificación del modelo. Además, las variables sólo pueden ser continuas. Las series temporales requieren procesos estabilizados y con intervenciones muy limitadas. En nuestro entorno no está de más recordar que la educación no sólo no es un proceso estabilizado, sino que no debe serlo (Sáenz, González, 1990). Pero, al margen de estas consideraciones y limitaciones, muestra del interés que ofrecen son los trabajos de Arnau, Ruiz y Salvador (1988) que examinan la evaluación del tratamiento en diseños conductuales mediante estas técnicas; o el propio Alvira (1991) que dedica un apéndice a discutir la aplicabilidad de estos modelos. Al hablar de la evaluación económica, citamos el estudio de Plotnick y Deppman (1999). En el mismo llevan a cabo un análisis de costo-beneficio de programas de prevención del maltrato infantil y proponen, entre otras, técnicas analíticas basadas en series temporales para la extrapolación de los posibles beneficios de un programa.

De cualquier forma, a la vista de la problemática de estas dos técnicas, proponemos como salidas a este callejón dos modelos que entre sus especificidades también dan cabida a la multidimensionalidad y al tratamiento de diseños longitudinales.

Modelos de ecuaciones estructurales

Según Batista y Coenders (2000), ya que los fenómenos de interés son complejos, tienen muchos aspectos, obedecen a múltiples causas y están frecuentemente medidos con error, identificar el origen de su variabilidad requiere servirse de métodos multivariantes adecuados, como los modelos de ecuaciones estructurales, que permiten incorporar el error de medida y considerar relaciones recíprocas entre constructos.

Al contrario del análisis de la varianza con medidas repetidas, en que sólo es viable una única relación entre variables dependientes e independientes, los modelos de ecuaciones estructurales posibilitan especificar las variables independientes que predicen cada una de las variables dependientes. Operando con datos no experimentales y relaciones lineales se puede entender que surgen como alternativa al modelo lineal general permitiendo además una visión holista.

Fundamentalmente se centra en la especificación del modelo a voluntad y su modificación para mejorar el ajuste a los datos (esto quiere decir confirmatorio: especificar el modelo antes de su estimación); y ello basado en la descomposición de varianzas y de covarianzas. Otro aspecto básico es el tratamiento del efecto de los errores de medida de las relaciones entre las variables, para lo cual hay que basarse en medidas repetidas. Es decir, se pueden contrastar relaciones causales entre variables, así como evaluar la calidad de la medida, en lo que se entiende como modelo de análisis factorial confirmatorio, como paso previo a las especificaciones del modelo.

Una aplicación de este tipo de modelos (en concreto del de crecimiento latente) nos lo ofrecen Wang y otros (1999) evaluando programas de intervención de reducción de conductas de riesgo con respecto al SIDA.

Los modelos jerárquicos lineales

De la Orden (1999), al introducir los trabajos sobre esta cuestión presentados al último congreso de AIDIPE, afirma que «No resulta excesivamente arriesgado afirmar que el principal avance metodológico de los últimos años en investigación en Ciencias Sociales son los Modelos Jerárquicos Lineales» (1999: p. 425).

Ciertamente, constituye una enorme ventaja el poder disponer de un conjunto de técnicas que permiten analizar los datos tomando en consideración la estructura multinivel o anidada de los mismos, habida cuenta que en el ámbito educativo ésta suele ser su conformación habitual, casi diríamos natural.

Este tipo de modelos resulta de aplicación más general, aunque en su aspecto más sencillo no cabe la multidimensionalidad. Se trata de explicar y/o predecir una variable dependiente o respuesta, que en el caso más sencillo es normal. Los individuos están agrupados según uno o más criterios o niveles entre los que se establece, habitualmente de forma natural, una jerarquía. Dado que los sujetos dentro de jerarquías tienden a ser más homogéneos que los obtenidos aleatoriamente de la población, conviene que cada nivel tenga asociado su propio modelo, que representa las relaciones estructurales así como la variabilidad residual dentro del nivel (Bryk, Raudenbush, 1992). Dicho de otra manera, la noción central es que modelos lineales diferentes para el primer nivel o micro-nivel deben ajustarse para cada contexto (macro-nivel) (De la Orden, 1999). Los coeficientes de regresión del primer nivel se analizan como variables aleatorias en el segundo nivel, descomponiéndose la varianza en la de ambos niveles. Se obtienen estimaciones estadísticamente eficientes, y desde el punto de vista práctico, se pueden ordenar los elementos del segundo nivel y superiores en función de la variable respuesta. Para profundizar sobre estos modelos remitimos a Goldstein (1995).

