META-ANÁLISIS Y MODELOS JERÁRQUICOS LINEALES: ESTUDIO ANALÍTICO DE LAS APORTACIONES DE LOS ÍNDICES DE MAGNITUD DEL EFECTO INDIVIDUAL CLÁSICO Y EMPÍRICO BAYESIANO

MARÍA CASTRO MORERA (*)
JOSÉ LUIS GAVIRIA SOTO (*)

RESUMEN. Se analizan los procedimientos sistemáticos para la síntesis de resultados procedentes de distintas investigaciones. En concreto, se centra en las aportaciones del procedimiento de síntesis más consolidado, el meta-análisis. En este trabajo se analizan comparativamente las aportaciones del índice meta-analítico por excelencia: la magnitud del efecto individual. En la literatura científica aparecen perfilados dos tipos diferenciados de estimadores de la magnitud del efecto individual: clásico y empírico-bayesiano. Sin embargo, no se encuentra un estudio analítico de las aportaciones de ambos estimadores. Los resultados muestran el mejor comportamiento del estimador empírico-bayesiano. Este estimador, aunque tiende a infravalorar levemente el valor paramétrico central, es clara y sistemáticamente más eficiente que el clásico. Se comprueba que los estimadores empíricos-bayesianos, propuestos dentro de la perspectiva de los modelos jerárquicos lineales, suponen una clara aportación frente a los estimadores clásicos. Su comportamiento estadístico es aceptable y mejor que el de los estimadores tradicionalmente empleados.

INTEGRACIÓN DE RESULTADOS DE INVESTIGACIÓN COMO MEDIDA DE RACIONALIDAD ANTE EL AVANCE CIENTÍFICO

La historia de la investigación educativa está compuesta por cientos de miles de experimentos con el objetivo común de mejorar la educación en algunos de sus ámbitos. En cada una de las investigaciones se ha invertido una no desdeñable cantidad de recursos económicos y, más importante aún, de recursos humanos. Sin embargo, teoría y práctica educativa se han visto muy poco alteradas por la aplicación de los resultados encontrados. En muchas áreas educativas, actualmente, el problema no radica en la carencia de evidencias experimentales que validen o falseen una determinada hipótesis, sino precisamente en lo contrario, en el exceso de resultados, normalmente incongruentes entre sí. Esta ausencia de resultados coherentes dificulta la toma de decisiones para el desarrollo de un cuerpo teórico sistemático y para la puesta en marcha de progra-

(*) Universidad Complutense de Madrid.

mas de intervención realmente eficaces, puesto que, desde la investigación no se puede ofrecer una respuesta unívoca.

En cualquier ámbito de estudio, es fácil que el investigador caiga en la tentación de plantearse realizar nuevos estudios que aclaren «de una vez por todas» la situación. Sin embargo, en Ciencias Sociales y, más concretamente, en Ciencias de la Educación el mito del único y definitivo estudio resulta una completa farsa (Rosnow y Rosenthal, 1989). Los resultados de una nueva investigación pasan a la cada vez más abultada lista de conclusiones incongruentes entre sí, sin aportar mucho más y necesitando siempre una nueva investigación, la *n+1*, que resuelva el problema educativo planteado. De tal modo que, en lugar de formar un ovillo ordenado y estructurado, los nuevos esfuerzos pueden liar aún más la madeja. Dentro de esta dinámica, es fácil caer en la «falacia epistemológica», por la cual la realidad o la naturaleza de los fenómenos que investigamos es igual a lo que se observa. Desde el realismo científico, se apunta que el papel de la ciencia y del conocimiento es el desarrollo de teorías que permitan explicar los fenómenos estudiados y probar esas teorías con criterios racionales (House, 1991).

La solución, por tanto, no pasa por rechazar anteriores esfuerzos a través de un «borrón y cuenta nueva» sino por utilizarlos. Así, es posible afirmar que el avance de la Ciencia, y fundamentalmente en las Ciencias Sociales, debe proceder de forma acumulativa, sin que esto signifique, en ningún caso, la renuncia a la realización de estudios individuales. Más aún, si se consideran algunas de las características propias de la investigación educativa como son su naturaleza multidisciplinar y multivariada, se hace más evidente la necesidad de acumular los resultados y conclusiones procedentes de distintas investigaciones individuales, puesto que, por el objeto de estudio, resulta prácticamente imposible diseñar un estudio que refleje y controle la complejidad de los fenómenos educativos.

De esta manera, el problema no es la carencia de informaciones, de datos, sino la falta de conversión de esa información en conocimientos útiles que permitan desarrollar cuerpos coherentes de conocimientos relativos a los procesos educativos contrastados con la realidad y, de manera subsidiaria, mejorar la práctica educativa. En palabras de Glass (1976, p. 4), nuestro problema es encontrar el conocimiento en la información.

La revisión de resultados de investigación se perfila como una actividad científica relevante e imprescindible puesto que se constituye una medida de racionalidad ante el avance científico en Ciencias Sociales.

Ahora bien, la realización de síntesis de resultados provenientes de distintas investigaciones supone prestar una especial atención a los supuestos implícitos y explícitos de cada una de ellas. Cada estudio revisado parte de planteamientos conceptuales diferentes, emplea únicamente algunas variables comunes al conjunto de estudios que componen la muestra y aplica distintas técnicas de recogida de información, de tratamiento de la misma, siendo éstos sólo algunos de los aspectos que se podrían enumerar. Parece entonces que, es necesario ser especialmente cuidadoso en el proceso de resumir la información contenida en las distintas investigaciones, con el objeto de no desvirtuar sus planteamientos, sus constructos medidos o sus resultados. La atención al proceso técnico de desarrollo y realización de una revisión de resultados de investigación constituye una de las primeras necesidades de las revisiones de resultados de investigación. No se puede volver la espalda a los riesgos técnicos que supone la integración de resultados procedentes de distintas investigaciones sobre el mismo tema, dado que de ellos depende la calidad y exactitud de las conclusiones que se extraigan. *Aprehender el significado de un amplio conjunto de investigaciones sobre un determinado problema educativo se convierte básicamente en una cuestión de* *carácter técnico o metodológico* (Glass, 1976; Pillemer y Light, 1980).

