

## **EL ANÁLISIS ESTADÍSTICO Y LAS CIENCIAS SOCIALES**

Enrique Crivisqui

Abstract. There is a latent agreement between these two epistemologies when they reach their form - limit: the recognition that at the same time as they are different exists the necessity of the complementing of the statistical method and the sociological one. The social scientists accept of relative good degree this complementing and differentiates understanding it like a subordination of the instrument approach and formal from the social processes to the social theory. Nevertheless, it is common to find in the sociologists, critics not only to the explanations obtained like the imperfections of the statistical instrument. However, the statistical ones do not criticize the conceptual constructions of the sociological speech.

Hay un acuerdo latente entre estas dos epistemologías corporativas cuando alcanzan su forma - límite: el reconocimiento que a la vez que son diferentes existe la necesidad de la complementación del método estadístico y el sociológico.

Los científicos sociales aceptan de relativo buen grado esta complementación y diferencia entendiéndola como una subordinación de la aproximación instrumental y formal de los procesos sociales a la teoría social. Sin embargo, es común encontrar en ellos (los sociólogos) críticas no solo a las explicaciones obtenidas como a las imperfecciones del instrumento estadístico, lo que éste no hace de las construcciones conceptuales del discurso sociológico.

Muchas de las discusiones sobre las relaciones entre la estadística y las ciencias sociales parten de la idea de lo que Passeron (1991) denomina *“teoría imperial del*

*conocimiento sociológico*"<sup>1</sup>, en donde la significación del razonamiento estadístico solo es útil, ya que permite ilustrar las explicaciones sociológicas<sup>2</sup>.

Por su lado, los estadísticos tampoco se quedan atrás en materia de imperialismo corporativo. Acostumbrados a las exigencias de la representatividad de los datos y del síndrome de la confiabilidad y validez de la información y conociendo el costo para lograr homogeneizar datos sociales, resultan inevitablemente propensos, a una desconfianza generalizada respecto de todo lo que signifique una interpretación sociológica de ese dato manifestando una reticencia principista en contra de la *interpretación* conceptual, siempre sospechosa de sobreinterpretación política<sup>3</sup>.

Por lo contrario, aquí se parte del supuesto epistemológico de que *todas las conceptualizaciones operadas a partir de la observación del mundo empírico poseen, en tanto abstracciones científicas, una pertinencia empírica común.*<sup>4</sup> Es simplemente plantear que todas las enunciaciones formuladas por las diferentes ciencias sociales no pueden tener más que una única fuente: *la observación histórica por cualquier método que se la opere*<sup>5</sup>. Al igual se parte de la idea de que el razonamiento sociológico se distingue del razonamiento estadístico no por referirse a constataciones de una naturaleza distinta de las que son accesibles a la observación o medición de los fenómenos históricos, sino porque debe, para formular proposiciones dotadas de alguna generalidad, "adoptar un procedimiento específico de composición de los "enunciados de observación", que nunca es íntegramente reducible a un razonamiento experimental".<sup>6</sup>

La diferencia entre el razonamiento estadístico y el sociológico se sintetiza así en los dos usos que pueden hacerse *del análisis estadístico* cuando se lo utiliza en las

---

<sup>1</sup> Passeron, Jean-Claude, Lo que un cuadro dice y lo que de él se dice. El lenguaje de las variables y la interpretación en las ciencias sociales, 'Capítulo V, 1991, *Le raisonnement sociologique. L'espace non-poppérien nature*. Paris: Nathan, pp. 111-133 (versión castellana de Denis Barangcr).

<sup>2</sup> Passeron las denomina "constataciones ilustrativas.

<sup>3</sup> Idem. pág. 2

<sup>4</sup> Idem. pág. 3

<sup>5</sup> Idem. Pág. 4

<sup>6</sup> Idem. pág. 3

condiciones de la investigación social. Se puede, utilizarlo en forma plana o descontextualizada, o *bien* incorporando o contextualizando, a los resultados que de él se extraen, las características *socioculturales inherentes al dato sociológico*.

En las ciencias sociales, mantenerse en el marco de **un razonamiento estadístico estricto** supone omitir *momentáneamente* las condiciones contextuales que involucra cualquier dato extraído de la realidad social. Por lo tanto, cuando se quiere formular una generalidad o una interpretación sociológica, es necesario retornar al contexto socio cultural para incorporarlas a la formulación descriptiva del dato.