Con los modelos jerárquicos lineales se resuelve el problema de la unidad de análisis, puesto que se incluyen distintos niveles jerarquizados. Las variables asociadas a los distintos niveles jerarquizados afectan a los resultados individuales. Si la unidad de análisis es el individuo, se asignan las características del nivel correspondiente a cada sujeto de él, dando origen a la no independencia y a eludir los efectos de grupo. Al contrario, si se considera que la unidad de análisis es un determinado nivel, se incluyen en él las características medias de los niveles inferiores, perdiéndose así la variabilidad entre esos niveles inferiores. Además, los resultados del modelo lineal general dentro de cada elemento del segundo nivel, por ejemplo, no son generalizables más que si tales elementos constituyen toda la población.

Entre los distintos aspectos ventajosos que el modelo jerárquico lineal presenta cabe mencionar que precisa de menos condiciones de aplicación en comparación con los modelos clásicos. Además proporciona modelos más intuitivos, pues refleja el proceso de muestreo, que a menudo es por conglomerados. Es una alternativa a la técnica clásica del análisis de la varianza con medidas repetidas, que precisa de las mismas mediciones para cada sujeto y un cierto balance en la estructura; los sujetos se incluyen en el segundo nivel y las distintas mediciones en el primero, esto es, se da un anidamiento dentro de los sujetos; los resultados obtenidos con el modelo jerár-

quico lineal son más sencillos de interpretar y de comunicar. Se pueden tratar datos de respuesta discreta y variables de respuesta cuantitativas y cualitativas simultáneamente.

Si de cada individuo se miden diversas variables de respuesta (modelos multivariantes), se sitúa al individuo en el segundo nivel y las variables de respuesta en el primero; se da un correcto tratamiento a los datos ausentes, que en muchas ocasiones se deben al propio diseño, pues hay individuos en que sólo se puede medir un subgrupo de variables respuesta; los resultados son fácilmente interpretables, en comparación con los de los análisis multivariantes basados en el modelo lineal general multivariante.

Con todo lo dicho, no es de extrañar que aquí sí sea fácil encontrar abundantes referencias. Los interesados encontrarán especialmente útil la página Web que sobre el proyecto multinivel tiene el Instituto de Educación de Londres. Su dirección es <http://www.ioe.ac.uk/multilevel/> y en la misma encontrarán abundante información bibliográfica, información sobre programas, ficheros de datos, ejemplos, etc.

Pero junto a esto, tanto aplicaciones de estos modelos como estudios de índole más metodológica podemos encontrar en los siguientes artículos:

El ya citado trabajo de Kreft (1998) que efectúa un re-análisis de investigaciones evaluativas sobre prevención del consumo de drogas usando, además del ya citado análisis de correspondencias, los modelos multinivel. Krull y MacKinnon (1999) realizan un estudio de simulación centrado en las variables mediadoras. Gaviria (1999) analiza las posibilidades del enfoque multinivel en la evaluación de sistemas educativos y Murillo (1999) hace lo propio en el ámbito de los estudios sobre eficacia escolar.

Además, cada vez es más frecuente encontrar aplicaciones de estos modelos al meta-análisis: McNeal y Hansen (1995) realizan un estudio meta-analítico de las diferentes evaluaciones de los programas DARE usando técnicas multinivel. Desde un punto de vista más metodológico, en Castro y De la Orden (1998) y en Castro (1999) se analizan y comparan los diferentes modelos de meta-análisis y se estudian las aportaciones que a los mismos pueden prestar los modelos jerárquicos lineales.

Es habitual encontrar evaluaciones que consisten en distintas réplicas de un programa, ya que han sido implementados en diversos lugares y/o a distintos grupos, así como en distintos períodos de tiempo. Por otra parte, también es habitual aplicar un programa a grupos reducidos, lo que origina la práctica imposibilidad de rechazar la hipótesis nula. Todo esto es suficiente para darle al meta-análisis un lugar específico en el análisis de datos en evaluación.

El meta-análisis

El meta-análisis (Hedges y Olkin, 1985) es un conjunto de técnicas estadísticas de integración de los resultados de distintas investigaciones sobre un mismo problema. Se trata de buscar la consistencia de los mismos, de examinar las posibles contradicciones, de buscar patrones comunes, o, sin más, de resumir sus resultados cuantitativos. Para ello se parte de utilizar una métrica común entre ellos. Más técnicamente es el

análisis de un conjunto de estimaciones independientes con distribución normal y varianza conocida.