En este artículo, el centro de interés se desplaza al análisis de procedimientos sistemáticos para la síntesis de resultados. En concreto, nos centraremos en el estudio de las aportaciones del procedimiento de síntesis más consolidado, el meta-análisis. El objetivo del trabajo que se presenta es analizar comparativamente las aportaciones del índice meta-analítico por excelencia: la magnitud del efecto individual. En la literatura científica aparecen perfilados dos tipos diferenciados de estimadores de la magnitud del efecto individual. Sin embargo, no se encuentra un estudio analítico de las aportaciones de ambos estimadores. Para cumplir con este objetivo, este artículo se organiza en cinco grandes epígrafes. El primero está dedicado a hacer una breve descripción de la herramienta meta-analítica. En el segundo epígrafe nos centraremos en la descripción de los principales modelos de síntesis meta-analítica que determinan los estimadores. A continuación, se presentan con detalle los dos índices de magnitud del efecto individual. El cuarto apartado, muestra el análisis comparativo de las características técnicas de ambos estimadores. Y, por último, se presentan las conclusiones de este trabajo.

EL META-ANÁLISIS COMO PERSPECTIVA METODOLÓGICA PARA LA SÍNTESIS DE RESULTADOS Y LA MAGNITUD DEL EFECTO INDIVIDUAL

De una manera genérica se puede decir que los principales problemas metodológicos con los que se encuentran los investigadores a la hora de perfilar un método coherente para la integración de resultados de investigación son dos, la existencia de distintas unidades de medida de los estudios a integrar y la presencia de variaciones en las características básicas de los distintos estudios. Un grupo de estudios puede analizar la misma cuestión, incluso emplear métodos de análisis similares, pero el revisor puede no estar seguro de que las conclusiones alcanzadas tras la integración de los resultados de las investigaciones sobre ese tema no estén distorsionadas debido a las diferencias en los criterios de medida o las características propias de cada uno de los estudios.

Hasta 1976 no aparece en la literatura sobre revisiones, una alternativa metodológica que supere los dos problemas antes mencionados: el *meta-análisis*. Este término procede de una clasificación, realizada por Glass (1976), de los distintos análisis de datos en función del tipo de datos que emplea, distinguiendo tres niveles: primario, secundario y meta-análisis. El *análisis primario* es el tratamiento estadístico original de los datos de una investigación. El *análisis secundario* se define como el re-análisis de los datos procedentes de una investigación original con el objeto de responder a las cuestiones inicialmente planteadas en ella con mejores herramientas estadísticas o responder a nuevos objetivos de investigación con «viejos» datos. La principal novedad es la introducción del término *meta-análisis*, acuñado por el autor, refiriéndose a un *análisis de los análisis*. El meta-análisis consiste en el tratamiento estadístico de los resultados de un conjunto amplio de estudios primarios con el objeto de integrar estos resultados tomando como referencia un conjunto de criterios objetivos y comunes para el conjunto de investigaciones a integrar. De esta manera, se establece un procedimiento de síntesis netamente distinto a los empleados hasta ese momento.

El meta-análisis es un procedimiento sistemático de integración de resultados de investigación y pretende, globalmente, el análisis y síntesis de los resultados obtenidos en un conjunto de estudios experimentales que abordan la misma problemática de investigación, para comprobar si la convergencia de hipótesis y la variada gama de diseños de investigación conducen a los mismos o similares resultados para el problema objeto de estudio. En definitiva, trata de buscar patrones de solución comunes en un cuerpo de estudios que abordan la misma problemática desde distintas ópticas y con distintos procedimientos.

Desde el punto de vista metodológico, el meta-análisis supone un hito en la evolución procedimental de las revisiones de resultados de investigación. La introducción de un estadístico, la *magnitud de los efectos*, que permite poner en una misma escala los resultados procedentes de distintas investigaciones primarias es una de sus principales aportaciones.

Los distintos modelos meta-analíticos nacen como un intento de reflexión, profundización y mejora metodológica de las técnicas empleadas para la integración de resultados de investigación. Esta reflexión ha dado lugar a modelos matemáticos complejos para la síntesis cuantitativa de la evidencia. Estos distintos esfuerzos de modelización de la síntesis han abierto un extenso debate científico sobre la adecuación y pertinencia de los modelos y estimadores empleados.

PRINCIPALES MODELOS DE SÍNTESIS META-ANALÍTICA

El proceso general de síntesis de resultados procedentes de distintas investigaciones supone la integración de los índices de la magnitud de los efectos de cada investigación primaria. Los distintos modelos de síntesis constituyen las distintas alternativas metodológicas para el tratamiento de problemas meta-analíticos.

En la literatura meta-analítica aparecen descritos dos grandes tipos de modelos empleados para sintetizar evidencias:

- Los modelos clásicos, planteados por Hedges y Olkin (1985).
- Los modelos jerárquicos lineales (Raundenbush y Bryk, 1985; Bryk y Raundenbush, 1992) como alternativa teórica a las limitaciones manifestadas por los modelos clásicos.

Las alternativas propuestas suponen un intento de superación del modelo de la media global que ha venido dominando en la literatura meta-analítica, ya que son necesarias aproximaciones más sensibles que reflejen que los efectos posibles de un tratamiento están influenciados por distintos factores. Así, se puede asumir que la variabilidad entre magnitudes del efecto individuales es totalmente aleatoria y, por tanto, inexplicable. Por este motivo, el modelo aleatorio de Hedges (1983) y Rubin (1981) es parcialmente adecuado. La variabilidad entre resultados dentro de este modelo es igual a la variabilidad debida al muestreo entre sujetos y a la variabilidad debida al muestreo entre estudios. En este caso, tanto la variable independiente como la dependiente tienen una variación aleatoria. Sin embargo, también se puede pensar que la determinación de la variable independiente no es aleatoria, sino fija, agrupando a los estudios en distintos bloques que presentan cierta homogeneidad interna por compartir características similares. Desde este punto de vista, la variabilidad entre resultados se debe únicamente al muestreo entre sujetos, ya que no reconoce la existencia de variabilidad entre estudios. Este modelo de efectos fijos también puede ser factible y la variabilidad observada entre resultados atribuirse a las distintas características de los estudios (Hedges, 1982; Hedges y Olkin, 1985). Sin embargo, ambos modelos parecen limitados, puesto que la variación entre resultados puede atribuirse a la variación entre sujetos, a la variación debida al muestreo entre estudios y a la variación atribuible a las características de los estudios. Parece que una solución puede venir de la mano de un modelo de efectos mixtos, ya que incorpora todas las posibilidades y aporta una estructura cuando la variación entre efectos es parcialmente explicable. Este modelo se integra dentro de la lógica de los modelos jerárquicos lineales. El *modelo lineal mixto* permite (Raundenbush y Bryk, 1985):

