

Una técnica para analizar los efectos de composición grupal

JAMES A. DAVIS

JOE L. SPAETH

CAROLYN HUSON

Se propone una técnica para aislar la contribución hacia la probabilidad de ocurrencia de un atributo dependiente de: 1) atributos individuales y 2) composición grupal en términos de los atributos de los miembros. Se propone una clasificación de posibles tipos de relaciones que se pueden identificar mediante la técnica, se discuten problemas estadísticos y se presentan dos ejemplos empíricos.

Ni el problema que comentaremos ni su solución general son nuevos. De hecho, el germen de la idea completa está en *The Rules of Sociological Method*¹ de Durkheim. El problema es el siguiente: Contando con datos sobre individuos con agregados específicos (grupos, condados, naciones), ¿qué inferencias se pueden hacer sobre la naturaleza de un efecto a distintos niveles de agregación? Por lo tanto, por ejemplo, el hallazgo que las naciones católicas tienen tasas más bajas de suicidio es consistente con cualesquiera de las siguientes inferencias: 1) los católicos tienden menos a suicidarse independientemente de la composición religiosa de la nación; 2) en países con muchos católicos, tanto los católicos como los protestantes tienden menos a suicidarse, pero dentro de un país determinado no hay diferencia religiosa en el suicidio; 3) los católicos tienden más a suicidarse que los protestantes, pero ambas religiones tienen tasas más bajas de suicidio donde hay muchos católicos, y 4) independientemente del país, los católicos tienen una tasa constante de suicidio, pero la tasa de suicidio para los protestantes baja con la proporción de católicos. Todas estas aseveraciones declaran cierto tipo de relación entre la religión y

*Esta investigación fue conducida como parte de un estudio del programa de Great Books, bajo una subvención de The Fund for Adult Education. Se leyó una versión preliminar de este trabajo en las reuniones de The American Sociological Association en 1959.

el suicidio, pero difieren mucho en sus implicaciones substantivas, y en sus suposiciones sobre la naturaleza del efecto al nivel individual (diferencias religiosas dentro de determinada nación) y al nivel del grupo (diferencias entre personas de la misma religión en distintas naciones).

El principio estadístico de estas ambigüedades ha sido reconocido por bastante tiempo. Ha sido expresado negativamente como "la falacia de correlación ecológica".² Patricia Kendall y Paul F. Lazarsfeld, sin embargo, al analizar problemas en *The American Soldier*, sugieren que el principio puede conducir a ganancias positivas si los datos permiten análisis simultáneos a nivel de grupos y de individuos. Dicen:

No existe razón alguna por la cual no se pueden usar datos unitarios para caracterizar a los individuos en la unidad. Una persona que no tiene paludismo en una unidad donde la incidencia de paludismo es muy baja, probablemente tiene una opinión muy distinta sobre su estado de salud que una persona que no tiene paludismo, pero presta servicios en una unidad con alta incidencia. . .

En términos de análisis real, se puede reformular la cuestión en los siguientes términos: al igual que podemos clasificar la gente por variables demográficas o por sus actitudes, también los podemos clasificar por el tipo de ambiente en que viven. Las variables adecuadas para dicha clasificación probablemente serán los datos unitarios. Un análisis por encuesta entonces abarcaría los datos personales y unitarios simultáneamente.³

Desde entonces varios estudiantes y colegas de Kendall y Lazarsfeld han acudido a este enfoque como estrategia de investigación. Por lo tanto, Berelson, Lazarsfeld y McPhee muestran que entre grupos de amistad el porcentaje que vota a favor del Partido Republicano aumenta conforme aumenta la proporción en el grupo cuya afiliación política es el Partido Republicano, tanto para individuos que son partidarios de los republicanos como para los que son partidarios de los demócratas.⁴ O bien, Lipset, Trow y Coleman⁵ en su estudio del reporte de I.T.U. muestran que en talleres donde hay un consenso político, el interés político es más alto (independientemente de preferencias políticas) que en talleres donde hay una división en lealtad política.

Estos hallazgos, como era de esperarse, no son nuevos. Durkheim mismo, en su trabajo sobre suicidio notó no solamente que las tasas de suicidio varían considerablemente entre distintas religiones, pero también que, para una religión dada, las tasas de suicidio son más bajas cuando sus miembros forman una clara minoría en la sociedad.⁶ Groves y Ogburn, en un libro publicado en 1928,⁷ mostraron que las tasas de matrimonio entre los hombres y mujeres varían en direcciones opuestas a las proporciones por sexo de las comunidades

en las que viven. Al igual, Faris y Dunham, en su estudio de 1939 de la distribución ecológica de los domicilios de psicóticos en Chicago descubrieron que algunas tasas psicóticas eran más altas para los negros que viven en áreas blancas y los blancos que viven en áreas negras que para las mismas razas que viven en áreas donde formaban la mayoría.⁸

Y, naturalmente, varios hallazgos en *The American Soldier* fueron de este tipo, incluyendo el análisis citado frecuentemente, que mostró que mientras que los soldados ascendidos criticaban menos el sistema de promoción militar que los soldados que no fueron ascendidos, existían mayores críticas hechas tanto por los soldados ascendidos como no ascendidos en unidades con una alta tasa de ascensos.⁹

En un artículo publicado en 1957, Peter Blau denominó dichas relaciones “efectos estructurales” y los definió de la siguiente manera:

El principio general es que si la X del ego afecta no solamente la Y del ego pero altera también la Y, se observará un efecto estructural, que significa que la distribución de X en un grupo está relacionada a Y aún cuando la X del individuo se mantiene constante. Dicho hallazgo indica que la red de relaciones en el grupo con respecto a que X influye sobre Y. Aísla los efectos de X sobre Y que se deben completamente a o son transmitidos por el proceso de interacción social.¹⁰

Nuestro análisis surge de las ideas presentadas por el grupo de Columbia University, y difiere solamente en la siguiente forma: en primer lugar, llamaremos a estos fenómenos “composicionales” en vez de “estructurales”, porque pensamos que sólo hay un traslazo parcial entre estas relaciones y lo que los sociólogos consideran ser la estructura social; en segundo lugar, sugeriremos algunos procedimientos formales y sencillos para el análisis estadístico; en tercer lugar, en vez de considerar nada más los efectos individuales contra los efectos de grupo, vamos a sugerir una tipología de relaciones posibles.

Procedimientos

Para pasar de un análisis esencialmente verbal a operaciones estadísticas concretas, tendremos que empezar con una breve serie de definiciones y anotaciones.

Consideraremos dos o más poblaciones o grupos, que vamos a identificar con números arábigos. Dentro de cada población, se pueden caracterizar los individuos por la presencia o ausencia de un atributo *independiente* (A ó \bar{A}). Por consiguiente, cada población puede caracterizarse por la proporción de sus miembros que tienen el atributo, que también es la probabilidad de que un miembro escogido al

azar es una "A". Kendall y Lazarsfeld se refieren a dicho dato de nivel de grupo como una tasa.¹¹ Denominaremos estas tasas o probabilidades P_1, P_2, P_3 , etcétera, el suscrito designando la población de la que se trata. Dentro de cada población también podemos determinar la probabilidad de que un miembro de una subclase definida por la presencia o ausencia de A tiene un atributo *dependiente* dado. Por lo tanto, vamos a definir D_{a1} como la probabilidad de que una A en la población, tenga el atributo dependiente. Finalmente, la probabilidad de que cualquier miembro no escogido de un grupo específico tenga el atributo dependiente se identificará mediante D_1, D_2, D_3 , etcétera.

La estrategia subyacente de cualquier análisis de esta naturaleza consiste en comparaciones entre las probabilidades D en distintas categorías.¹² En nuestra situación, hay dos ejes básicos de comparación: a) podemos comparar miembros de un grupo específico que pertenecen a distintas subclases del atributo independiente, y b) podemos comparar miembros de la misma subclase que pertenecen a distintos grupos. El resultado terminal será una serie de diferencias en probabilidades o porcentajes, una serie de *diferencias dentro del grupo* y una segunda serie de *diferencias entre grupos*.

Un ejemplo hipotético puede aclarar estas diferencias. Supongamos que hemos conducido una encuesta en cuatro comunidades distintas y estamos investigando la relación entre la religión y el egoísmo de acuerdo con Durkheim. Los resultados de este estudio hipotético se presentan en la siguiente tabla ficticia.

TABLA I
RESULTADO DE ENCUESTA HIPOTETICA

<i>Ciudad</i>	<i>Oneville</i> <i>Prot.Cat.</i>	<i>Twoville</i> <i>Prot.Cat.</i>	<i>Threeville</i> <i>Prot.Cat.</i>	<i>N-ville</i> <i>Prot.Cat.</i>
Porcentaje	77 43	68 32	59 21	86 54
egoístas	600 400	400 600	200 800	800 200

La tabla I puede describirse con la terminología delineada antes en la forma siguiente:

D = El porcentaje sufriendo del egoísmo (por ejemplo, 59 por ciento entre protestantes en Threeville)

A = protestantes

\bar{A} = católicos

P = El porcentaje de la ciudad que es protestante (por ejemplo, 20 por ciento en Threeville).

Por lo tanto $D_{\bar{A}2}$ es el porcentaje de egoístas en la comunidad 2

(32 por ciento); D_{A1} es el porcentaje de egoístas entre los protestantes en la comunidad $_1$ (77 por ciento).

Hagamos ahora los dos tipos de comparación. Dentro de cada comunidad, cuando comparamos los católicos y los protestantes, vemos que los protestantes tienden más a ser egoístas que los católicos. Entonces, cuando comparamos los protestantes en distintas comunidades, notamos una gama de valores (77-68-59-86); al igual que cuando comparamos los católicos en distintas comunidades (43-32-21-54).

Concluimos que dentro de una comunidad dada, existe una diferencia relacionada con la religión, y entre comunidades hay una diferencia que controla el factor religión. ¿Hemos pues demostrado un efecto de composición? Aún no, puesto que no hemos demostrado que cualquiera de estas dos diferencias esté relacionada a la composición de población. Para expresarlo formalmente:

Existe un efecto de composición cuando el tamaño absoluto de ya sea a) las diferencias dentro del grupo y/o b) la diferencia entre grupos para las A's y/o las \bar{A} 's puede describirse como una función de P.

Una forma de hacer esto es cambiar de una tabla de porcentajes a una gráfica en la cual el eje vertical es la probabilidad D, el eje horizontal es P, y los puntos son los valores D para las subclases en las poblaciones bajo consideración. La figura 1 ilustra la gráfica para nuestra encuesta hipotética.

