

El Método Experimental

Por F. STUART CHAPIN

Profesor emérito de Sociología de la Universidad de Minnesota, Estados Unidos de América. Colaboración especial para el Número Jubilar de la Revista Mexicana de Sociología, vertida del inglés por Oscar Uribe Villegas.

EL método experimental, que se inició con las investigaciones de la ciencia física y que se asocia con el ilustre Galileo (1564-1642), se ha convertido en el método básico de la ciencia, pero sólo recientemente se ha aplicado al estudio de las relaciones humanas.

Un resumen detallado de los veinticinco años recién pasados de investigación sociológica con el método experimental, fue el proporcionado por Ernest Greenwood en su libro *Experimental Sociology* (15) en 1945. En esa monografía, de entre los 150 títulos de la bibliografía, hay 86 que corresponden a experimentos reales. No es necesario que repitamos esta información tan accesible. En general, con todo, ha habido un uso creciente del método experimental, desde 1930. A pesar de que Comte (7), hace ya un siglo, mencionó la experimentación como una de las tres técnicas generales de investigación en la dinámica social, parece que tenía en mente la observación de los llamados "experimentos naturales" en los que ciertas formas de la patología individual eran observados por el sociólogo que informaba de ellas en alguna publicación. No fue sino hasta 1931 cuando S. C. Dodd (9) realizó el primer estudio realmente vigilado y señoreado de la relaciones comunitarias, conforme al informe que proporciona en su libro sobre *A Controlled Experiment on Rural Hygiene in Syria*, en 1934. Este estudio midió ciertos cambios en las prácticas personales higiénicas, ocurridos entre 1931 y 1932, en una aldea experimental que recibió servicios de salubridad de una clínica viajera de la Fundación para el Cercano Oriente. Tres aldeas que no recibieron estos servicios constituyeron los grupos

contrastantes o de toque. Se construyeron y normalizaron escalas elaboradas, con el fin de medir las prácticas higiénicas. Puesto que este extenso estudio es muy accesible en las grandes bibliotecas, no lo resumiré. Más recientemente (1947), mi libro *Experimental Designs in Sociological Research*, revisado en 1955 (5), y el de René König: *Beobachtung und Experiment in der Sozialforschung* (18), publicado en 1956, describen estudios específicos y descienden al detalle práctico. De ahí que, a partir de este punto, omita toda presentación del desarrollo histórico y me concentre en las dificultades prácticas presentes y en sus posibles soluciones, puesto que sólo dicho énfasis parece que puede resultar útil para estudiosos e investigadores.

Los estudios sistemáticos de poblaciones relativamente grandes y/o de aquellas que abarcan muchos factores, están siendo analizados actualmente, con ayudas de computadoras o mediante cerebros mecánicos. Quienes lean este trabajo indudablemente contribuirán a este enfoque moderno. Por mi parte, considero que mi tarea debe consistir en analizar algunos breves estudios exploratorios adaptados para la investigación de los estudiosos. Estos estudios son útiles en cuanto pueden sugerir nuevas hipótesis, y proporcionar incluso pruebas de dichas hipótesis conduciendo hacia estudios más amplios y cuidadosamente diseñados.

La investigación sociológica trata de obtener resultados que permitan generalizaciones acerca de los efectos de la exposición del comportamiento humano a algunas influencias o tratamientos. Una base firme para tal razonamiento puede depender de la habilidad del estudiante para vigilar y señorear, mediante el método experimental, aquellas variables que confunden las relaciones entre la presunta causa y el efecto. Algunas de estas variables son conocidas; otras son desconocidas. Pero, todas estas variables pueden ser vigiladas y señoreadas mediante una selección aleatoria de los sujetos humanos que hay que estudiar. Esto es así porque la selección aleatoria es una elección no prejuiciada o sesgada, y, de este modo, todos los tipos de variables, conocidas o desconocidas, tienen iguales oportunidades de ser incluidas o de estar representadas en la muestra que ha de investigarse. La aleatorización inicial es básica para el cómputo de estimaciones probabilitarias válidas. Esto es de importancia científica, porque, en el caso de las medidas sociales, tenemos pocas oportunidades de conseguir, en nuestras aproximaciones, una plena certeza, de modo que los grados de probabilidad están vinculados y dependen del criterio principal de interpretación válida.