- La estimación de la varianza del parámetro de la magnitud de los efectos a través de medias de máxima verosimilitud.
- Proponer un conjunto de modelos lineales para explicar los efectos de la varianza paramétrica.
- Estimar la varianza residual de los parámetros de la magnitud del efecto para cada modelo lineal.
- Emplear la información sobre las características de las investigaciones para obtener estimaciones *empírico bayesianas* mejoradas de las magnitudes del efecto individuales.
- Examinar la sensibilidad de las inferencias sustantivas con relación a los errores en la estimación de los componentes de varianza.

Así, la aportación está centrada en dos puntos: en primer lugar, el reconocimiento de la doble fuente de variabilidad aleatoria, que además tiene la ventaja de ser un modelo más generalizable que los modelos meta-analíticos clásicos, a pesar de tener una varianza de error menor que el modelo de efecto aleatorios; y, en segundo lugar, la consideración de la existencia de una distribución de parámetros, δ_i, rompiendo la lógica clásica que prevé un único valor paramétrico central para cualquier estimación de la magnitud del efecto individual.

En la propuesta de *modelo de efectos mixtos* de Raundenbush y Bryk (1985), la variación entre magnitudes del efecto individuales se entiende como aleatoria, para proponer un modelo lineal que permita la explicación de la variación entre magnitudes del efecto como una función de las características de las investigaciones más error. En el primer nivel, la variación entre resultados se define como aleatoria. Mientras que en el segundo nivel, la variación inter-estudios se trata como un resultado, dependiente de las características de los estudios, se entiende, por tanto, como un efecto fijo más un término de error de muestreo inter-estudios. Los modelos jerárquicos proponen un conjunto de modelos lineales para explicar la variabilidad de los parámetros de magnitud del efecto, respetando la estructura jerárquica de los datos meta-analíticos. Cada investigación individual refleja una determinada relación entre tratamientos y efectos, relación que está anidada en una investigación concreta, definida por un conjunto de características. El problema de investigación se repite constantemente en un conjunto de investigaciones, cada una con sus propias particularidades. Los modelos jerárquicos lineales reconocen este segundo nivel, lo que permite evaluar la consistencia de la relación efectos-tratamientos en todo el conjunto de investigaciones. Proceden descomponiendo la variabilidad de los resultados en fuentes aleatorias y sistemáticas a través de un modelo lineal mixto.

En un modelo de efectos mixtos se identifican tres fuentes de variabilidad. Por una parte, la de origen sistemático, producida por la relación entre w (una característica de los estudios) y δ. Por otra parte, dado un estudio primario con un valor dado de w, w_i, ese estudio puede considerarse como un elemento extraído al azar de una distribución de estudios que comparten un valor común en w_i, pero difiriendo en otras variables (como p.e., aptitudes de los sujetos experimentales). Hay por tanto una variabilidad debida a esas diferencias: muestreo entre estudios.

Todos esos estudios que comparten una misma característica (w_i) están representados por una variable fija, δ_i, que es una función de la característica, así δ_i = f (w_i). Ahora, si se seleccionara un estudio concreto de esa distribución es: $\delta_i = \delta + u_i$ donde: $u_i \sim N(0, \sigma_u^2)$ y siendo u_i una variable aleatoria que corresponde al error debido al desconocimiento de las otras características del estudio que afectan a su efecto paramétrico.

Por último, hay otro error que se debe al muestreo entre sujetos. Por tanto, cuando se calcula un valor d_i de un estudio primario, ese valor incluye un error debido a este muestreo entre sujetos, por lo que $d_{ij} = \delta_{ij} + e_{ij}$, donde $\delta_i \sim N(\delta, \tau_{11})$ y $e_{ij} \sim N(0, \sigma_e^2)$.

En un solo modelo se tiene que $d_{ij} = \delta_i + u_i + e_i$ y como $\delta_i = f(w_i)$, entonces: $d_{ij} = f(w_i) + u_{ij} + e_{ij}$. Su variabilidad es: $var(d_{ij}) = var(w_i) + var(u_{ij}) + var(e_{ij})$. Ahora bien, si $f(w_i)$ es una función lineal, $\delta_i = \gamma_0 + \gamma_1 w_i$ (es decir, medido sin error, por tanto la característica w explica completamente δ_i), entonces: $d_{ij} = \gamma_0 + \gamma_1 w_i + u_{ij} + e_{ij}$.

El meta-análisis con modelos jerárquicos lineales se concreta matemáticamente en un modelo de efectos mixtos. En un modelo de efectos fijos, la única variabilidad de d_i es la debida al muestreo de sujetos en cada estudio primario, puesto que se supone que los estudios han sido extraídos de una población de estudios idénticos, salvo en la muestra empleada. De ahí que su poder de generalización alcance sólo a aquellos estudios iguales a los incluidos en la muestra. En un modelo de efectos aleatorios, tenemos la variabilidad de los sujetos y la debida al muestreo de los estudios, de ahí que su poder de generalización sea mayor, sin embargo toda la variabilidad es aleatoria, no refleja la posibilidad de que características de las investigaciones primarias funcionen como fuentes de variación sistemática. En un modelo de efectos mixtos, además de la variabilidad de los sujetos, identifica dos componentes. Uno, explicado por la variable independiente de segundo nivel y es, por tanto, varianza de origen conocido o sistemática (a pesar de ser debida al muestreo de estudios) y otro componente que es aleatorio (u_i), debido al muestreo entre investigaciones primarias. Tiene la ventaja de que es más generalizable que los enfoques clásicos y de tener una varianza de error menor que los aleatorios. Se puede decir que el modelo de efectos mixtos es un modelo exhaustivo, ya que permite reflejar distintas posibilidades de explicación de la variabilidad. En este sentido se puede decir que es un modelo flexible y con capacidad de adaptación a distintas situaciones de investigación. Ahora bien, es un modelo que incluye una gran cantidad de parámetros, es decir, que es menos parsimonioso que los otros dos, que tratan de explicar los fenómenos con el menor número de parámetros.