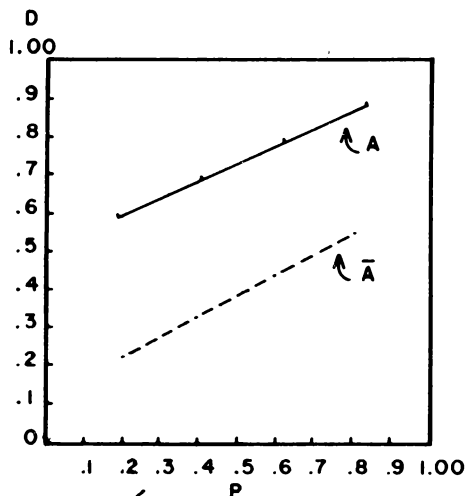


FIG.1 GRÁFICA DE EFECTO DE COMPOSICIÓN HIPOTÉTICO.

Puesto que la figura 1 es ficticia, es agradable y clara. Al examinarla cuidadosamente, notamos lo siguiente:

a) Entre ambas A's (línea sólida) y las \bar{A} 's (línea punteada) hay un aumento lineal de D conforme P aumenta.

b) Las dos líneas que conectan las observaciones son paralelas.

De estas dos observaciones, podemos llegar a las siguientes inferencias sobre nuestro efecto hipotético composicional.

a) Hay un efecto de composición muy marcado sobre la diferencia entre grupos, porque tanto entre las A's y las \bar{A} 's, D aumenta uniformemente con P; por lo tanto, la variación en la composición social produce un efecto sobre el comportamiento aún cuando se controla el atributo individual.

b) No hay efecto de composición sobre la diferencia dentro del grupo, puesto que para todos los niveles la diferencia entre D_A y $D_{\bar{A}}$ es constante; por lo tanto, la variación en la composición social no produce efecto alguno sobre el comportamiento dentro de un grupo, fuera del efecto explicado por la diferencia de nivel individual común en todo el universo.

Como sucede a menudo cuando uno procede a formalizar una aseveración verbal, las posibilidades resultan ser más numerosas de lo esperado. Por lo tanto, parece ser que puede haber siquiera dos tipos de efectos de composición bastante distintos, uno, un cambio en las diferencias entre grupos, y otro, un cambio en la diferencia dentro del grupo, conforme P varía. Puesto que, además, las líneas no tienen que ser rectas, es evidente que potencialmente, siquiera, en vez de un efecto de composición "a", puede haber una gran variedad de tipos. En la siguiente sección presentaremos una clasificación de tipos de efectos de composición.

Una clasificación de efectos de composición

Cuando definimos los efectos de composición como funciones tomamos el cuidado de no fijar límites sobre el tipo de función involucrada, pero simplemente declaramos que dichos efectos serían reconocidos cuando una o ambas de nuestras diferencias varían con P. Por lo tanto, no existe razón, en principio, para que el efecto de composición no tome la forma de cualquiera de los espirales y curvas que se proliferan en textos matemáticos.

Sin embargo, aun en trabajo preliminar con una técnica, algún intento de clasificación ayudará a estructurar el análisis. Por lo tanto, vamos a sugerir que de las posibilidades infinitas, hay ocho que son de importancia estratégica. Nuestro sistema de clasificación combina

algunas consideraciones matemáticas con especulación sobre los procesos sociales que pudieran generar las formas matemáticas.

Empecemos con la matemática. Una de las características atractivas del análisis de composición es que en el proceso, los datos cualitativos se trasmutan a datos cuantitativos. Es decir, aunque A es un atributo u observación cualitativa, P puede considerarse como un número real. Tiene un punto cero significativo, las distancias entre los valores P pueden tener números reales, y los valores pueden elevarse al cuadrado legítimamente. Dada esta característica, es posible describir las funciones mediante las cuales se definen los efectos de composición en términos de análogos geométricos estándar de ecuaciones algebraicas —lineales, monotónicos, parabólicos, etcétera.

Ahora, consideremos los efectos posibles de A sobre D.

1) El atributo puede aumentar o disminuir la probabilidad de D, independientemente de P. Esta es la diferencia de nivel individual.

2) P puede influir sobre la probabilidad de D entre las A's y \bar{A} 's igualmente. Esta es la diferencia de nivel de grupo.

3) Puede haber una interacción en los efectos. Aunque lógicamente son equivalentes, hay dos formas diferentes de describir el efecto. En primer lugar, uno puede decir que P influye sobre la probabilidad de D, pero diferencialmente para las A's y las \bar{A} 's. Podemos considerar que esto es una "susceptibilidad diferencial" puesto que dichas relaciones sugieren que las A's y las \bar{A} 's difieren en sus relaciones a la variación en composición. Alternativamente, podemos decir que A puede aumentar o disminuir la probabilidad de D pero diferencialmente a distintos niveles de P. Podemos considerar esto como "diferencias individuales condicionales" puesto que dichas relaciones sugieren que el tamaño del efecto de nivel individual está relacionado con la composición de grupo. Estas interacciones ni son diferencias de nivel de grupo ni diferencias de nivel individual, pero más bien una combinación de ambas.

¿Qué deben mostrar nuestras gráficas para estos tres tipos de efectos?

a) Si está operando un efecto de nivel individual, algunos de los parámetros, de las funciones para las A's y las \bar{A} 's van a diferir, y, geoméricamente, serán descritos por dos líneas distintas o curvas. Si un proceso como éste no está operando, las funciones para las A's y las \bar{A} 's serán idénticas, y, geoméricamente, solamente habrá una línea, independientemente de que haya o no un efecto de nivel de grupo.

b) Si hay un efecto de nivel de grupo, P o una función de P será un

parámetro en la función, para las A's y las \bar{A} 's, y, geoméricamente, ambas líneas se alejarán de una línea recta paralela al eje P. Si no hay un efecto como éste, ambas líneas serán paralelas al eje P.

c) Si este tercer tipo de efecto ocurre, P o una potencia de P será un parámetro en las funciones, pero P tendrá distintos valores para las A's y las \bar{A} 's. Gráficamente, habrá dos líneas o curvas que no son paralelas.