En los estudios agrícolas y en las ciencias biológicas, que tienen que ver con organismos vivos, el método de análisis de la variancia y de la

covariancia, inventado por R. A. Fisher (10 y 11) en la tercera década de este siglo, ha tenido una aplicación amplia y fructífera bajo la forma de diseños para someter experimentalmente a prueba datos obtenidos de la observación. En estas ciencias de la vida, el estudio de los fenómenos propios de los organismos vivos colocados en sus habitats naturales es un aspecto importante de la investigación. Mientras que ciertos problemas muy específicos del comportamiento pueden estudiarse ventajosamente bajo una vigilancia y señorío rígidos de la situación artificial propia del laboratorio, hay otros problemas principales que necesitan estudiarse y que se disgregan cuando se analizan fuera del ambiente natural en el que ocurren (como cuando se les lleva al laboratorio). Entre los grandes méritos del análisis de la variancia y de la covariancia se encuentra el de que dicho análisis permite no sólo el estudio experimental de la vida de plantas y animales *in situ*, sino también el de que es un método de vigilancia y señorío de las variables perturbadoras a través de la aleatorización, puesto que considera muchos factores que varían simultáneamente. También hay que considerar que el mismo proporciona estimaciones probabilitarias.

Ahora bien, es característico de los problemas de relación humana el que son muchos los factores que se presentan juntos, en patrones complicados e interrelacionados o en racimos de características interconectadas. Esta red de influencias recíprocas y de fuerzas sociales exhibe su forma más natural en la situación familiar de la comunidad. Consecuentemente, podemos considerar el análisis de la variancia como un modelo que nos puede guiar en el estudio experimental de situaciones humanas reales, vitales. Así, por ejemplo, podemos diseñar experimentos para que éstos se realicen en la comunidad; experimentos que nos permitan medir los efectos esperados de ciertos programas para el tratamiento social de problemas como el de la delincuencia juvenil, el de la conducta criminal, el de los malos alojamientos, etcétera. Hasta aquí, nuestra evaluación de la efectividad de tales programas comunitarios referentes a servicios especiales para individuos desajustados han gravitado sobre la mera opinión o, en el mejor de los casos, sobre las "opiniones expertas" acerca de "hechos" que, por sí mismos, resultan frecuentemente dudosos. Y, sin embargo, anualmente se gastan, en todo el mundo, millones en ciertos esfuerzos para mejorar, aliviar, vigilar y señorear o prevenir la aparición de problemas sociales mediante servicios especiales que se consideran apropiados para quienes los necesitan. Una contribución importante a la ciencia sociológica y a la política social sería la que consistiera

en adaptar fructíferamente el método experimental a fin de procurar estimaciones basadas en algo más sistemático y que, especialmente, tuviera una base científica; estimaciones que pudieran dar una medida de los presuntos efectos de muchos programas aprobados socialmente y que, hasta ahora, no han sido revalidados.

Puede notarse que hablo sólo de “adaptaciones” y no de una transferencia directa de la experimentación de laboratorio y del análisis de la variancia al marco comunitario. Como ya se ha establecido, una transferencia directa de las complejas situaciones vitales al ambiente artificial del laboratorio (cuando esto es posible, en experimentos más o menos simples) o sobresimplifica el problema en estudio o rompe la fina malla de las relaciones humanas; con esto, el problema se transforma en nuestras manos, a causa de nuestros esfuerzos para estudiarlo. Sin embargo, podemos escapar de estas complicaciones de laboratorio “adaptando” el análisis de la variancia a estudios experimentales del comportamiento humano *in situ*. ¿Qué quiere decir “adaptar” en este caso? Necesitamos recordar, tan sólo, que un análisis válido de la variancia debe satisfacer las condiciones siguientes: 1. Que los casos por estudiar se seleccionen al azar; 2. Que las variables medidas se distribuyan normalmente en la población más amplia a la que deseemos generalizar los hallazgos que obtengamos a partir de la muestra; 3. Que la desviación media cuadrático en los subgrupos aleatorios (o réplicas) de la muestra, sea aproximadamente igual en su valor (homoesquedasticidad), y 4. Que las intercorrelaciones entre las variables sean lineales.

