

INTERPRETACIÓN DE LA FIABILIDAD EN EL ANÁLISIS DOCUMENTAL MEDIANTE ELECCIONES DICOTÓMICAS Y SU INCIDENCIA EN LA INVESTIGACIÓN CUALITATIVA

Santiago Nieto Martín
Universidad de Salamanca

En el ámbito de la investigación educativa (y, en general, de las ciencias sociales), cuando de planteamientos metodológicos se trata, nos encontramos, en demasiadas ocasiones, ante una serie de dilemas intuitivamente explicados/resueltos por parte del investigador, que, más que puramente técnicos, tienen su anclaje en cierta contaminación formativa del propio ambiente investigador, y que, en todo caso, «facilitan» el proceso de trabajo.

En ese sentido, resulta interesante considerar las ideas manifestadas por Cicourel (1982), de quien Habermas (1988: 182) se hace eco en los siguientes términos: ...«nuestra falta de refinamiento metodológico significa que los procedimientos de decisión en orden a categorizar los fenómenos sociales están llenos de supuestos implícitos de sentido común acerca del actor, las personas concretas, y de los propios puntos de vista del observador acerca de la vida diaria. Los procedimientos parecen intuitivamente «correctos» o «razonables» porque están enraizados en la vida diaria».

Tal expresión adquiere singular relevancia cuando se trata de metodologías aplicadas a las distintas manifestaciones del comportamiento humano. La aplicación de técnicas de análisis documental en la investigación educativa requiere refinamientos técnicos, insuficientemente resueltos, a los que pretendemos aportar nuestra reflexión teórico-práctica en torno a un aspecto clave como es la fiabilidad de los análisis que se realizan.

1. LA FIABILIDAD Y VALIDEZ EN EL ANÁLISIS DOCUMENTAL

Muchas decisiones de trascendencia en educación se basan en análisis informales que carecen de fiabilidad y validez a causa de un alto nivel de subjetividad y de la ausencia de un enfoque sistemático en el análisis. Aunque en la medición educativa, fiabilidad y validez están relacionadas entre sí por diversas proposiciones, en el presente trabajo pretendemos poner de manifiesto algunas cuestiones metodológicas para la interpretación más adecuada de lo que debe ser la fiabilidad referida al análisis documental aplicado a la investigación educativa.

«La importancia de la fiabilidad procede de la seguridad que ofrece en cuanto a que los datos han sido obtenidos con independencia del suceso, instrumento o persona que los mide. Por definición, los datos fiables son aquellos que permanecen constantes en todas las variaciones del proceso de medición» (Kaplan y Goldsen, 1965: 83). La fiabilidad, por tanto, mide el grado en el cual, cualquier diseño

de investigación (total o parcialmente) o los datos resultantes de la misma, «representan variaciones en los fenómenos reales, en lugar de representar las circunstancias extrínsecas de la medición, las idiosincrasias ocultas de cada uno de los analistas o las tendencias subrepticias de un procedimiento» (Krippendorff, 1990: 192).

Sin duda, tenemos que admitir que para verificar la fiabilidad del análisis documental *se requiere una considerable duplicidad de esfuerzos*. En dichos análisis distinguimos, al menos, tres tipos distintos de fiabilidad: estabilidad, reproducibilidad y exactitud. Cada uno de ellos requiere un diferente diseño, pero, en cuanto a su eficacia, podemos establecer tres niveles:

ESTABILIDAD < REPRODUCIBILIDAD < EXACTITUD

El diseño que requiere el análisis de la fiabilidad como *estabilidad*, parte del comúnmente denominado «test-retest», mediante el cual, no obstante, aceptamos el razonable supuesto de que se pueden poner de manifiesto las sucesivas incongruencias del observador; por ello, se considera que es el menos eficaz de todos.

La fiabilidad como *reproducibilidad* («test-test»), requiere, al menos, dos observadores. Ello no evita los errores de incongruencia que puedan cometer los observadores, además del desacuerdo que se puede producir entre ellos. No obstante, es más eficaz que el anterior, siendo, a partir de este nivel, donde plantearíamos nuestra reflexión sobre el tema.

Sin duda, el más eficaz de los tipos de fiabilidad mencionados es el de *exactitud*. Para ello se establece un diseño «test-norma», donde sin embargo, no quedan eliminadas ni las incongruencias de los observadores, ni los desacuerdos entre ellos, y, donde, además, pueden aparecer desviaciones sistemáticas respecto a la norma.

2. ÍNDICE DE FIABILIDAD: CÁLCULO E INTERPRETACIÓN

Los datos sobre fiabilidad exigen que dos codificadores, como mínimo, describan de forma independiente un conjunto, probablemente elevado, de unidades de registro en términos de un lenguaje de datos común, que bien pueden ser en términos de categorías descriptivas, valores numéricos de una variable o, en muchas ocasiones, complejos esquemas clasificatorios.