A su vez, ciertas corrientes de la investigación social, que excluyen el control formal o analítico de los instrumentos metodológicos que permiten una mayor validez, confiabilidad y representatividad de la información solo aceptarán ese mismo dato como parte de un discurso particular, generalmente especulativo y generalizante. Las condiciones de la investigación social no excluyen según Passeron (1991) “momentos de razonamiento experimental, pero obligan, si se quiere aplicar los resultados de ese razonamiento al mundo del cual proviene la información que procesa, a razonamientos sobre las condiciones limitativas de cada uno de los razonamientos experimentales, a constataciones sobre las condiciones de las constataciones. La composición de estos diferentes razonamientos no es otra cosa que un razonamiento sociológico”<sup>7</sup>.

El análisis estadístico se convierte *ipso facto* en formulación sociológica aún tratándose de la tabla de contingencia más simple, cuando apenas nos proponemos determinar la relación de una variable con otra o afirmar o negar algo. En el lenguaje de las variables: el simple hecho de tratar de determinar la lista de sus características o categorías de respuestas pertinentes se convierte en un proceso cualitativo donde siempre intervendrán, indirecta o subrepticamente signos culturales.

---

<sup>7</sup> Ibid. pág. 5

Cientistas sociales y estadísticos comparten la necesidad de prevenir a la vez, lo que se ha venido en llamar, la *ilusión del estadístico* y la *ilusión del sociólogo*. En otras palabras, deben protegerse de la *ilusión* que la estadística o cualquiera de las ciencias sociales podrían por sí sola, enunciar una generalidad cualquiera que fuera sobre el mundo de lo social. Una tabla estadística no dice jamás por sí sola cuando se le hace decir algo acerca del mundo de lo social, hay que hacerla hablar sociológicamente. Es por ello que el análisis e interpretación de una relación estadística demanda necesariamente operaciones de interpretación sociológica

### **DEL CONCEPTO AL INDICADOR**

El lenguaje de las variables si bien permite hacer lecturas de los datos más representativos o más interesantes presenta algunos problemas que es necesario tener en cuenta. Uno de ellos es la necesaria definición instrumental de los conceptos utilizados en las categorías o listados de sus características de lo que se estudia<sup>8</sup>.

El ejemplo típico es el de la lectura del sentido de las variaciones que aparecen en una tabla cruzada. Por ejemplo, en una tabla en la que la edad es tomada como variable independiente respecto a la cual se miden los diferentes valores tomados por una "variable dependiente" (**la apreciación sobre la solución de la crisis política**). Cuantitativamente la tabla nos dará unos porcentajes leibles. Pero, cuando nos detenemos a analizar las características que estamos tratando de medir de una variable como "edad" sería recomendable pensarlo dos veces. Se sabe que una variable como la edad puede tener por lo menos dos sentidos: puede estar registrando ya sea el "efecto de envejecimiento", ya sea el "efecto de generación". Es evidente que esta distinción no es inmediatamente codificable en el lenguaje de las variables; pero se impone apenas se intenta interpretar una tabla cruzada según

---

<sup>8</sup> Entre las numerosas operaciones involucradas en la construcción de una información cuantitativa, como las decisiones operacionales de un proceso de investigación y el diseño y aplicación de los instrumentos de medición o de procesamiento, a continuación se trabajara aquí una sola. Se trata de la convención instrumental que permite tratar una distribución de propiedades sociales como una "variable".



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

la edad para formular la variación sociológica revelada por la relación estadística, es decir apenas se pretende enunciar sociológicamente hablando lo que una tabla dice estadísticamente.

¿Qué "quieren decir", en efecto, tablas que presentan distribuciones cruzadas como la **Tabla 1** en la que "x" es la **edad de las personas entrevistadas**, repartidos en grupos de edad (x1: jóvenes; x2: adultos jóvenes; x3: adultos mayores, e "y" es la **apreciación sobre la solución de la crisis política con tres categorías de posibles respuestas**: y1: **Gobierno autoritario**; y 2: **Cambio de gobierno**; y 3: **Negociación política**.

**TABLA 1**

		Edad			
Apreciación sobre la solución de la crisis política		y1	y2	y3	
<b>x1</b>	20%	20%	60%	100 (n <sup>1</sup> = ...)	
<b>80</b>		15	5		
<b>x2</b>	30%	<b>60</b> 40%	30%	100 (n <sup>2</sup> = ...)	
20			20		
<b>x3</b>	60%	20%	20%	100 (n <sup>3</sup> = ...)	

¿Cómo decir del modo más adecuado posible la muy fuerte correlación estadística que se lee entre edad y formas de solución a la crisis política?

¿Para formular lo que una tabla de este tipo dice, se *debe verdaderamente interpretar*? ¿No es posible contentarse, para formular una interpretación sociológica, con decir lo que dice estadísticamente una tabla de frecuencias, sin incurrir en ningún riesgo en la formulación interpretativa?