Independientemente de las diferencias en los modelos, se observan también diferencias en los distintos tipos de estimadores. La magnitud del efecto individual es, como ya se ha dicho, el estimador meta-analítico por excelencia. Es la base de todo el proceso de estimación y análisis, de ahí que sea especialmente importante tener buenos estimadores, estadísticamente hablando. Las distintas concepciones de la tarea de síntesis de evidencias (clásicas y basadas en los modelos jerárquicos lineales) han dado lugar a dos tipos de estimadores distintos que pasamos a describir.

ÍNDICES DE MAGNITUD DEL EFECTO INDIVIDUAL

La estimación de la magnitud de los efectos clásica se realiza por medio de la comparación de resultados procedentes de las condiciones experimental y de control. La medida que se obtiene es la diferencia de medias tipificada procedentes de las dos condiciones experimentales. Esta medida es conocida como *magnitud o tamaño del efecto individual* (δ_i) y viene dada por el siguiente estimador (Hedges, 1981):

$$d_i = \frac{\overline{X}_e - \overline{X}_c}{s}, \qquad \text{(Ecuación 1)}$$

donde s es la desviación típica intragrupos dada por:

$$s = \sqrt{\frac{(n_e - 1)s_e^2 + (n_c - 1)s_c^2}{n_e + n_c - 2}},$$

(Ecuación 2)

donde n_e, n_c, s_e y s_c son los tamaños muestrales y las desviaciones típicas respectivas de los grupos experimental y de control.

La magnitud del efecto representa la estandarización de la diferencia de medias encontradas entre tratamientos o distintas condiciones experimentales en cada investigación primaria. Así, la magnitud del efecto proporciona una medida estándar, en virtud de la cual se hacen directamente comparables los resultados de diferentes estudios. Además, el significado de la magnitud del efecto es fácilmente interpretable y comprensible. Dada la normalidad de la distribución de medias, las distintas magnitudes del efecto pueden representarse gráficamente en forma de distribuciones de puntuaciones y escalas percentiles comparables. Como se ha podido observar, la magnitud del efecto, tal y como ha sido definida en este caso, considera únicamente diseños experimentales con postest. Los estudios independientes también puede analizarse mediante correlaciones.

Los modelos clásicos emplean este estimador clásico insesgado de la magnitud del efecto. En cambio, los modelos jerárquicos lineales, que también utilizan una diferencia de medias estandarizada en todos sus análisis, proponen realizar una estimación mejorada de este índice. Esta estimación emplea toda la información procedente del conjunto de investigaciones. Ambos modelos toman el valor del estadístico de Hedges como estadístico básico para obtener estimaciones más precisas. Sin embargo, la principal diferencia entre ambos es que dentro de los modelos clásicos se utiliza únicamente la información procedente de cada investigación, como se ve en la *ecuación 1* sólo se utilizan los datos procedentes de cada investigación aisladamente. Mientras que, el estimador en el modelo jerárquico lineal es la distribución *a priori* de δ_i, ya que se intenta estimar el valor del parámetro en función del conjunto de datos procedentes de todas las investigaciones incluidas en el meta-análisis. Este estimador viene dado por la siguiente expresión:

$$\delta^* = \lambda_i d_i + (1 - \lambda_i)(\gamma_0^* + \gamma_1^* w_i)$$

(Ecuación 3)

En esta expresión, $\lambda_i = \tau / (\tau + v_i)$ es la precisión del estimador, d_i es el estimador clásico, γ_0 representa el valor de la gran media dentro de los modelos jerárquicos lineales y γ_1 representa la relación existente con una posible variable moderadora de los resultados, si el modelo es no condicionado, el valor de $\gamma_1 = 0$.

La *ecuación 3* es el estimador *empírico bayesiano* (James y Stein, 1961; Efron y Morris, 1975) que también es la media de la distribución posterior de δ, cuando la distribución *a priori* sobre γ es imprecisa. La estimación δ_i^* es una ponderación de d_i y $(\gamma_0 + \gamma_1 w_i)$. Para comprender el significado de esta composición de d_i y $(\gamma_0 + \gamma_1 w_i)$ se deben considerar dos situaciones. Primera, supóngase que se tiene únicamente los datos de un estudio, i, la estimación, δ^*, basada en estos datos sería d_i. Por otro lado, supóngase que se desconocen los datos del estudio i, pero conocemos algo sobre la muestra de investigaciones, la relación entre las características, $(\gamma_0 + \gamma_1 w_i)$, y los resultados de los estudios. En este caso, el mejor estimador de δ^* sería $(\gamma_0 + \gamma_1 w_i)$. Segunda, supongamos ahora que ambas estimaciones están disponibles. Cómo combinarlas para obtener un mejor estimador, superior a los anteriores. La ecuación 3 muestra una óptima combinación de las dos estimaciones. Es óptima porque la media bayesiana posterior minimiza la pérdida cuadrática de error. Las ponderaciones para d_i son proporcionales a v^{-1}, que es la precisión en la estimación de δ^* basada sólo en los datos del estudio i. Las ponderaciones para $(\gamma_0 + \gamma_1 w_i)$ son proporcionales a τ^2, que son las medidas de concentración de δ alrededor de $(\gamma_0 + \gamma_1 w_i)$. Por tanto, se obtiene el resultado más preciso intra-estudios, las ponderaciones se realizan en torno a d_i.

Este índice supera el estimador clásico, puesto que a la información que aporta cada investigación primaria, incorpora la información procedente del conjunto de investigaciones a integrar.

Así, las diferencias en los estimadores de la magnitud del efecto clásico y empírico bayesiano son especialmente destacables. Desde un punto de vista teórico, la estimación empírico bayesiana permite incorporar las informaciones de todo el conjunto de investigaciones primarias a la estimación de cada una de las magnitudes del efecto de cada investigación, frente al estimador clásico que únicamente utiliza la información procedente del estudio individual de referencia.