El proceso es análogo al de una investigación primaria. Se definen los objetivos y se clarifican los criterios de selección de estudios y datos. Tras codificar la estructura de las investigaciones en función de los objetivos de la revisión y para evitar la ausencia de validez, se unifica, como se ha mencionado, bajo una métrica común, habitualmente el tamaño del efecto.

De entre los distintos enfoques, no excluyentes, destacamos que en el basado en el *Escrutinio* el problema es que se confunden el efecto de la intervención y el del tamaño de la muestra, además de que no es considerado el impacto del programa evaluado. En el *Clásico o de Glass* (Glass, McGaw y Smith, 1981) se caracteriza cada resultado (como unidad de análisis) de cada estudio por el tamaño del efecto. La crítica se basa en que se da más peso a los estudios con más resultados, los datos no son independientes, pues cada estudio puede traducirse en distintas unidades de análisis y la fiabilidad de los resultados está condicionada por la inclusión de estudios metodológicamente defectuosos. En lo que se refiere a la conocida como efecto del estudio, que incluye criterios de selección de estudios, éstos quedan caracterizados por el tamaño del efecto; como aspectos negativos cabe mencionar los sesgos en la inclusión de estudios y el reducido número de unidades de análisis. Otro enfoque es el de las *Pruebas de homogeneidad*, desarrolladas para determinar si la varianza entre tamaños del efecto es aleatoria. El *meta-análisis Psicométrico* (Hunter y Schmidt, 1990), según el cual se recogen todos los estudios independientemente de su calidad y se corrige la distribución de los tamaños del efecto según una serie de factores que influyen sobre ella, presenta la limitación de que de cada estudio se precisa la suficiente información, no siempre disponible, para proceder a la corrección mencionada. Por último en el *Modelo jerárquico lineal*, como enfoque más novedoso, se considera que la organización de los datos meta-analíticos responde a una estructura jerárquica o anidada. (Castro, 1999) (el anidamiento se produce dentro de los estudios).

Por último, mencionemos las dificultades genéricas en el meta-análisis. Así, cabe indicar la relativa a la selección de estudios, en que el debate se establece en cuanto a si sólo incluir solamente los publicados o en función de su calidad; por otra parte surge el problema de los estudios con resultados ausentes en que las alternativas son eliminarlos totalmente contra proceder a imputación. También la presencia de múltiples tamaños del efecto en un mismo estudio, lo que da lugar a la no independencia, es resuelto de diversas formas, contradictorias entre ellas. En lo que se refiere a la magnitud del efecto, la discusión es la de siempre: a partir de qué valor un número es pequeño y a partir de cuál grande; se barajan distintas alternativas que se pueden encontrar en Cooper (1981), Cohen (1988) y Lipsey (1990).

Tal y como apuntábamos al inicio de este apartado, somos conscientes de que hemos dejado al margen otros modelos y técnicas de interés, como el escalamiento multidimensional, el logarítmico lineal, el análisis conjunto, etc.; pero creemos que con los que hemos presentado, bien exponiéndolos, bien señalando sus limitaciones, el objetivo inicialmente marcado queda, a nuestro juicio, suficientemente cumplido.

DIFUSIÓN Y COMUNICACIÓN DE LOS RESULTADOS DE LA EVALUACIÓN DE PROGRAMAS

Si repasamos algunas de las notas distintivas con que hemos caracterizado a la evaluación de programas educativos, e incluso si examinamos algunas de las técnicas analíticas que hemos presentado (como lo tocante al meta-análisis), podemos darnos cuenta que las cuestiones relativas a la difusión y comunicación de los datos y resultados de una investigación evaluativa constituyen un aspecto crucial del proceso.

Aquí hemos introducido dos términos a los que, a pesar de su evidente vecindad semántica, asignamos significados muy distintos. Se trata de los de *difusión* y *comunicación*. La diferenciación básica que establecemos es la siguiente:

Por *difusión* entendemos la presentación del proceso, los datos y resultados de la evaluación al resto de colegas, de miembros de la comunidad científica. Esta *diseminación* (traducción literal del término inglés) es llevada a cabo fundamentalmente a través de las revistas científicas especializadas.

Por *comunicación* entendemos la presentación de los resultados de la evaluación, y de cualquier otra información que se considere relevante, a los clientes, interesados, implicados, etc. Esta presentación suele efectuarse de forma oral en una o varias sesiones e igualmente suele elaborarse un informe escrito.

Queda pues claro que la distinción básica que establecemos entre estos dos términos se basa en los destinatarios o audiencias de la información. La asignación de dichos términos es poco menos que arbitraria aunque *difusión* o *diseminación* se suelen asociar a la literatura científica, de ahí su elección por nuestra parte. En cualquier caso, el significante elegido es lo de menos. Lo importante son las consecuencias que se derivan de tener que dirigirse a dos audiencias tan distintas como las que acabamos de apuntar. Y en ellas vamos a centrarnos ahora.