La interpretación de los datos - formulada ya sea por un estadístico o por un sociólogo- comienza cuando se intenta contextualizar la información con otras constataciones teóricas o empíricas reconocidas o aceptadas, por ejemplo, retomar o aceptar una formulación conceptual que especifique lo que realmente mide "la edad de los sujetos".

En la tabla 1, ninguna lectura de la información permitirá hacerle decir a la variable “edad” si la variación de las distintas opiniones sobre la solución a la crisis política es producto<sup>9</sup> de:

- (a) un *efecto de envejecimiento*, es decir unas opiniones o actitudes que supuestamente variarían de acuerdo a los diferentes grupos de edad si se la midiera sucesivamente en distintas épocas, o
- (b) si se trata de un *efecto de generación*, es decir de la variación de unas opiniones o actitudes de una cohorte, independiente de la edad biológica.

En síntesis, de los datos de la Tabla 1, la única pregunta interpretativa que permitiría decir de ellos algo interesante, radica en la posibilidad de tener claridad conceptual sobre el real significado de las categorías que enmarcan las posibles respuestas o el origen metodológico de los datos: por ejemplo si son producto de la utilización de una encuesta panel o longitudinal o de tipo transversal.

Se podría mostrar, en nuestro ejemplo, el de una encuesta longitudinal que seguiría en el tiempo a una cohorte de individuos. La dificultad de interpretación se solucionaría en el seguimiento durante *n* tiempo (días, semanas, semestre) de los mismos sujetos lo que permitiría sin duda, al medir en el tiempo biográfico la variación de la actitud en las diferentes edades en la misma generación o cohorte **neutralizando** formalmente “el efecto de la generación”.

### ***El sentido de las relaciones entre variables***

El lenguaje de las variables impone no solamente un código o término alfanumérico que codifica unívocamente las posibles respuestas, sino también una forma de “leer” cuya operación central define siempre una forma de relación entre variables en una tabla de contingencia. Examinemos uno de los problemas que se plantean en la

---

<sup>9</sup> En la tabla 1, los valores numéricos que figuran en enunciados tabulares permitan afirmar, sin recurrir a otras informaciones fuera de las contenidas en la tabla, que existe una relación “estadísticamente significativa” entre las variables X e Y.

interpretación de un cruce de variables, aún cuando nos conformemos con leer lo que quiere decir, para una variable “x”, estar en relación con una variable “y” en una correlación estadística, es decir cuando nos limitamos al sentido poco exigente que encierra una afirmación del tipo: *los elementos de una población tienen, por el hecho de poseer ciertas propiedades, “probabilidades” más o menos grandes de tener tales o cuales otras propiedades.*

El ejemplo típico puede ser aquí el de la interpretación de los resultados producidos por el procesamiento en el “análisis multivariado” de los efectos de una o varias variables sobre una variable dependiente.

Así, la lectura de los resultados de un sondeo de opinión sobre la problemática de la educación, que apunta a establecer la acción de factores de selección como la estratificación social en los diferentes niveles de estudio, no puede interpretar *directamente* las frecuencias diferenciales de retención (deserción) escolar de los sujetos de distinta estratificación social, tales como se las observa en las etapas sucesivas del proceso educativo.

**TABLA II: RETENCION – DESERCIÓN ESCOLAR SEGÚN ESTRATO**

	1er CICLO retención escolar		2° CICLO retención escolar		3er CICLO retención escolar	
	Alta	Baja	Alta	Baja	Alta	Baja
Estratos Altos	70 99% 30	30 1% 70	55 90% 45	45 10% 55	30 80% 70	70 20% 30
Estratos Bajos	86%	14%	78%	22%	80%	20%

Al analizar lo que dice la tabla II, se aprecia una relación que une la pertenencia a un estrato a la probabilidad relativamente más o menos fuertes de retención escolar, la interpretación sociológica debería tener en cuenta **otros hechos ausentes de las relaciones explícitamente formulada** por la propia tabla cruzada de que las poblaciones sobre las cuales se establece la variación de la relación entre la pertenencia a una categoría de estrato y retención escolar, “habitan” condiciones de escolaridad diferentes (infraestructura), formas de escolaridad diferentes, procesos de socialización diferenciados, etc.