Sin embargo, la aportación descrita, que desde el punto de vista teórico incorpora el estimador empírico bayesiano, no ha sido comprobada. Al revisar la literatura de investigación se constata la ausencia de comparaciones analíticas y empíricas entre ambos modelos. Se hace necesario un análisis comparado y pormenorizado para poder ofrecer a los distintos usuarios del meta-análisis criterios sobre los que justificar la adopción de una determinada orientación. El estudio del funcionamiento comparado de estos dos estimadores es el objeto de estudio de este trabajo. En este artículo se presentan los resultados del estudio del comportamiento de los índices de la magnitud del efecto individual dentro de las alternativas clásica y empírico bayesiana para conocer de forma analítica y comparada sus aportaciones.

APORTACIONES DEL ESTIMADOR EMPÍRICO-BAYESIANO FRENTE AL CLÁSICO

El estudio analítico de las propiedades estadísticas de los estimadores de la magnitud del efecto individual clásica y empírico-bayesiana es uno de los objetos de atención de este artículo. De forma más específica, se comparan estos estimadores en dos cuestiones fundamentales: sesgo y eficiencia relativa de ambos estimadores.

El *sesgo* de un estimador se define genéricamente como la esperanza de la diferencia entre el estimador y el parámetro. Para el estudio analítico del sesgo de los estimadores clásico y empírico-bayesiano de la magnitud del efecto individual se han desarrollado las distintas formalizaciones matemáticas de la esperanza de la diferencia entre estadísticos y parámetros para ambos tipos de estimadores.

Al estudiar la esperanza matemática del sesgo del estimador clásico, se observa cómo el sesgo de este estimador es cero:

$$\varepsilon(d_i - \delta_i | \delta_i) = \varepsilon\,(d_i | \delta_i) - \varepsilon\,(\delta_i | \delta_i) = \delta_i - \delta_i = 0$$

Como el sesgo es cero, entonces: $(di - \delta_i) \rightarrow 0$ cuando el tamaño de las muestras tiende a infinito (es decir, cuando es suficientemente grande).

El análisis de la esperanza matemática del estimador empírico-bayesiano muestra cómo el sesgo de este estimador nunca podrá ser cero incluso cuando N sea suficientemente grande, así:

$$\overline{(\delta_i^{*} - \delta_i)} \equiv \varepsilon(\delta_i^{*} - \delta_i | \delta_i) = \varepsilon(\delta_i^{*} | \delta_i) - \varepsilon(\delta_i | \delta_i) = \varepsilon(\delta_i^{*} | \delta_i) - \delta_i$$

Por tanto, la esperanza matemática del estimador empírico-bayesiano sólo podrá ser cero cuando $\varepsilon(\delta_i^{*} | \delta_i) = \delta_i$. Al desarrollar el primer término de esta expresión, se obtiene:

$$\begin{aligned}\varepsilon(\delta_i^{*} | \delta_i) &= \lambda_i\, \varepsilon(d_i | \delta_i) + (1 - \lambda_i)\, \varepsilon(\gamma_0^{*} + \gamma_1^{*}\, w_i) = \\ &\lambda_i \delta_i + (1 - \lambda_i)\, \varepsilon(\gamma_0^{*} + \gamma_1^{*}\, w_i) = \\ &\lambda_i \delta_i + (1 - \lambda_i)\, \varepsilon(\gamma_0^{*}) + (1 - \lambda_i) \\ &\varepsilon\, (\gamma_1^{*}\, w_i)\end{aligned}$$

Igualando a cero:

$$\lambda_i \delta_i + (1 - \lambda_i)\, \varepsilon(\gamma_0^{*}) + (1 - \lambda_i)\, \varepsilon\, (\gamma_1^{*}\, w_i) = 0$$
$$\lambda_i \delta_i = - (1 - \lambda_i)\, \varepsilon(\gamma_0^{*}) + (1 - \lambda_i)\, \varepsilon\, (\gamma_1^{*}\, w_i)$$

Se comprueba que sólo podrá valer cero cuando λ_i sea uno, puesto que $(1 - \lambda_i) = 0$ y exige, además, que δ_i sea igual a cero. λ_i es la precisión del estimador y viene dada por la expresión: $\lambda_i = \dfrac{\tau^2}{\tau^2 + \nu_i}$. Esta expresión se hace uno cuando ν_i es igual a cero, situación en la práctica imposible, dado que supone asumir que no existe variabilidad entre los sujetos de cada muestra, ya que es el término de error debido a las diferencias entre sujetos.

Desde un punto de vista empírico, la eficiencia de las estimaciones empírico-bayesianos es mucho mayor que la de las estimaciones clásicas, tal y como ya señalan Raundebush y Bryk (1985) y Bryk y Raundenbush (1992). Esta menor dispersión de los valores del estimador empírico-bayesiano es un claro indicador de una concentración en torno a la gran media, debida al factor de concentración $(1 - \lambda_i)$ (James y Stein, 1961; Efron y Morris, 1975). Ya se ha señalado que λ_i puede interpretarse como la fiabilidad de d_i como estimación de δ_i. Así, si la estimación de d_i fuera perfectamente fiable ($\lambda_i = 1$) no habría ningún tipo de concentración ni diferencias en la dispersión de ambos estimadores. Por tanto, se puede decir que las estimaciones empírico-bayesianas producen mejores inferencias sobre la magnitud del efecto de cada estudio individual como también muestran los trabajos de Louis (1991), Louis y Zelterman (1994) y Castro (1997).

De este modo, es necesario hacer notar que este comportamiento sesgado del estimador δ_i^{*} no es tan perjudicial. Ya hemos dicho que δ_i^{*} es la alternativa a d_i cuando este último es un mal estimador, es decir cuando $\lambda_1 \neq 1$, es decir, casi siempre. Se sabe que d_i es insesgado. Sin embargo, cuando se calcula una estimación se obtiene un único valor, del que no se conoce exactamente en qué punto se sitúa respecto al parámetro, dentro del posible rango de variación del error. Lo que sí se conoce es que el rango de variación del error del estimador clásico es mucho más amplio que el rango de variación del estimador empírico-bayesiano. Por otra parte, hay evidencia de que δ_i^{*} es un estimador sesgado, tendiendo a infravalorar el valor del parámetro, como máximo dos centésimas (Castro, 1997). Valor que debe ser interpretado a la luz del valor de τ^2. Sin embargo, al estimar un único valor del estadístico se sabe que siempre tenderá a ser un valor más cercano al parámetro frente al posible error del clásico, puesto que su rango de variación es claramente menor. De este modo se puede concluir que *el estimador* δ_i^* *está ligeramente más sesgado que el estimador clásico, pero tiene una varianza de error menor.*

La *eficiencia relativa* es la varianza del error del efecto medio estimado. Un aumento de los valores de la varianza se interpreta como una disminución de la eficiencia. En este estudio se muestra que los estimadores empírico-bayesianos son más eficientes que los clásicos, en cualquiera de las dos situaciones posibles, por las que aparece o no la influencia de una variable moderadora de los resultados.