Desgraciadamente, en contraste con lo que ocurre en la investigación agrícola, en la investigación sociológica es común el que una o más de estas condiciones no puedan satisfacerse. Por fortuna, la invención de los análisis no paramétrico de la variancia, realizada por Milton Friedman (13) en 1937, ayuda a superar algunas de estas dificultades, pues libra al investigador social de restricciones como las que se refieren a la normalidad de la distribución o a la homoesquedasticidad y a la linealidad. Pero, junto con tales simplificaciones del problema del método analítico de los datos, nos podemos privar de muchas de las ventajas técnicas de una teoría bien desarrollada de la factura estadística de hipótesis (8, 13) y de su docimasia basadas en los parámetros.

Un investigador de los procesos sociales ¿puede seleccionar casos al azar, para sus estudios experimentales, en vez de elegirlos porque los necesite? Aquí nos encontramos, de inmediato, con las objeciones expresadas por los dirigentes morales, por los políticos y por los administradores de los programas de servicio. Tanto las mores locales como

las prescripciones legislativas insisten en que el tratamiento debe de aplicarse a causa de la necesidad humana, y este principio se ha interpuesto en el camino de las aplicaciones del método experimental en las ciencias sociales. A más de este obstáculo, existe también oposición a que se realice una experimentación directa con los seres humanos. Volveré a referirme a estas dificultades en otro contexto, para explicar cuál es la forma en que pueden superarse sin lesionar los principios humanitarios o las libertades humanas individuales.

Las consideraciones anteriores sobre los problemas prácticos y técnicos de aplicación del método experimental en sociedad, me hicieron sugerir, en 1947, que podría ser útil el que el estudioso adoptara un enfoque diferente. Sugerí que, lógicamente, hay tres clases muy amplias de diseño experimental que aplicar en la investigación sociológica (5, pp. 29-33). La primera consiste en el análisis de un corte transversal, hecho en una época dada, en el que se usa la lógica de la correlación parcial y de la regresión múltiple. Este tipo de diseño no establece la existencia de relaciones causales, debido, en gran parte, a la ausencia del factor tiempo en los estudios comparativos del corte. En el análisis causal debemos hipotetizar (o postular) antecedentes, y dar los acontecimientos subsecuentes en calidad de presuntos efectos o consecuencias. El método del corte transversal tiene, con todo, el valor de sugerir cuáles son las variables que pueden requerir vigilancia y señoría y, en ocasiones, sugiere cuál de entre las varias hipótesis puede ser más fructífera (5, pp. 135-6). Pero, este diseño transversal y sus limitaciones ahora nos son familiares (21) y, por ello, no me extenderé en él.

El segundo tipo de diseño experimental (6), puede designarse como “diseño *proyectado*” de experimentación social. Aquí aplicamos la lógica clásica de la experimentación mediante observaciones “ante-” y “post-” expositivas de la situación, a algún conjunto de influencias o tratamientos. El grupo tratado se describe, de acuerdo con medidas del presunto efecto, antes de que se aplique el tratamiento, en t_1 . En seguida, las segundas medidas se hacen al terminarse el tratamiento, en t_2 . La diferencia entre el valor de estas dos medidas se analiza después, para determinar si pueden aceptarse como medidas válidas de los presuntos efectos del tratamiento.