Difusión de los resultados

Tal y como en distintas ocasiones hemos afirmado, la evaluación de programas puede ser considerada como un caso o tipo específico de investigación aplicada. Como tal, sus resultados han de ser difundidos entre los miembros de la comunidad científica.

Y esta labor de difusión e intercambio se realiza mediante los canales habituales: de una parte las reuniones científicas, seminarios, congresos, etc., y de otra, las publicaciones en revistas especializadas. El contenido y formato de las mismas está suficientemente tratado en los manuales de investigación e incluso los editores o las sociedades científicas y profesionales establecen y publican normas al respecto.

Poco más tendríamos que decir al respecto si no fuese por una cuestión: como ya hemos tenido ocasión de comentar, en evaluación de programas la acumulación y contrastación de evidencia, o en general las labores de re-análisis, meta-análisis o metaevaluación, tienen una importancia especial. Y como ya hemos apuntado también, para llevar a cabo este tipo de análisis de segundo orden a veces no basta con la información que suele aparecer publicada.

Dicho en otros términos, en evaluación de programas —como en otros muchos ámbitos de la investigación en ciencias sociales y humanas— hay ocasiones en que, además de la información contenida en los informes de investigación en sus diferentes formatos, *es preciso o deseable disponer de acceso a los datos*.

De la Orden y Martínez de Toda (1992) insisten en esta cuestión al hablar del análisis secundario como forma más completa de metaevaluación y añaden que la disponibilidad de los datos para el re-análisis es un indicador de credibilidad del evaluador (p. 519).

Pues bien, esta disponibilidad de los datos plantea cuestiones como las siguientes: el soporte o canal mediante el cuál los datos son accesibles, el formato y estructura de los mismos, y, por último, las garantías de confidencialidad y la ética de la investigación.

Con respecto al soporte parece claro que actualmente las redes informáticas constituyen el medio más idóneo para garantizar la accesibilidad de los datos a otros investigadores. De hecho cada vez es más frecuente encontrarnos en Internet con ficheros y conjuntos de datos disponibles para la comunidad científica. No vamos a detenernos en cuestiones técnicas, pero no está de más subrayar que, ya que se trata de facilitar las cosas a los colegas, el formato de los ficheros debe ser aquél o aquellos que resulten más fácilmente *importables* por los programas y aplicaciones más usados. De esta forma se simplifica la tarea de lectura e incorporación de la información.

Pero evidentemente, no basta con incluir un fichero de datos en una página Web. Para que sean útiles, tales datos han de estar adecuadamente documentados. Y con esto entendemos que ha de explicitarse qué variables son; en qué momento o momentos, cómo y con qué instrumentos la información ha sido recogida; qué transformaciones han sufrido los datos originales; cómo han sido seleccionados los casos; qué posibles niveles de agregación existen; etc. Toda esta información (que algunos denominan *metadatos*) debe añadirse a lo relativo a los resultados que ya hayan sido generados: cuáles se han obtenido y mediante el empleo de qué técnicas.

Por último, en lo que se refiere a las cuestiones de ética de la investigación, señalemos que, además del respeto a los acuerdos establecidos con los clientes o implicados, otro asunto muy importante es el que atañe a las garantías de confidencialidad, pues todo lo hasta aquí apuntado facilita los análisis secundarios, pero plantea graves problemas en lo que a esta cuestión se refiere.

No se trata de un conflicto insoluble, pero hay que prestar especial atención a los mecanismos y procedimientos tendentes a garantizar la confidencialidad bajo estos nuevos supuestos.

En primer lugar, la libre disponibilidad no está reñida con el establecimiento de filtros y niveles de acceso a la información que nos sirvan de razonable garantía de que nuestros datos van a ser usados con fines de investigación. Junto a este tipo de mecanismos propios de la seguridad informática, la propia disposición de los datos y ficheros puede también servir para garantizar la confidencialidad.

Aquí hay que prestar especial atención a aquellas variables que operan, o son susceptibles de operar, como identificadores. El problema no está tanto en los identificadores directos —pues lo normal es quitarlos— como en aquellas variables que tratadas

de forma conjunta pueden ser empleadas para identificar los casos. Según con qué muestras se opere y a qué niveles de agregación, con un cierto conocimiento de la población, esta posibilidad es bastante viable. Sieber (1998) describe, entre otros procedimientos, los que permiten introducir un error aleatorio en los datos de forma que no se interfiere con posibles análisis posteriores pero se dificulta o evita la identificación de casos o unidades.