Podemos considerar que la fiabilidad está garantizada cuando el acuerdo entre los codificadores es total; sin embargo, *si el acuerdo no es mayor que el correspondiente al azar, la fiabilidad es nula*. El problema radica pues en «medir» adecuadamente el acuerdo alcanzado por los observadores, codificadores o jueces al procesar de forma independiente la información científica que se somete a estudio e investigación.

Aunque nosotros vamos a exponer sobre un ejemplo concreto una metodología adecuada para valorar correctamente la fiabilidad de un análisis, hemos de manifestar que el hecho de que dos observadores (o más) *coincidan en un determinado porcentaje de acuerdo* en la asignación de los datos a las categorías, *no tiene ningún significado especial*. Y esto es muy importante considerarlo por cuanto en numerosos manuales sobre análisis documental se utiliza, como criterio de fiabilidad, el coeficiente porcentual alcanzado, y, sin duda, así considerado, resulta engañoso.

Supongamos, a modo de ejemplo, que queremos conocer ciertas características psicológicas de una persona a través del análisis de algunos documentos escritos y pertenecientes a esa persona. Para ello, dos codificadores con la debida preparación y formación para realizar dicho análisis, reciben el encargo de tratar de identificar una característica determinada en 15 documentos, marcándolos con el dígito «0» en el supuesto de no identificar nada relacionado con la instrucción recibida, o bien, con «1», en caso contrario.

Una vez realizado el mencionado trabajo, los datos de la situación «test-test» generada, se pueden representar de la siguiente manera:

Documentos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Codificador A	0	0	1	1	0	0	0	1	0	1	1	1	0	0	1
Codificador B	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0

El simple análisis de los datos muestra que los dos codificadores (u observadores) coinciden en once de un total de quince asignaciones (es decir, en el 73,3 %). Ahora bien, ese porcentaje que, en principio, no tiene por qué «explicarnos» mucho, *nada nos dice sobre si esa cifra es grande o pequeña, y, sobre todo, de que manera puede compararse con lo que hubiera sucedido por obra del azar*. Posiblemente no podamos dar respuesta convincente a esos interrogantes, pero, en cualquier caso, siempre hemos de tenerlos en cuenta. Para ello, los datos obtenidos en la matriz 2 x 15, los colocamos en una tabla de contingencia de frecuencias de 2 x 2. Así:

		Codificador B		
		0	1	
Codificador A	0	7	1	8
	1	3	4	7
		10	5	15

Deseamos obtener, a partir de los datos anteriores, una tabla semejante en la que el acuerdo sea producto del azar. Aunque ignoramos esa distribución, podemos estimarla a partir de la «experiencia» mostrada por los dos codificadores.

El Codificador A sostiene que ha identificado siete documentos conforme a la instrucción recibida, de un total de quince; sin embargo, el Codificador B mantiene que identificó cinco del total de quince. Por lo tanto, en conjunto, podemos decir que identificaron doce de un total de treinta (o sea, el 40 %: ese porcentaje se corresponde exactamente con el número de «1» en la matriz de 2 x 15).

Ahora bien, si los dieciocho «0» y los doce «1» los introdujéramos en una urna y fueran extraídos al azar por los dos codificadores, nos encontraríamos con las siguientes situaciones:

— La probabilidad de extraer dos «0» sin reposición (uno por cada observador), sería la siguiente: $(18/30)(17/29) = 0,351$; multiplicada esa cantidad por el número de unidades de registro (en nuestro caso, 15), obtenemos un valor aproximado de 5,3.

— Del mismo modo, al extraer dos «1», ésta sería la situación: $(12/30)(11/29) = 0,151$, que, multiplicado por 15, nos daría un valor aproximado de 2,3.

— Así mismo ocurriría al extraer «0» y «1»: $(18/30)(12/29) = 0,248$, con un valor aproximado de 3,7 al multiplicarlo por 15. Igual valor obtendríamos al extraer «1» y «0», respectivamente.

Es evidente que si se produjera un acuerdo perfecto, sólo «encontraríamos» valores en la diagonal principal, es decir, la que va del casillero 0-0 al casillero 1-1, dado que todo lo demás sería nulo; en ese sentido, el número máximo de veces que, sin reposición, obtendríamos el par 0-0, sería 9; así mismo, el número máximo de veces del par 1-1, sería 6. Por lo tanto, las tres tablas que nos han de servir de referencia para establecer una *medida adecuada de acuerdo* son:

<table style="width: 100%; text-align: center;"> <tr><td>9</td></tr> <tr><td>6</td></tr> </table>	9	6	<table style="width: 100%; text-align: center;"> <tr><td>7</td><td>1</td></tr> <tr><td>3</td><td>4</td></tr> </table>	7	1	3	4	<table style="width: 100%; text-align: center;"> <tr><td>5,3</td><td>3,7</td></tr> <tr><td>3,7</td><td>2,3</td></tr> </table>	5,3	3,7	3,7	2,3
9												
6												
7	1											
3	4											
5,3	3,7											
3,7	2,3											
Acuerdo máximo	Co-ocurrencias observadas	Acuerdo por azar										

Se considera que la medida más convincente para expresar la cantidad de acuerdo en los datos sobre fiabilidad resulta del *grado de semejanza entre la tabla de co-ocurrencias observadas y la tabla de acuerdo máximo*, en lugar del grado en que el acuerdo es meramente producto del azar.