Una alternativa familiar de este método de auto-comparación en dos fechas diferentes consiste en conducir el estudio simultáneamente en dos grupos: un “grupo experimental” que recibe el tratamiento, y un “grupo contrastante” o “de toque”, al que no se le aplica el tratamiento. Este

método del “grupo de toque” descansa en las diferencias esperadas entre los dos conjuntos de medidas en la fecha terminal (t_2). Se da por supuesto el que aunque ambos grupos muestren los mismos valores en la misma fecha, o sea, en la “ante-exposición”, o en t_1 , mostrarán diferencias significativas entre ellos en la fecha terminal (t_2), como en el método autocomparativo, y se supone también que, cuanto tales diferencias se encuentran en la fecha terminal (t_2), podemos valorar los presuntos efectos del tratamiento. En tales estudios, el grupo de toque nos da información o puntos de referencia, o una imagen de lo que pudo ocurrir a un grupo no tratado entre la primera fecha (t_1) y la fecha terminal (t_2). Puesto que el factor tiempo se introduce en este diseño experimental proyectado, podemos usarlo por lo menos como aproximación en nuestro deseo de entender la causación en la sociedad.

El tercer diseño experimental aplicado en la investigación sociológica, puede designarse como diseño *ex post facto*. Es el reverso del diseño proyectado en la dirección de nuestra búsqueda de posibles relaciones causales. Aquí, tomamos alguna diferencia observada y medida en el presente (t) entre el grupo experimental que ha recibido tratamiento por un cierto periodo y un grupo de toque que no lo ha recibido. En seguida, nos remontamos en el tiempo, hasta un punto arbitrario, de origen del tratamiento, para captar cuáles son los factores que han operado en ese intervalo, factores que pueden explicar las diferencias observadas entre el “grupo experimental” y el “grupo de toque”, y que se observan en el presente (t). Este procedimiento considera las variables perturbadoras que, en caso de no vigilarse y señorearse, tienden a oscurecer la verdadera relación entre el tratamiento y los presuntos efectos, que es el problema mismo que tratamos de resolver mediante el método experimental.

Volvámonos a considerar, ahora, las dificultades prácticas que se encuentran en los esfuerzos para aplicar estas dos clases de diseño experimental en la investigación sociológica. Mi discusión, aquí, se basa en una experiencia de primera mano. Los detalles ocuparían más tiempo y espacio de aquellos de los que dispongo, por lo que me veré obligado a concentrar mi atención en algunos principios lógicos del método experimental en el contexto de la comunidad. La aplicación de estos principios es difícil. Aquí, nuevamente, tendré que ser severamente ecléctico. Para hacerlo, limitaré mi presentación a algunas de estas dificultades. No puedo presentarlas todas, hasta llegar a incluir, por ejemplo, las que se encuentran durante las entrevistas investigatorias de las áreas industrialmente subdesarrolladas (9, pp. 12-4). Discutiremos cinco subtópicos, en

cuanto ejemplos interrelacionados: 1. muestras aleatorias; 2. vigilancia y señorío de las observaciones; 3. métodos de análisis objetivo (*passim*); 4. diseño *ex post facto*, y 5. modo de usar los resultados de los diseños experimentales imperfectos (*passim*).

Cuando se menciona el método experimental, se piensa —generalmente— en el experimento de laboratorio, puesto que en el laboratorio la variable perturbadora puede someterse a vigilancia y dominarse o mantenerse constante. Recientemente, se ha usado este procedimiento para la investigación con grupos pequeños. Muchas grandes universidades tienen ahora laboratorios de ciencias sociales. Estos están equipados con grandes espejos, con medios o instrumentos de grabación, con cámaras que sincronizan los registros sonoro y de movimiento.

En pleno contraste con tales medios de vigilancia y dominio de las observaciones en el laboratorio, subsiste —en buena parte intacto— el problema de la aplicación del método experimental a la situación natural comunitaria, para medir los efectos del tratamiento o alguna otra influencia que parezca estar modificando los patrones psicológicos o las estructuras sociales. Durante muchos años, nos hemos contentado con apoyarnos en observaciones no vigiladas ni señoreadas de dichos fenómenos a pesar de que, a causa de ello, se sigue desperdiciando gran cantidad de dinero porque hemos llegado a *creer* que los medios de rehabilitación que hemos elegido logran *verdaderamente* los propósitos que se buscan con su aplicación. Así es como seguimos invirtiendo el dinero de los impuestos en alojamientos públicos, en programas para evitar la delincuencia juvenil, en ayudar a personas necesitadas de rehabilitación, sin verdadera garantía de efectividad.