Comunicación o presentación de resultados

Comenzábamos el apartado anterior enfatizando el carácter investigador de la evaluación de programas con objeto de justificar la difusión de los datos entre la comunidad científica. Ahora aquí procede poner el acento en lo que de *aplicada* tiene.

Y, como ya vimos, de ahí se deriva el hecho de que su objetivo primordial no sea tanto la búsqueda de la verdad o la producción de conocimiento, como el uso, el empleo de dicho conocimiento para la acción, para la mejora. A tal fin, los resultados han de ser presentados y comunicados a los clientes, implicados o interesados. En definitiva, a quienes tienen competencia, responsabilidad o capacidad decisoria de cara a la puesta en marcha de los procesos de intervención o de las acciones de mejora.

Si queremos que la evaluación produzca efectos, que sea realmente usada, es necesario que este proceso comunicativo sea eficaz. Y para ello es crucial tomar en consideración el hecho de que este tipo de audiencias es distinto de la comunidad científica a quien antes hemos hecho referencia.

Tal y como venimos haciendo a lo largo de este trabajo, no vamos aquí a extendernos sobre la comunicación del informe en general, sino que nos vamos a centrar en una parte del mismo: en la comunicación y presentación de *los resultados del análisis*. Y aquí la cuestión básica es conseguir que los mismos sean *inteligibles* para los destinatarios de la información.

Actualmente, el desarrollo metodológico pone a disposición del evaluador complejas técnicas analíticas para procesar los datos. Esta complejidad y versatilidad pueden permitir un mejor tratamiento de los datos, pero a su vez llevan aparejada una mayor dificultad de uso, y lo que más nos importa aquí, de comunicación de los resultados a audiencias no expertas.

Tal y como apunta Brinkerhoff (1983), no se trata, por descontado, de no emplear técnicas complejas con la excusa de que son difíciles de entender. Este autor advierte contra las sofisticaciones *innecesarias*, pero su primera recomendación es no hacer análisis simplistas habida cuenta de la complejidad de la realidad a evaluar.

En definitiva, el reto que tiene el evaluador no es nada fácil: hacer comprensibles a los destinatarios los resultados de los análisis que haya sido necesario emplear por complejos que sean, sin que esto suponga distorsión o pérdida significativa de información. Y éste es un tipo de comunicación que exige un esfuerzo y planteamiento muy distinto al de la difusión entre colegas en que se da por supuesto el conocimiento previo de la audiencia.

En cualquier investigación, después de obtener los resultados éstos han de ser *interpretados* tomando en consideración las características de las técnicas analíticas usa-

das y, sobre todo, confrontándolos con la realidad y el contexto en que se insertan. De esta interpretación, de esta búsqueda de *significado*, surgirán las respuestas a las hipótesis y cuestiones planteadas.

Pero en evaluación de programas, después de esta *primera vuelta* de análisis e interpretación, es necesario *re-analizar* dichos resultados para hacerlos comprensibles en la presentación. Y nótese que hemos empleado el término re-analizar con lo que queremos significar que este trabajo es parte también de la fase del análisis de los datos.

El énfasis que Patton (1997) pone en el uso de la evaluación, hace que las ideas que aporta en lo tocante a estas cuestiones nos resulten especialmente adecuadas y sugerentes. Como ilustración de lo dicho, en el capítulo dedicado al análisis y comunicación de resultados afirma que los resultados de un análisis de regresión pueden ser reducidos a tablas de contingencia y a un conjunto de estadísticos descriptivos de forma que en la presentación se focalice la atención en los resultados importantes.

Este alegato a favor de la simplicidad en la presentación (nada fácil de conseguir por otra parte) es completado con un ejemplo de la información presentada a los clientes respecto a una evaluación en la que, entre otras, se emplearon técnicas factoriales y de regresión para analizar los datos (pp. 317-321). Como conclusión afirma que los implicados no expertos pueden comprender e interpretar datos y resultados cuando son presentados en forma clara y legible usando gráficos y tablas estadísticas simplificadas.

Esto supone que la fase del análisis *no acaba* con la obtención e interpretación de los primeros resultados, de los resultados *brutos* que nos proporcionan los programas informáticos. Es competencia y trabajo (arduo) del investigador re-elaborar y re-analizar dichos resultados de forma que puedan garantizar una *comunicación eficaz*.