La fórmula mediante la que calculamos el coeficiente de acuerdo buscado (α) es la siguiente:

$$\text{Co-ocurrencias observadas} = \alpha \left(\begin{matrix} \text{acuerdo} \\ \text{máximo} \end{matrix} \right) + (1 - \alpha) \left(\begin{matrix} \text{acuerdo} \\ \text{por azar} \end{matrix} \right)$$

Cuando la tabla de co-ocurrencias observadas sea igual a la tabla de acuerdo máximo, entonces $\alpha = 1$; por el contrario, $\alpha = 0$ cuando la tabla de co-ocurrencias observadas sea igual a la tabla de acuerdo por azar. Por lo tanto, tras diversas operaciones algebraicas, llegamos a la expresión:

$$\alpha = 1 - \frac{\text{discrepancia observada}}{\text{discrepancia prevista}}$$

Dado que en la tabla 2 x 2 las discrepancias aparecen en los casilleros 0-1 y 1-0, podemos establecer:

- Discrepancia observada: $3 + 1 = 4$
- Discrepancia prevista: $3,7 + 3,7 = 7,4$; luego:

$$\alpha = 1 - \frac{(3 + 1)}{(3,7 + 3,7)} = 0,4594$$

El coeficiente obtenido nos lleva a deducir que el acuerdo alcanzado resulta ser un 45,94 % por encima del obtenido al azar, lo que pone de manifiesto que los dos codificadores han leído e identificado correctamente la referencia señalada sólo el 45,95 % de todos los documentos. Este dato pone de relieve la necesidad de alejarse del uso indiscriminado de los acuerdos porcentuales como criterio de fiabilidad.

El coeficiente de acuerdo indica de qué manera es posible reconstruir la tabla original a partir de las dos tablas de acuerdo máximo y de acuerdo por azar, respectivamente:

A partir del acuerdo máximo: 0,4594	<table style="width: 100%; text-align: center;"> <tr><td>9</td></tr> <tr><td>6</td></tr> </table>	9	6	=	<table style="width: 100%; text-align: center;"> <tr><td>4,1</td></tr> <tr><td>2,8</td></tr> </table>	4,1	2,8				
9											
6											
4,1											
2,8											
A partir del acuerdo por azar: (1 - 0,4594)	<table style="width: 100%; text-align: center;"> <tr><td>5,3</td><td>3,7</td></tr> <tr><td>3,7</td><td>2,3</td></tr> </table>	5,3	3,7	3,7	2,3	=	<table style="width: 100%; text-align: center;"> <tr><td>2,9</td><td>2</td></tr> <tr><td>2</td><td>1,2</td></tr> </table>	2,9	2	2	1,2
5,3	3,7										
3,7	2,3										
2,9	2										
2	1,2										
Tabla simétrica equivalente a la observada:	<table style="width: 100%; text-align: center;"> <tr><td>7</td><td>2</td></tr> <tr><td>2</td><td>4</td></tr> </table>	7	2	2	4						
7	2										
2	4										

Las tablas siguientes son todas equivalentes entre sí en cuanto al grado de acuerdo/discrepancia:

7	4	7	3	7	2	7	1	7	0
0	4	1	4	2	4	3	4	4	4

Para finalizar, hemos de hacer necesaria referencia a que la cuestión planteada es la más elemental de cuantas pueden ofrecerse (dos codificadores y datos dicotómicos). Sin duda, resulta evidente pensar que la cuestión se complica a partir de lo que podemos considerar más normal en cualquier proceso de investigación; es decir, dos o más codificadores, y, sobre todo, si se amplía el número de categorías, o, incluso, el tipo de métrica. De ahí la prudencia que debemos conceder a la interpretación de coeficientes que en cada momento podamos obtener.

REFERENCIAS BIBLIOGRÁFICAS

- BARDIN, L. (1986): *Análisis de Contenido*. Madrid: Akal Universitaria.
 BERELSON, B. (1952): *Content analysis in communication research*. Glencoe: The Free Press.
 CICUOREL, A. (1982): *El método y la medida en sociología*. Madrid: Editora Nacional.
 FOX, D. (1981): *El proceso de investigación en educación*. Pamplona: EUNSA.
 HABERMAS, J. (1988): *La lógica de las ciencias sociales*. Madrid: Ed. Tecnos.
 KRIPPENDORFF, K. (1990): *Metodología de análisis de contenido*. Barcelona: Ed. Paidós Ibérica.