DEMOSTRACIÓN ANALÍTICA DE QUE LOS ESTIMADORES EMPÍRICO-BAYESIANOS SON MÁS EFICIENTES QUE LOS ESTIMADORES CLÁSICOS

CASO 1: *Modelo que no incluye una variable moderadora o no condicionado.*

$(\delta_i^* = \lambda_i\, d_i + (1 - \lambda_i)\, \gamma_0)$

La varianza del estimador empírico bayesiano es una función de la varianza del estimador no bayesiano:

$$v(\delta_i^*) = f(v(d_i))$$
$$v(\delta_i^*) = \varepsilon(\delta_i^{*2}) - \varepsilon(\delta_i^*)^2$$
$$v(\delta_i^*) = \varepsilon(d_i^2) - \varepsilon(d_i)^2$$

Por otro lado, también se sabe que: $\delta_i^* = \lambda_i\, d_i + (1 - \lambda_i)\, \gamma_0$. Así, al desarrollar estas expresiones se obtiene que:

$$\varepsilon(\delta_i^*) = \lambda_i\, \varepsilon(d_i) + (1 - \lambda_i)\, \gamma_0$$
$$\varepsilon(\delta_i^*)^2 = \lambda^2_i\, \varepsilon(d_i)^2 + (1 - \lambda_i)^2\, \gamma_0^2 + 2\,\lambda\, (1 - \lambda_i)\, \gamma_0\, \varepsilon(d_i)$$

$$\varepsilon(\delta_i^{*2}) = \varepsilon[\lambda_i^2\, d_i^2 + (1 - \lambda_i)^2\, \gamma_0^2 + 2\lambda(1 - \lambda_i)\, \gamma_0 d_i] = \lambda_i^2\, \varepsilon(d_i^2) + (1 - \lambda_i)^2\, \gamma_0^2 + 2\lambda\, (1 - \lambda_i)\, \gamma_0\, \varepsilon(d_i)$$

Sustituyendo en la expresión inicial:

$$v(\delta_i^*) = \varepsilon(\delta_i^{*2}) - \varepsilon(\delta_i^*)^2 = [\lambda_i^2\, \varepsilon(d_i^2) + (1 - \lambda_i)^2\, \gamma_0^2 + 2\lambda_i\, (1 - \lambda_i)\, \gamma_0\, \varepsilon(d_i)] - [\lambda_i^2\, \varepsilon(d_i)^2 + (1 - \lambda_i)^2\, \gamma_0^2 + 2\lambda\, (1 - \lambda_i)\, \gamma_0\, \varepsilon(d_i)] = \lambda_i^2\, \varepsilon(d_i^2) - \lambda_i^2\, \varepsilon(d_i)^2 = \lambda_i^2\, [\varepsilon(d_i^2) - \varepsilon(d_i)^2] = \lambda_i^2\, v(d_i)$$

Así se comprueba que el estimador bayesiano es más eficiente que el clásico ya que λ es el valor de la fiabilidad y está comprendido entre 0 y 1. Por tanto, la varianza de d_i es sistemáticamente mayor que la de δ_i^*.

CASO 2: *Modelo que incluye una variable moderadora o condicionado.*

$(\delta_i^* = \lambda_i\, d_i + (1 - \lambda_i)\, (\gamma_0 + \gamma_1\, w_i))$

Como en el caso anterior, se quiere probar que la varianza del estimador empírico bayesiano es una función de la varianza del estimador clásico, ahora incorporando la presencia de una variable moderadora, como en el caso anterior:

$$v(\delta_i^*) = f(v(d_i))$$
$$v(\delta_i^*) = \varepsilon(\delta_i^{*2}) - \varepsilon(\delta_i^*)^2$$
$$v(\delta_i^*) = \varepsilon(d_i^2) - \varepsilon(d_i)^2$$

Por otro lado, también se sabe que: $\delta_i^* = \lambda_i d_i + (1 - \lambda_i)\, (\gamma_0 + \gamma_1 w_i)$ estimador que incluye la presencia de una variable moderadora. Así, al desarrollar estas expresiones se obtiene que:

$$\delta_i^* = \lambda_i d_i + (1 - \lambda_i)\, (\gamma_0 + \gamma_1 w_i)$$
$$\delta_i^{*2} = \lambda_i^2 d_i^2 + (1 - \lambda_i)^2\, (\gamma_0 + \gamma_1 w_i)^2 + 2\lambda_i d_i\, (1 - \lambda_i)\, (\gamma_0 + \gamma_1 w_i)\; \varepsilon(\delta_i^{*2}) = \lambda_i^2\, \varepsilon(d_i^2) + (1 - \lambda_i)^2\, \varepsilon[(\gamma_0 + \gamma_1 w_i)^2] + 2\lambda_i\, \varepsilon(d_i)\, (1 - \lambda_i)\, \varepsilon(\gamma_0 + \gamma_1 w_i)$$

Y,

$\varepsilon(\delta_i^*) = \lambda_i^2\ \varepsilon(d_i) + (1 - \lambda_i)\ \varepsilon(\gamma_0 + \gamma_1 w_i)$
$[\varepsilon(\delta_i^*)]^2 = \lambda_i^2\ [\varepsilon(d_i)]^2 + (1 - \lambda_i)^2\ [\varepsilon(\gamma_0 + \gamma_1 w_i)]^2 + 2\lambda_i\ \varepsilon(d_i)\ (1 - \lambda_i)\ \varepsilon(\gamma_0 + \gamma_1 w_i)$