Y, además de Patton, son muchos los autores (con Tukey como ilustre precursor) que apuntan que la mejor manera de comunicar eficazmente es usando tablas y, sobre todo, *gráficos*. Pero para que la transmisión visual de información sea eficaz, no basta con apelar a la inteligencia espacial del observador confiando en que, por sí misma, será capaz de deducir de un gráfico los datos subyacentes y las relaciones que entre ellos se establecen. Si están mal diseñados pueden conseguir el efecto contrario: ocultar los datos (esto a veces se hace intencionadamente) o dispersar la atención del observador.

La capacidad de producir fácilmente gráficos cada vez más complejos se ha visto notablemente incrementada por las posibilidades que brindan la mayoría de los programas informáticos. Además de los programas orientados exclusivamente a este fin, la gran mayoría de los paquetes de análisis de datos incorporan potentes módulos gráficos con lo que la labor del investigador se ve notablemente facilitada.

De esta manera, es posible generar gráficos que cubren un amplio espectro de enfoques analíticos: desde la clásica descripción univariada o bivariada, hasta las tendencias, las series temporales o las nubes multivariadas.

Pero al igual que el uso de un paquete estadístico no exime del conocimiento de la estadística, aquí también es preciso conocer las normas y criterios básicos que garantizan una comunicación visual eficaz. En nuestro campo, autores como Henry (1993,

1998) han estudiado el asunto y examinado los distintos tipos de gráficos y las condiciones que deben reunir.

Con respecto al contenido mismo de los gráficos la primera recomendación habitual estriba en que, además del gráfico en sí, en el mismo deben aparecer datos. El objetivo es claro: el gráfico actúa como soporte y la información textual o numérica permite leer e interpretar correctamente la información iconográfica. En conclusión, los gráficos deben ser ricos en información mostrando la complejidad subyacente a los datos. Pero a la vez, conviene no recargar excesivamente cada gráfico pues dificulta su lectura. En vez de intentar mostrar toda la información en un solo gráfico, es mejor emplear varios: por ejemplo, una serie para las descripciones univariadas y otra para mostrar las relaciones que pudieran existir entre variables. En el delicado y difícil equilibrio entre estos dos polos radica una de las claves de una comunicación visual eficaz.

Pero además de las cuestiones de contenido, al hablar de información gráfica no podemos obviar lo tocante a las propias formas, a las cuestiones de formato. En palabras del propio Henry (1998: p. 537), aquí hay cuestiones tan importantes que no conviene dejarlas tal y como aparecen en las opciones *por defecto* pre-establecidas por los programas. Entre ellas cita lo relativo a la orientación, los ejes, las tramas y el relleno o las proporciones. Con el desarrollo actualmente alcanzado por monitores e impresoras, nosotros nos atrevemos a añadir el tratamiento del color. El conocimiento de las normas básicas derivadas de la teoría del color, la complementariedad y la saturación, creemos que contribuirían al menos a evitar gráficos que —por la gama de colores usados— no sólo son de gusto estético más que discutible sino que —lo que es más grave— dificultan cuando no impiden la aprehensión de los datos.

El resto de las posibilidades que brinda la informática gráfica (animación, efectos tridimensionales, etc.) están contribuyendo al auge y desarrollo del campo de la *visualización de los datos*. Por razones obvias no podemos extendernos más en estas cuestiones. Además del ya citado Tukey, la obra de Bertin (1973), *Sémiologie graphique* es todo un clásico en la materia. Junto a ellas, nos permitimos recomendar a los interesados las obras de Wallgren y otros (1996) y de Jacoby (1998).

Muy útil y completa es la página Web sobre visualización de datos del *Statistical Consulting Service* de la universidad canadiense de York. Su dirección es la siguiente: <http://www.math.yorku.ca/SCS/Gallery/>. En la misma aparecen ejemplos de buen y mal uso de los gráficos, además de muchos enlaces con páginas dedicadas a los gráficos y a la visualización de datos.

BIBLIOGRAFÍA

- Alvira, F. (1991): *Metodología de la evaluación de programas*. Madrid: CIS.
- Apodaca, P. M. (1999): Evaluación de los resultados y del impacto. *Revista de Investigación Educativa*, 17, 2, 363-377.
- Arnau, J.; Ruiz, M. y Salvador, F. (1988): Evaluación del tratamiento en diseños conductuales mediante técnicas de análisis de series temporales: una aproximación computacional. *Evaluación Psicológica*, 4, 1, 51-62.