Ahora estamos listos para examinar los tipos específicos de los efectos de composición. Empecemos dividiendo las posibilidades entre dos grupos grandes: primero, los casos lineales, en los cuales las ecuaciones que describen D como función de P no involucran potencias de P más grandes o menos grandes que uno, y por lo tanto pueden graficarse con una línea recta, y en segundo lugar, los casos en que una o más de las funciones no es lineal.

EFECTO NIVEL INDIVIDUAL	INTER-ACCIÓN	EFECTO NIVEL DE GRUPO	
		NO	SI
NO	NO	TIPO 0 	TIPO II
SI	NO	TIPO I 	TIPO III a TIPO III b
SI	SI	LOGICAMENTE IMPOSIBLE	TIPO IV a TIPO IV b

FIG. 2 CLASIFICACIÓN DE RELACIONES PARA EL CASO LINEAL.

La figura 2 resume las cinco posibilidades básicas en el caso lineal. Dentro de cada celda hemos colocado una gráfica que ilustra los resultados geométricos. En todas nuestras gráficas observaremos estas convenciones: a) el eje vertical es la probabilidad de D de 0 a 1.00; b) el eje horizontal es el valor de P de 0 a 1.00; c) la línea sólida conecta puntos que representan valores de D entre las A's para valores de P particulares; d) la línea punteada conecta puntos representando los valores de D entre las \bar{A} 's, para valores de P particulares.

Consideremos cada uno de los cinco tipos, uno por uno.

Tipo 0. Tipo 0 es un caso trivial. Tanto A como \bar{A} producen la misma línea horizontal paralela al eje P. En el tipo 0, A no tiene efecto alguno sobre D. No hay diferencia de nivel individual, no hay diferencia entre grupos, y no hay diferencia de tercer tipo.

Tipo I. En una relación tipo I vemos dos líneas horizontales, paralelas la una a la otra y al eje P. Esto significa que en cada nivel de P hay una diferencia dentro del grupo entre las A's y las \bar{A} 's, y esta diferencia es constante, la relación interpretándose como lo que hemos denominado un efecto puro de nivel individual.

Tipo II. Una relación tipo II está polarmente opuesta al tipo I, correspondiente a un efecto de grupo "puro", en el cual observamos una diferencia entre grupos relacionada a P, pero ninguna diferencia dentro de grupo. El resultado final es una línea recta sola tanto para las A's y las \bar{A} 's, según la ilustración en la figura 2.

Las relaciones del tipo II presentan algunos problemas interesantes para su interpretación. El proceso de influencia aquí puede considerarse como actuando en la ausencia de cualquier diferencia de característica individual. El efecto posiblemente pueda considerarse como "católico", en el cual la influencia de un atributo funciona afectando el clima del grupo o medio sin influir sobre los individuos directamente.

Tipo IIIa. En una relación tipo IIIa vemos una combinación del tipo I y el tipo II. Es decir, hay una diferencia individual constante, junto con un efecto lineal sobre la composición del grupo. En la figura 2 vemos que la representación gráfica de un efecto como éste consiste en dos líneas paralelas que suben o bajan con P. Substantivamente sugiere una especie de efecto de acarreo.

Tipo IIIb. Tipo IIIb es igual que el tipo IIIa, excepto por el hecho paradójico: el nivel individual es opuesto al efecto de nivel de grupo.¹³ En la figura 2 vemos un ejemplo donde, dentro de grupos, el ser A conduce a menores probabilidades de D, mientras que entre grupos, la proporción de las A's lleva a un aumento en D entre ambas subclases. Uno pensará que es una de aquellas cosas poco probables arrojada al enumerar las posibilidades lógicas, excepto que éste es el

modelo del hallazgo clásico de la Fuerza Aérea-Policía Militar: Dentro de unidades, los soldados ascendidos se sienten más contentos con el sistema de ascensos, pero entre unidades, la proporción que aprueban disminuye con el porcentaje ascendido, entre los dos tipos de soldados. Como Blau señala en su discusión de efectos estructurales, estos fenómenos son sumamente importantes para el sociólogo puesto que son confirmaciones notables de su aseveración que los efectos de nivel de grupo difieren de los efectos de nivel individual.