El *contextualizador* de la interpretación es evidentemente aquí es el concepto de “estratificación social” que evidentemente tiene tantas acepciones como uso se le da. Importaran entonces las diferentes acepciones y conceptualizaciones teórico metodológicas que existan lo cual constituirá la contextualización necesaria a la interpretación:

Se observa en efecto retornando a la Tabla II que dos lecturas son posibles:

a) Atenerse al razonamiento empírico que sugiere la tabla estadística cuando se la autonomiza. Se puede entonces hacerle decir - lo que es verdadero si se prescinde de las condiciones en las cuales “viven” las poblaciones sometidas a la encuesta - que los estratos bajos otorgan menores probabilidades de retención escolar en el primer ciclo de la enseñanza formal, por lo contrario ofrece mayores chances en el último. Más brevemente: *la relación entre estratificación social y retención escolar se invierte de la etapa 1 a la etapa 3.*

Es este un enunciado exacto mientras nos limitamos a la información sobre la cual se ha construido la tabla y **no le planteamos ninguna pregunta a partir de otras informaciones.**

El riesgo, del cual no nos escapamos mediante ese enunciado condicional que omite enunciar una de las condiciones esenciales de la formulación, consiste evidentemente, puesto que todo enunciado de apariencia sociológica tiende a través de la generalidad del concepto, en dar como verdadero y estadísticamente fundado un enunciado que en sí sería falso: *el concepto de estratificación social denota características que, siendo intrínsecamente favorables a la retención escolar en los primeras etapas de enseñanza formal, se convierten en desfavorables cuando se trata de las competencias escolares requeridas en el más alto nivel de los pensum de la educación superior.*

b) Recurrir al razonamiento - sociológico que permite tomar en cuenta el hecho, ausente de la tabla estadística, de que la mortalidad escolar diferencial ha deformado a través de las etapas la estructura interna de las poblaciones sobrevivientes sobre las cuales se desea establecer la influencia de la estratificación social. La tabla dice entonces (y en consecuencia resulta engañosa cuando no se le hace decirlo) que *la estratificación social limita siempre la retención escolar de los sujetos de origen*

*popular considerados en su conjunto; pero que, si se aísla del conjunto de su cohorte los sujetos de extracción popular que en el estadio 3 no han resultado pura y simplemente eliminados y que por este hecho han sido sobre seleccionados, la relación se invierte en la sub-cohorte de los sobrevivientes de origen popular.*

En resumen, la interpretación sociológica siempre tiene por función el interrogarse sobre las condiciones sociales, económicas, demográficas, culturales, etc., que configuran los sujetos de estudio.

Ello es totalmente evidente, por ejemplo, en el caso de un sondeo de opinión sobre prácticas lectoras de una la población de los usuarios de centros de documentación en la Universidad de Antioquia (de facultades, de escuelas, etc.). Las relaciones que se observan en el sondeo, por una parte, estudiante según programa académico, grupos de edad, estrato o de sexo y, por la otra, las prácticas o actitudes lectoras, si bien se las podría generalizar fuera de la institución a los grupos de usuarios de otros centros externos se correría el riesgos de generalizaciones “fuera de contexto”. Lo que se podría aceptar sería realizar algunas generalizaciones más con el objeto de ilustrar que de dar por cierto para todos algo que lo es solo para una población específica.

### **RELACIONES ENTRE VARIABLES SOCIALES**

Generalmente se podría estar tentado de creer que todas las dificultades en la interpretación de las relaciones entre variables obedecen al hecho de que ellas no representan la real riqueza de lo observado o medido. En la investigación social, es el “diseño cuantitativo” a menudo concebido como una forma de verificación de las hipótesis es él que mejor ilustraría esta paradoja metodológica.

La lectura o formulación estadística de los efectos cruzados puede realizarse olvidando la estructura de la *población de referencia*. Pero la interpretación propiamente sociológica de las variaciones no puede efectuarse más que

restituyendo a los diferentes valores asumidos por una variable los diferentes sentidos que le marcan, en el contexto real de las multicausalidades.

La interpretación sociológica por lo tanto sólo puede ser definida como un razonamiento condenado a moverse **y** a operar compromisos argumentativos entre las exigencias lógicas del razonamiento empírico y las exigencias descriptivas de la contextualización social. La investigación social que recurre al lenguaje de las variables en el análisis de los hechos sociales deja en efecto de aparecer como un *dilema* o, al menos, una *tensión metodológica* entre la descripción socioeconómica y cultural, que impone el conocimiento de lo social y el razonamiento empírico, concebidos para establecer la independencia o detallar la interacción de las variables en juego<sup>10</sup> o las constataciones de relación fundadas sobre esas co-ocurrencias o como confirmaciones generales que tratan acerca de correlaciones entre *variables puras*. Evidentemente no es posible privarse de los servicios que el recurso del lenguaje de las variables presta a la investigación aplicada y al control de la comprobación de relaciones, pero tampoco se puede ignorar las limitaciones **y** las tareas impuestas a esta metodología de reconstrucción **y** de procesamiento de los hechos.

---

<sup>10</sup> Lemaine G., Lemaine J-M., *Psychologie sociale et expérimentation*. París. Mouton. pp. 19-26.