Sustituyendo en la expresión inicial:

$\nu(\delta_i^*) = \varepsilon(\delta_i^{*2}) - [\varepsilon(\delta_i^*)]^2 = \lambda_i^2\ \varepsilon(d_i^2) + (1 - \lambda_i)^2\ \varepsilon[(\gamma_0 + \gamma_1 w_i)^2] + 2\lambda_i\ \varepsilon(d_i)\ (1 - \lambda_i)\ \varepsilon(\gamma_0 + \gamma_1 w_i) - \lambda_i^2\ \varepsilon(d_i^2) - (1 - \lambda_i)^2\ [\varepsilon(\gamma_0 + \gamma_1 w_i)]^2 - 2\lambda_i\ \varepsilon(d_i)\ (1 - \lambda_i)\ \varepsilon(\gamma_0 + \gamma_1 w_i) = \lambda_i^2\ \varepsilon(d_i^2) + (1 - \lambda_i)^2\ \varepsilon[(\gamma_0 + \gamma_1 w_i)^2] - \lambda_i^2\ \varepsilon(d_i^2) - (1 - \lambda_1)^2\ [\varepsilon(\gamma_0 + \gamma_1 w_i)]^2 =$

Operando y sacando factor común, se obtiene que:

$= \lambda_i^2\ (\varepsilon(d_i^2) - [\varepsilon(d_i)]^2 + (1 - \lambda_i)^2\ [\varepsilon[\gamma_0 + \gamma_1 w_i)^2] - [\varepsilon(\gamma_0 + \gamma_1 w_i)]^2] =$
$= \lambda_i^2\ \nu(d_i) + (1 - \lambda_i)^2\ \nu(\gamma_0 + \gamma_1 w_i) =$
$\lambda_i^2\ \nu(d_i) + (1 - \lambda_i)^2\ \gamma_1^2\ \nu(w_i)$

Luego la $\nu(\delta_i^*)$ es un valor comprendido entre los límites $\gamma_1^2\nu(w_i)$ y $\nu(d_i)$. Sin embargo, para comprobar la tesis inicial es necesario probar ahora que $\gamma_1^2\nu(w_i) \leq \nu(d_i)$. O, lo que es lo mismo, que $\gamma_1^2\nu(w_i) \leq \nu(\delta_i^*) \leq \nu(d_i)$ supone probar que la varianza del estimador empírico bayesiano es como máximo igual a la del estimador clásico.

Se sabe que $\nu(d_i) = \varepsilon(d_i^2) - \varepsilon(d_i)^2$ y que $d_i = \gamma_0 + \gamma_1 w_i + u_i + e_i$. Al desarrollar esta expresión se obtiene el valor de d_i^2:

$d_i^2 = \gamma_0^2 + (\gamma_1 w_i + u_i + e_i)^2 + 2\gamma_0\ (\gamma_1 w_i + u_i + e_i) = \gamma_0^2 + 2\gamma_0\ (\gamma_1 w_i + u_i + e_i) + \gamma_1^2 w_i^2 + (u_i + e_i)^2 + 2\gamma_1 w_i\ (u_i + e_i) = \gamma_0^2 + 2\gamma_0\gamma_1 w_i + 2\gamma_0 u_i + 2\gamma_0 e_i + \gamma_1^2 w_i^2 + u_i^2 + e_i^2 + 2u_i e_i + 2\gamma_1 w_i u_i + 2\gamma_1 w_i e_i$

El valor de la esperanza matemática de d_i^2 es entonces:

$\varepsilon(d_i^2) = \gamma_0^2 + 2\gamma_0\gamma_1\varepsilon(w_i) + 0 + 0 + \gamma_1^2\ \varepsilon(w_i^2) + \varepsilon(u_i^2) + \varepsilon(e_i^2) + 0 + 0 + 0 = \gamma_0^2 + 2\gamma_0\gamma_1\ \varepsilon(w_i) + \gamma_1^2\ \varepsilon(w_i^2) + \sigma^2_{ui} + \sigma_{ei}^2 + 0$

Los términos que se hacen cero en la expresión anterior es debido a la ausencia de relación entre los términos de error y el resto de los factores. Se quiere hacer notar que las varianzas σ_{ui}^2 y σ_{ei}^2 son los valores τ_{11} y τ_{00}.

Por otra parte, es necesario calcular el valor de la esperanza matemática al cuadrado de d_i. Si $\varepsilon(d_i) = \gamma_0 + \gamma_1\ \varepsilon(w_i)$, entonces:

$$\varepsilon(d_i)^2 = \gamma_0^2 + \gamma_1^2\ \varepsilon(w_i)^2 + 2\gamma_0\gamma_1\ \varepsilon(w_i)$$

Al sustituir estas dos expresiones en la inicial $\nu(d_i) = \varepsilon(d_i^2) - \varepsilon(d_i)^2$:

$$\begin{aligned}\nu(d_i) = {} & \gamma_0^2 + 2\gamma_0\gamma_1\ \varepsilon(w_i) + \gamma_1^2\ \varepsilon(w_i^2) + \sigma_{ui}^2 + \\ & \sigma_{ei}^2 + 0 - \gamma_0^2 + \gamma_1^2\ \varepsilon(w_i)^2 + 2\gamma_0\gamma_1\ \varepsilon(w_i) = \\ & \gamma_1^2\ [\varepsilon(w_i^2) - \varepsilon(w_i)^2] + \sigma_{ui}^2 + \sigma_{ei}^2 + 0 = \\ & \gamma_1^2\ \nu(w_i) + \sigma_{ui}^2 + \sigma_{ei}^2 + 0\end{aligned}$$

Dado que $\sigma_{ui}^2 + \sigma_{ei}^2 \geq 0$ y que $2|\sigma_{uei}| \leq \sigma_{ui}^2 + \sigma_{ei}^2$, entonces se comprueba que $\nu(d_i) \geq \gamma_1^2\ \nu(w_i)$, verificándose así que el valor de la varianza del estimador empírico bayesiano va a tener como máximo el valor de la varianza del estimador clásico. Con lo que se demuestra que el estimador empírico bayesiano es, como mínimo, tan eficiente como el clásico.