Tipo IV. Las relaciones del tipo IV consisten en aquellas formas donde hay dos líneas rectas que no son paralelas. Esto corresponde a lo que hemos denominado susceptibilidad diferencial o diferencias individuales condicionales. La diferencia en pendiente de las dos líneas puede interpretarse como susceptibilidad diferencial, al proceso de las A's y de las \bar{A} 's, o como un cambio en la correlación de nivel individual. Puesto que hay muchas formas en que las líneas pueden tener divergencias, evidentemente hay muchos tipos distintos de relaciones tipo IV. En la figura 2, sin embargo, hemos escogido tres que son interesantes. En el ejemplo IVa, vemos una situación donde ambos grupos muestran un aumento lineal con P, pero la tasa de incremento, o pendiente, de la línea es mayor para un grupo (aquí, las \bar{A} 's) que para el otro. Es decir, en el ejemplo, las \bar{A} 's son más sensibles a la variación en composición que las A's. En el tipo IVb, vemos que la línea A está paralela al eje P, y la línea \bar{A} muestra un aumento lineal. Esto sugiere que solamente las \bar{A} 's son sensibles a compo-

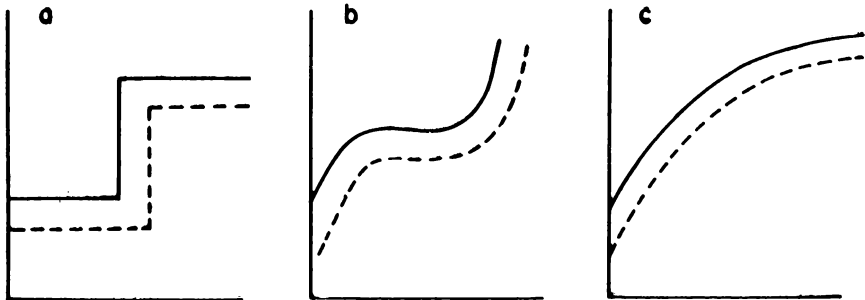


FIG. 3 FORMAS POSIBLEMENTE MONOLÍTICAS

sión, mientras que las A's son inmunes a este efecto. Finalmente, en el tipo IVc, vemos una situación donde ambos grupos muestran aproximadamente la misma cantidad de efecto, pero la dirección del efecto se invierte. Es decir, la tasa de D aumenta con P para un grupo y disminuye para el otro. De hecho, una relación tipo IV no es tan eso-

térica como parece ser, porque simplemente se puede interpretar como decir que para cada grupo D es una función lineal de su posición de mayoría-minoría.

A pesar de sus diferencias evidentes, todas las relaciones lineales tipo IV comparten una característica: en cada una la *diferencia en porcentaje* entre las A's y las \bar{A} 's es una función lineal de P. Es decir, la correlación de nivel individual varía directamente como función de P. En los tipos IVa y IVb la correlación mantiene dirección, pero varía en magnitud, y en el caso de IVc, vemos una inversión del signo de la correlación junto con una disminución lineal en la diferencia de porcentajes.

Esto completa nuestra descripción de las cinco formas lineales principales.¹⁴ No hemos podido dar una codificación completa de los tipos no lineales, pero bosquejaremos algunos tipos que merecen discusiones específicas, aunque breves.

En ciertos casos las formas que aparecen no serán lineales, pero mostrarán una dirección definitiva. Desde un punto de vista técnico, las podemos llamar funciones monotónicas e incluyen todas las formas en las que la gráfica no es lineal, pero donde las curvas resultantes nunca suben y bajan. Las funciones de etapas, en las que para ciertos rangos de P no hay efecto, pero en otras hay un aumento precipitado; curvas acelerantes y desacelerantes y ciertas combinaciones de las dos podría incluirse aquí. La figura 3 ilustra algunas posibilidades *a priori*.

El ejemplo "a" en la figura 3 ilustra una función escalonada, que sugeriría que hay algo de umbral necesario antes de que ocurra el efecto de influencia de composición. En el ejemplo "b" vemos una situación donde la tasa de cambio es mayor en los extremos que en el medio del rango, que en términos de un modelo de influencia sugeriría una situación donde el proceso de influencia era muy efectivo cuando uno de los grupos de "influenciadores" dominaba claramente, una situación a veces sugerida en la investigación de propaganda y comunicaciones. El ejemplo final, "c", simplemente es un caso donde D aumenta a una velocidad decreciente con un aumento en P, una situación que sugiere substantivamente algo que se parece a un proceso de saturación.

Las formas monotónicas pueden además incluir cualquiera de los tipos básicos de relaciones como las hemos definido para los casos lineales; todos los ejemplos en la figura 3 son del tipo IIIa.

Un segundo tipo de efecto composicional no lineal es uno en el que las curvas producidas tienen sólo una moda o pico, es decir, donde la curva que describe la relación tiene un solo doblez, la familia general de parábolas es un buen ejemplo. Substantialmente, sugiere

ren ya sea un punto “óptimo” o el resultado de consenso. Por lo tanto, por ejemplo, si (como en el ejemplo citado de mera *Union Democracy*) la unanimidad era importante para que D ocurriera, deberíamos esperar encontrar los niveles de D a valores de P sumamente altos o bajos serían idénticos, pero que la curva alcanzara un mínimo más o menos de .5 donde la heterogeneidad era la más alta.

Técnicas de análisis

Las consideraciones presentadas anteriormente sugieren que la técnica básica para el análisis de efectos de composición es la regresión bivariada estándar y el análisis de covariaciones. Sin embargo, los problemas de las N's y grados de libertad son difíciles con estos datos, y la estrategia de análisis vale la pena discutirse.

Empecemos con el enfoque general. Suponemos que en cualquier situación de investigación práctica, los datos serán una muestra de probabilidades o, si es un universo, nunca encajarán en cualquiera de las formas de modelos sugeridas sin cierto grado de error. Por lo tanto, la cuestión básica en el análisis de composición no es: “¿qué forma tiene?”, sino más bien: “¿a qué formas se aproximan los datos en forma razonable? Sugerimos que las hipótesis nulas implicadas son las siguientes: a) Estos datos se podían haber muestreado de un universo en el cual no hay efecto de composición, y b) si la hipótesis nula a) debe rechazarse, se podrían haber muestreado estos datos de un universo en el cual el efecto es el menos complicado desde un punto de vista matemático.