- Batista J.M. y Coenders G. (2000): *Modelos de ecuaciones estructurales*. Madrid: La Muralla, Hespérides.
- Bertin, J. (1973): *Sémiologie graphique*. Paris: Mouton.
- Bickman, L. y Rog, D. J. (1998): Introduction: Why a Handbook of Applied Social Research Methods? en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, ix-xix.
- Botella, J. y Barriopedro, M^a. I. (1995): Análisis de datos. En Fernández-Ballesteros, R. (Ed.). *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis, 173-203.
- Brinkerhoff, R.D. y otros (1983): *Program Evaluation. A practitioner's guide for trainers and educators*. La Haya: Kluwer-Nijhoff.
- Bryk, A., Raudenbush, S.W. (1992): *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Carballo, R. (1990): Algunas aplicaciones del análisis de correspondencias a la interpretación de tablas de contingencia en la evaluación de un programa de formación profesional ocupacional. *Revista de Investigación Educativa*, 8, 16, 537-550.
- Castro, M. (1999): Modelos multinivel aplicados al meta-análisis. *Revista de Investigación Educativa*, 17, 2, 445-453.
- Castro, M. y De la Orden, A. (1998): Análisis y comparación de las características de los principales modelos de síntesis meta-analítica: aportaciones de los modelos jerárquicos lineales. *Bordón*, 50, 2, 119-134.
- Cohen, J. (1988): *Statistical power analysis in the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, H.M. (1981): On the effects of significance and the significance of effects. *Journal of Personality and Social Psychology*, 41, 1013-1018.
- Cornejo, J.M. y Martínez, J.J. (1992): Análisis de datos e intervención social en la infancia. *Anuario de Psicología*, 53, 47-59.
- De la Orden, A. (1985): Investigación evaluativa. En De la Orden, A: *Investigación Educativa. Diccionario de Ciencias de la Educación*. Madrid: Anaya, 133-137.
- De la Orden, A. (1999): Los modelos jerárquicos lineales y su aplicación a la investigación educativa. *Revista de Investigación Educativa*, 17, 2, 425-427.
- De la Orden, A. y Martínez de Toda, M. (1992): Metaevaluación educativa. *Bordón*, 43, 4, 517-527.
- De Miguel, M. (1999): La evaluación de programas: entre el conocimiento y el compromiso. *Revista de Investigación Educativa*, 17, 2, 345-348.
- Echeburua, E. y otros (1996): Tratamiento cognitivo-conductual del trastorno de estrés postraumático crónico en víctimas de maltrato doméstico: un estudio piloto. *Análisis y Modificación de Conducta*. 22, 85, 627-654.
- Escofier B. y Pagès J. (1990): *Analyses factorielles simples et multiples*. Paris: Dunod.
- Fernández-Ballesteros, R. (Ed.). (1995): *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.
- Fetterman, D.M. (1998): Ethnography, en Bickman, L. y Rog, D. J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 473-504.

- Filella, G. (2000): Evaluación del programa de orientación para la inserción laboral en los centros penitenciarios. *Revista de Investigación Educativa*, 18, 1, 83-96.
- Francis, E. y otros (1999): Improving cost-effectiveness in a substance abuse treatment program. *Psychiatric Services*, 50, 5, 633-635.
- Freixa, M. y otros (1996): Una propuesta de entrevista semi-estructurada para la evaluación de futuros padres adoptivos: aplicación y estudio mediante el análisis textual. *Anuario de Psicología*, 71, 37-49.
- Gaviria, J.L. (1999): El enfoque multinivel en la evaluación de sistemas educativos. *Revista de Investigación Educativa*, 17, 2, 429-443.
- Glass G.V., McGaw B. y Smith M.L. (1981): *Meta-analysis in social research*. Thousand Oaks, CA: Sage.
- Goldstein, H. (1995): *Multilevel Statistical Models*. London: Edward Arnold; New York: Halsted Press.
- González Ramírez, T. (2000): Metodología para la enseñanza de las matemáticas a través de la resolución de problemas: un estudio evaluativo. *Revista de Investigación Educativa*, 18, 1, 175-199.
- Hannan, P.J. y Murray, D.M. (1996): Gauss or Bernoulli? A Monte Carlo Comparison of the performance of the linear mixed-model and the logistic mixed-model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Review*, 20, 3, 338-352.
- Hair J.F., Anderson R.E., Tatham R.L. y Black W.C. (1999): *Análisis multivariante*. Madrid: Prentice Hall Iberia.
- Hedges L.V. y Olkin I. (1985): *Statistical methods for meta-analysis*. London: Academic Press.
- Henry, G.T. (1993): Using graphical displays for evaluation data. *Evaluation Review*, 17, 60-78.
- Henry, G.T. (1998): Graphing Data, en Bickman, L. y Rog, D. J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 527-556.
- Hernández, P.; Aciego, R. y Domínguez, R. (1994): Crecimiento personal y educación en valores sociales: evaluación del programa PIECAP. *Revista de Psicología General y Aplicada*, 47, 3, 339-347.
- Herraiz Gonzalo, F. (1998): Revisión de un programa cognitivo-conductual en el departamento de Psiquiatría del Centro Penitenciario de Hombres de Barcelona. *Psiquis*, 19,10, 425-432.
- Hunter J.E., Schmidt F.L. (1990): *Methods of meta-analysis*. Newbury Park: Sage.
- JACOBY, W.G. (1998): *Statistical Graphics for Visualizing Multivariate Data*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (1994): *The Program Evaluation Standards*. Thousand Oaks, CA: Sage.
- Knight, C.C. y Kuleck, W.J. (1999): The value of multimethod qualitative/quantitative research methodology in an educational program evaluation: a case study. *Mid-Western Educational Researcher*, 12, 2, 6-14.
- Kreft, I.G.G. (1998): An illustration of item homogeneity scaling and multilevel analysis techniques in the evaluation of drug prevention programs. *Evaluation Review*, 22, 1, 46-77.