CONCLUSIONES

La magnitud del efecto individual es calculada a través de los estimadores clásicos y los empírico-bayesianos. El estimador clásico es el índice *d* de Hedges. Dentro de la perspectiva tradicional, este estimador se perfila como el estadístico menos sesgado y más eficiente (Hedges y Olkin, 1985; Marín, 1996).

Al mismo tiempo, el estimador empírico-bayesiano es una alternativa al clásico, que se aplica cuando es previsible un mal funcionamiento del estadístico d_i debido a distintas circunstancias de la investigación. Viene dado por la expresión:

$$\delta_i^* = \lambda_i d_i + (1 - \lambda_i)(\gamma_0^* + \gamma_1^* w_i)$$

La distribución de este estimador, también es normal, su media es δ_i y la varianza es τ_2. Este estadístico representa la estimación óptima de la magnitud del efecto individual desde la perspectiva de los modelos jerárquicos lineales. Esto es debido a que permite integrar en la misma expresión el conocimiento sobre un estudio individual ($\lambda_i d_i$) y el conocimiento sobre la relación de la muestra de estudios con características sustantivas para la explicación ($\gamma_0^* + \gamma_1^* w_i$). El estadístico δ_i^* no es otra cosa más que una media bayesiana que minimiza la pérdida cuadrática del error.

Así, parece que hay indicios para pensar en un posible distinto comportamiento de ambos índices de estimación de la magnitud individual. No se encuentra en la literatura meta-analítica ningún estudio de comparación del funcionamiento entre estimadores clásicos y empírico-bayesianos, por tanto no se tiene ninguna información procedente de estudios análogos.

En este trabajo se ha comparado el comportamiento del estimador empírico-bayesiano con el que se ha manifestado mejor estimador clásico pudiendo decirse que el estimador empírico-bayesiano:

- Tiende a infravalorar levemente el valor paramétrico central, aunque esto no es un claro indicador de mal funcionamiento, puesto que su rango de variación mucho menor que el del estimador clásico.
- Es clara y sistemáticamente más eficiente que el clásico.

En síntesis, se comprueba que los estimadores empírico-bayesianos, propuestos dentro de la perspectiva de los modelos jerárquicos lineales, suponen una clara aportación frente a los estimadores clásicos. Su comportamiento estadístico es aceptable y mejor que el de los estimadores tradicionalmente empleados, a pesar de la infravaloración del valor paramétrico. Además, es necesario destacar algunas de las virtualidades teórico-prácticas comprobadas de este tipo de estimadores, como es la mejora de la estimación a través de la consideración de la información procedente del conjunto de investigaciones primarias frente a la estimación clásica que incluye solamente la información procedente de un único estudio. Permiten superar el modelo de la media global, en tanto que son capaces de reflejar la presencia de una variable moderadora de los resultados, lo que supone la inclusión dentro del modelo la posibilidad de que no exista un único valor paramétrico central sino una distribución de parámetros. Por último, parece que el empleo de estimadores empírico-bayesianos puede suponer la necesidad de emplear menos estudios primarios para obtener estimaciones ajustadas, en tanto que el uso de la información *a priori* sobre el conjunto permite optimizar el uso de la información de cada investigación individual. Así, el estimador empírico-bayesiano de la magnitud del efecto individual se muestra como una alternativa clara a los estimadores clásicos.

Es especialmente destacable la aportación que supone la estimación empírico-bayesiana de la magnitud del efecto individual, tanto desde el punto de vista conceptual como desde el punto de vista empírico. Se dibuja, por tanto, una clara alternativa a la estimación clásica de la magnitud del efecto individual.

BIBLIOGRAFÍA

BRYK, A.S.; RAUNDENBUSH, S.W.: *Hierarchical linear models: applications and data analysis methods*. London, Sage, 1992.

CASTRO, M.: *Meta-análisis, aportaciones metodológicas a la síntesis cuantitativa de la evidencia. Un estudio de simulación Monte Carlo.* Madrid, Universidad Complutense, 1997.

EFRON, B.; MORRIS, C.: «Data analysis using Stein's estimator and its generalizations», en *Journal of the American Statistical Association,* 74(1975), pp. 311-319.

GLASS, G.V.: «Primary, secondary, and meta-analysis of research», en *Educational Researcher,* 5(10) (1976), pp. 3-8.

HEDGES, L.V.: «Distribution theory for Glass' estimator of effect size and related estimators», en *Journal of Educational Statistics,* 6 (2) (1981), pp. 197-228.

— «Fitting categorical models to effect sizes from a series of experiment», en *Journal of Educational Statistics,* 7 (1982), pp. 490-499.

— «A random effects model for effect size», en *Psychological Bulletin,* 93 (1983), pp. 388-395.

HEDGES, L.V.; OLKIN, I.: *Stadistical Methods for meta-analysis.* Orlando, Academic Press, 1985.

HOUSE, E.R.: «Realism in research», en *Educational Researcher,* august-september (1991), pp. 2-25.

JAMES, W.; STEIN, C.: «Estimation with quadratic loss», en *Fouth Berkeley Symposium on Mathematical Statistics and Probability,* (vol. 1). Berkeley, University of California, 1961.

LOUIS, T.A.: «Using empirical bayes methods in biopharmaceutical research», en *Statistics in Medicine,* 10 (1991), pp. 811-82.

LOUIS, T.A.; ZELTERMAN, D.: «Bayesian approaches to research synthesis», en H. COOPER y L.V. HEDGES (Eds.): *The handbook of research synthesis.* New York, Russell Sage Foundation (1994), pp. 411-422.

MARÍN, F: *Enfoques meta-analíticos: un estudio comparativo mediante simulación Monte Carlo.* Tesis Doctoral. Murcia, Universidad de Murcia, 1996 (inédito).

PILLEMER, D.B.; LIGHT, R.D.: «Synthesizing outcomes. How to use research evidence from many studies», en *Harvard Educational Review,* 50(2) (1980), pp. 176-195.

RAUDENBUSH, S. ; BRYK, A.: «Empirical Bayes meta-Analysis», en *Journal of Educational Statistics,* 10(2) (1985), pp. 75-98.

ROSNOW, R.L.; ROSENTHAL, R.: «Statistical procedures and justification of knowledge in psychological science», en *American Psychologist,* 44 (1989), pp. 1276-1284.