Estas dos consideraciones sugieren que los pasos naturales en dichos análisis son los siguientes:

- 1) Adaptando la línea más recta posible para las A's y también para las \bar{A} 's, usando el criterio de cuadrados mínimos.
- 2) Si no se pueden rechazar las hipótesis nulas de linealidad, probando el distanciamiento de la(s) pendiente(s) de cero, probando la posibilidad de una diferencia en pendientes, y probando si las líneas, en caso de ser “paralelas”, son significativamente diferentes en su punto de intercepción.
- 3) Si se rechaza la hipótesis nula de linealidad, probando el ajuste de funciones sucesivamente más complicadas hasta lograr ajuste bastante razonable o hasta abandonar el análisis.

El enfoque de cómputo general es aquel de procedimientos de textos estándar para regresión, pero surge un problema difícil cuando uno considera el problema de N. Si se muestrea más de un elemento de una colectividad dada, la muestra está muy apiñada, y el tamaño

de muestra efectiva es mucho menor que el número de individuos muestreados. Por otra parte, si se muestrean los individuos con procedimientos de muestreo sencillos y aleatorios dentro de grupos, el tamaño de muestra efectiva es mayor que el número de grupos, aunque menor que el número de individuos. Los manuales de estadísticas de muestreo sí nos dan fórmulas para dichas situaciones, pero nuestra experiencia práctica ha sido que los cálculos son demasiado complicados como para que sean útiles para situaciones de encuestas donde no se están probando hipótesis específicas. Nuestra convención actual es seguir adelante como si N equivaliera al número de agregados,¹⁵ pero es una convención conservadora y desperdicia datos. Hemos podido hacerlo porque hemos usado la técnica con situaciones donde el número de grupos era bastante grande. Esperamos que si la técnica resulta ser útil para estudios con un pequeño número de grupos se pueden elaborar pruebas precisas pero sencillas.

Ejemplos

Modelos del tipo que hemos descrito sólo son útiles en investigación si el mundo real se comporta un poco como el álgebra. Mientras que los ejemplos citados anteriormente de la literatura ilustran que hay efectos de nivel de grupo, esto no significa que todas o aun una gran parte de las relaciones sociológicas muestran estos efectos, o que todos los modelos abstractos tienen un referente empírico. Podemos decir que en la National Opinion Research Center, en los últimos dos años hemos aplicado esta técnica en dos estudios, uno de grupos de discusión de educación para adultos y uno de estudiantes graduados de humanidades y ciencias (donde la unidad de análisis es el departamento). En ambos estudios han surgido muchos efectos de composición, y aunque todavía no hemos enfrentado un tipo IVc claro, hemos visto más de uno de cada uno de los otros tipos. En nuestra investigación, el tipo IIIa parece ser el más común, aunque esto puede ser un artefacto de los problemas y variables escogidos en nuestros estudios.

Para ilustrar el grado de aproximación que uno puede esperar en la investigación real, quisiéramos presentar dos ejemplos empíricos breves. Ambos son de un estudio del programa The Great Books, un programa de educación adulta a nivel nacional.¹⁶ Los datos están basados en cuestionarios autoadministrados de 1 909 miembros de 172 grupos de discusión de The Great Books.

La figura 4 es una relación empírica del tipo II o un efecto de nivel de grupo puro. D es la probabilidad de que un miembro de un grupo de discusión específico lo abandonara entre la encuesta original en 1957 y un año de observación en 1958. El atributo A está basado

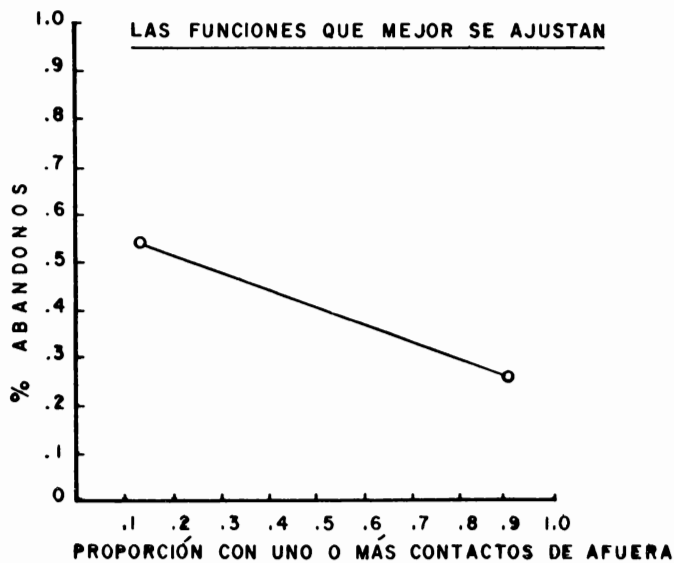
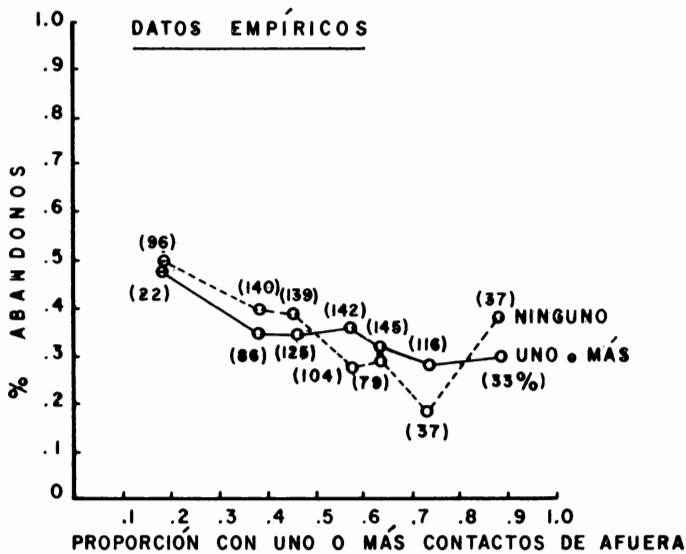