- Krull, J.L. y Mackinnon, D.P. (1999): Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 4, 418-444.
- LAVIT C. (1988): *Analyse Conjointe de Tableaux Quantitatifs*. Paris: Masson.
- Lazaro Martínez, A. (1992): La formalización de indicadores de evaluación. *Bordón*, 43, 4, 477-494.
- Lebart L., Morineau A. y Piron M. (1995): *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Linn R.L. (2000): Assessments and Accountability. Disponible en: <http://www.aera.net/pubs/er/arts/29-02/linn01.htm>
- Lipsey M.W. (1990): *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Mcneal, R.B. y Hansen, W.B. (1995): An examination of strategies for gaining convergent validity in natural experiments. *Evaluation Review*, 19, 2, 141-158.
- Murillo, F.J. (1999): Los modelos jerárquicos lineales aplicados a la investigación sobre eficacia escolar. *Revista de Investigación Educativa*, 17, 2, 453-458.
- Patton, M.Q. (1997): *Utilization-Focused Evaluation. The New Century Text (3rd ed.)*. Thousand Oaks, CA: Sage.
- Pérez, M. y Salvador, L. (1998): Evaluación experimental del proyecto de inteligencia Harvard. *Papeles del Psicólogo*, 71, 33-39.
- Pérez Juste, R. (1995): Metodología para la evaluación de programas educativos. En Medina, A. y Villar, L.M. (coord.) *Evaluación de programas educativos, centros y profesores*. Madrid: Universitas, 109-144.
- Plotnick, R.D. y Deppman, L. (1999): Using benefit-cost analysis to assess child abuse prevention and intervention programs. *Child Welfare*, LXXVIII, 3, 381-407.
- REPETTO, E. y otros (1994): Últimas aportaciones en la evaluación del programa de orientación metacognitiva de la comprensión lectora. *Revista de Investigación Educativa*, 23, 314-323.
- Sáenz O. y González A. (1990): Aplicación de cadenas Markovianas a los procesos de intervención educativa. *Revista de Investigación Educativa*, 16, 2, 493-499.
- Sánchez Algarra, P. y Anguera Argilaga, M.T. (1993): Aproximación al PERT en evaluación de programas desde las técnicas matemáticas de análisis de grafos. *Anales de Psicología*, 9, 2, 213-226.
- Sieber, J.E. (1998): Planning Ethically Responsible Research, en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 127-156.
- Sogunro, O.A. (1997): Impact of training on leadership development. *Evaluation Review*, 21, 6, 713-737.
- Tejedor, F.J.; García-Valcárcel, A. y Rodríguez Conde, M^a. (1994): Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo. *Revista de Investigación Educativa*, 23, 93-127.
- Tukey, J.W. (1977): *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Uriel, E. (1992): *Análisis de series temporales. Modelos ARIMA*. Madrid: Paraninfo.
- Wallgren, A. y otros (1996): *Graphing Statistics and Data*. Thousand Oaks, CA: Sage.
- Wang, J. y otros (1999): Evaluation of HIV risk reduction intervention programs via latent growth model. *Evaluation Review*, 23, 6, 648-662.

Ximénez C. y San Martín R. (2000): *Análisis de varianza con medidas repetidas*. Madrid: La Muralla, Hespérides.

Yin, R.K. (1998): The Abridged Version of Case Study Research: Design and Method, en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 229-259.

DIRECCIONES INTERNET:

<http://www.ioe.ac.uk/multilevel>

<http://www.math.yorku.ca/SCS/Gallery/>

<http://www.aera.net/pubs/er/arts/29-02/linn01.htm>