FIG. 4 CONTACTOS DE AFUERA Y ABANDONOS

en respuestas a la pregunta: “¿Cuántos miembros de su grupo (excluyendo su cónyuge) ve usted regularmente fuera de las discusiones del grupo?” Las A’s son las personas que reportan uno o más contactos de afuera, las \bar{A} ’s son las personas que reportan cero contactos de afuera. P, por consiguiente, es la proporción de miembros que tienen

uno o más contactos de afuera. La figura 4 muestra la relación (los resultados se presentan para intervalos agrupados de P ; los números entre paréntesis son el número de individuos sobre los cuales está basado un valor dado de D). La cifra superior muestra los resultados empíricos. No hay diferencia de nivel individual consistente dentro de grupos de una composición dada, pero conforme aumenta la proporción de un contacto de afuera, decaen las tasas de pérdidas. Por lo tanto, "contactos de afuera" muestra una relación de nivel de grupo con retención en estos grupos de discusión, pero no muestra diferencias de nivel individual. Bajo los datos, vemos la función con mejor ajuste, como una sola línea recta.

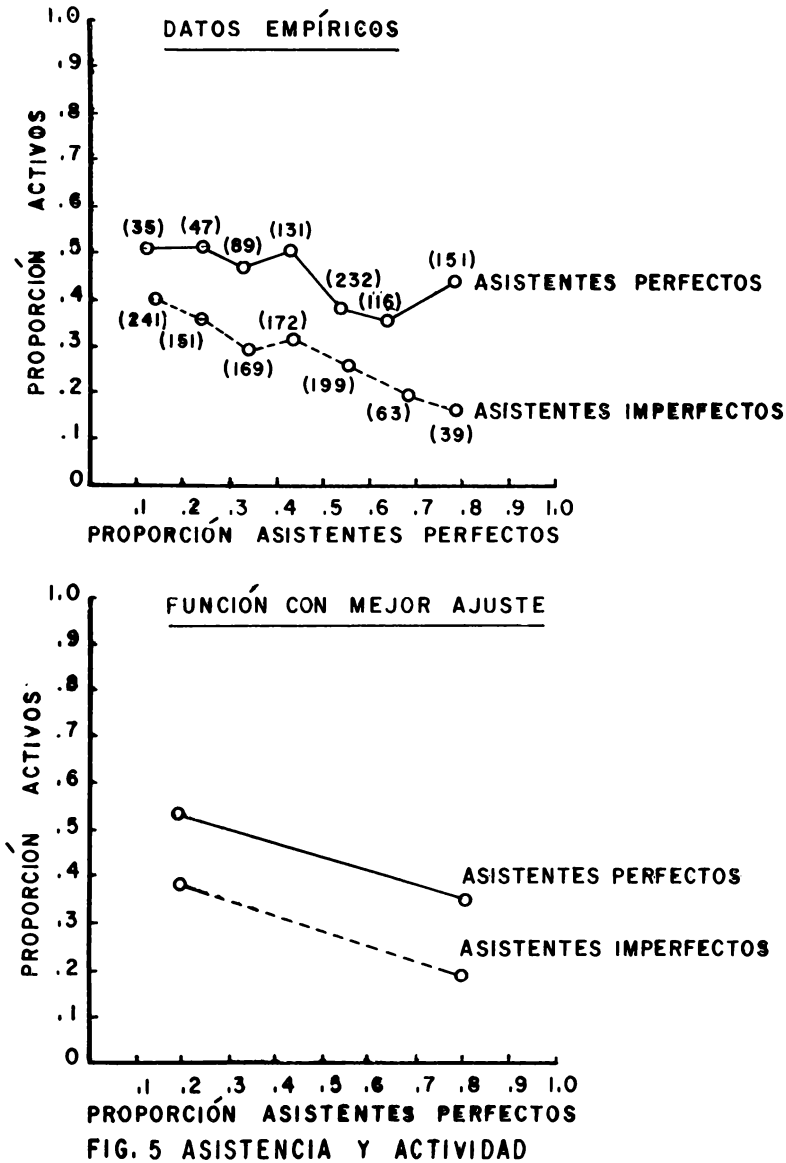
La figura 5 ilustra una relación del tipo IIIb, una relación de bumerang en la cual la diferencia individual se va en un sentido y la diferencia de nivel de grupo en otro. D , aquí, es la probabilidad de que un miembro dado será mencionado por los miembros de su grupo como un participante activo en las discusiones. El atributo A es si el miembro se reporta a sí mismo como siendo un asistente con asistencia perfecta en una reunión, en contraposición a reportarse como faltando a las reuniones ocasional o frecuentemente. Dentro de grupos de una composición dada, los asistentes perfectos tenderán más a ser citados como participantes activos; pero entre mayor la proporción de asistentes perfectos, menor será la proporción de aquellos citados como activos. Creemos que la relación ocurre debido a las siguientes razones: grupos con muchos miembros denominados como participantes activos tienen una tasa de supervivencia más alta que los grupos con participación más limitada. Cuando los grupos empiezan a titubear, los miembros ocasionales son los que lo abandonan primero.

Por lo tanto, los grupos con una alta proporción de asistentes perfectos son aquellos que se han reducido a su membresía seria, debido a una baja participación en las discusiones, aunque dentro de grupos, los asistentes perfectos tenderán más a ser los participantes activos.

Resumen

Hemos bosquejado una técnica cuya intención es codificar análisis en situaciones donde uno desea evaluar el efecto de un atributo dado, tanto en términos de características de individuos como en términos de una característica de agregados a los que pertenecen los individuos. El análisis de las posibilidades formales sugiere que son posibles una gran variedad de relaciones como éstas. Sugerimos un esquema para clasificar algunos tipos de efectos de composición como éstos, definiéndolos en términos de las funciones matemáticas involucradas, y sugiriendo algunos de los procesos sociales subyacentes que los pudieran generar. Ejemplos empíricos de un estudio del programa The

Great Books demostró que estos efectos sí existen empíricamente, aunque sería necesario llevar a cabo una gran variedad de estudios antes de poder llegar a conclusiones generales sobre las contribuciones relativas de los factores de niveles individual y de grupo en fenómenos sociológicos.



NOTAS

¹Emile Durkheim, *The Rules of Sociological Method* (publicado por primera vez en 1895), 8ª edición, traducido por S.A. Solovay y J.H. Mueller, y publicado por G.E.G. Catlin, Glencoe, Ill., The Free Press, 1950.

²W.S. Robinson, "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, 15 (junio de 1950), pp. 351-357.

³P.L. Kendall y P.F. Lazarsfeld, "Problems of Survey Analysis", en R.K. Merton y P.F. Lazarsfeld, editores, *Continuities in Social Research: Studies in the Scope and Method of the American Soldier*, Glencoe, Illinois, The Free Press, 1950, pp. 195-196.

⁴B.R. Berelson, P.F. Lazarsfeld, y W.N. McPhee, *Voting: A Study of Opinion Formation in a Presidential Campaign*, Chicago, Illinois, University of Chicago Press, 1954, pp. 100-101.

⁵S.M. Lipset, M.A. Trow y J.S. Coleman, *Union Democracy*, Glencoe, Illinois, The Free Press, 1956, pp. 163-171.

⁶Emile Durkheim, *Suicide. A Study in Sociology* (publicado por primera vez en 1897), traducido por J.A. Spaulding y G. Simpson, Glencoe, Illinois, The Free Press, 1951.

⁷Ernest R. Groves y William F. Ogburn, *American Marriage and Family Relationships*, Nueva York, Henry Holt y Cía., 1928, pp. 193-205.

⁸R.E.L. Faris y H.W. Dunham, *Mental Disorders in Urban Areas*, Chicago, Illinois, The University of Chicago Press, 1939, pp. 110-123.

⁹S.A. Stouffer, *et al.*, *Studies in Social Psychology in World War II*, vol. 1, *The American Soldier During Army Life*, Princeton, New Jersey, Princeton University Press, 1949.

¹⁰Peter M. Blau, "Formal Organization: Dimensions of Analysis", *American Journal of Sociology*, 63 (julio de 1957), p. 64.

¹¹P.L. Kendall y P.F. Lazarsfeld, *op.cit.*, p. 191.

¹²Mientras que la discusión en este artículo se limita al caso donde la característica dependiente es una dicotomía (estadísticamente, una proporción), no hay motivo, en principio, para que no se pueda aplicar nuestra técnica a atributos dependientes politomos, o a otras estadísticas (medias, variancias, medianas) que se aproximan a los datos. Para simplificar la exposición, sin embargo, sólo consideraremos dicotomías. El uso de estadísticas más sofisticadas hará que el investigador entre más en problemas aún más complejos al aplicar "pruebas de significancia" a los resultados.

¹³Una teoría formal que intenta explicar un efecto tipo IIIb ha sido sugerida por uno de los autores. Cf. James A. Davis, "A Formal Interpretation of the Theory of Relative Deprivation", *Sociometry*, 22 (diciembre, 1959), pp. 280-296.

¹⁴Robert McGinnis, en una comunicación privada a los autores, notó que para el caso lineal, un esquema más fácil de clasificación resultaría de una comparación de pares de coeficientes de las ecuaciones lineales para las A's y las \bar{A} 's. En términos geométricos esto equivaldría a preguntar si las pendientes eran iguales, cero, o desiguales, y si la intercepción de A es mayor que, menor que, o igual que la intercepción para las \bar{A} 's. Hemos conservado nuestro marco original para clasificación porque creemos que es más sugestivo de interpretaciones substantivas para el investigador, pero debemos notar que es menos parsimonioso.

¹⁵En cualquier serie de datos hay, naturalmente, más de una manera de formar dichos agregados. Una técnica de esta naturaleza es agregar datos en base a los valores de P agrupados. Por lo tanto, si se han agrupado los datos en décimas (deciles), N será diez. Se pueden basar otros métodos en membrecías de grupos como condados, compañías, o unidades del ejército.

¹⁶Un análisis de estos datos, usando esta técnica, se reporta en *Great Books and Small*

Groups: Studies of the Dynamics of Participation in a Program for Adult Liberal Education,
National Opinion Research Center, Report 72.