Si se tienen en mente estas consideraciones, no puede extrañar que el propósito de este artículo consista en llamar la atención para analizar y describir la forma de lograr adaptaciones apropiadas del método experimental en la evaluación de tales programas. Los problemas implícitos en los experimentos comunitarios constituyen una difícil tarea; pero una que es digna de estudio, a causa de su importancia para la vigilancia y señorío de la vida grupal. Es, asimismo, un reto intelectual que no puede considerarse pequeño. Comenzamos con la proposición de que la lógica esencial del método experimental, aplicado en la situación comunitaria, consiste en haber observaciones sobre cambios o acontecimientos de la vida grupal, bajo condiciones de vigilancia y señorío de las variables sobresalientes que aparecen en el comportamiento humano *in situ*. Muchas de ellas son de carácter cuantitativo. Así, las diferencias de especie y no de grado se presentan bajo la forma de

dicotomías sexuales, condiciones matrimoniales y categorías nacionales, así como afiliaciones religiosas. Algunas de éstas pueden ordenarse jerárquicamente; así, por ejemplo, las categorías ocupaciones pueden colocarse en un orden dado por la inteligencia, los valores de prestigio, la posición social. Otras son realmente cuantitativas, como ocurre con el ingreso, la riqueza, la edad mental, etcétera. Hay otras aun, como las actitudes, las preferencias, los papeles, etcétera, que pueden convertirse en cuantitativas mediante procedimientos escalares o escalas sociométricas.

Siempre que las variables cualitativas se pueden colocar en orden jerárquico por razones que importan para el propósito elegido, o cuando se presentan variables cuantitativas reales, se pueden conseguir aproximaciones a la vigilancia y señorío efectivos de la variación mediante la igualación de las distribuciones de frecuencias de un grupo experimental y de uno de toque, o mediante pareación de sus parámetros (como la media y las desviaciones medias cuadráticas) o, en las muestras pequeñas, mediante un pareamiento de individuos idénticos o mediante un pareamiento de casos de acuerdo con varios rasgos medidos.

Debe de señalarse, en este punto, que la vigilancia y señorío de las variables se logran no mediante la manipulación de los sujetos humanos o a través de interferencias con su libertad de acción, sino manteniendo —simplemente— como constantes entre los dos grupos los valores de los rasgos medidos que se quieren vigilar y señorear. Entonces es posible observar cómo se comportan las variables no vigiladas y señoreadas, puesto que ahora no estarán contaminadas por las otras. Por ello afirmo que el problema medular en los intentos para adaptar el método experimental en la situación comunitaria consiste en realizar una observación bajo condiciones vigiladas y señoreadas. Cuando la vigilancia y señorío se logran por aleatorización esto es de lo mejor; pero, cuando una selección aleatoria no es posible, se puede recurrir aun al pareo para lograr la vigilancia y señorío necesarios, siendo éste un procedimiento alterno pero no un procedimiento sustitutivo.

Si bien el estudio del grupo de toque con base en el pareamiento no es en forma alguna un buen sustituto de la aleatorización en los estudios experimentales, lo cierto es que proporciona una alternativa que puede constituir el óptimo asequible bajo ciertas circunstancias, como cuando el prejuicio, la interferencia o el malentendimiento comunitarios están presentes.

Si hay fundamentos semánticos para objetar mi uso del estudio del grupo de toque en cuanto método experimental, no objetaré si no es para reiterar mi opinión de que la lógica real de la experimentación con-

siste en proporcionar un método mediante el que las observaciones se puedan hacer bajo condiciones vigiladas y señoreadas. El uso de la del grupo de toque permite —en efecto— que se mantengan constantes ciertas variables. Por razones como éstas, he usado la frase “diseños experimentales en la investigación sociológica” para designar estos procedimientos empíricos en el estudio de los grupos de vigilancia y señorío, reconociendo sus limitaciones en términos de requerimientos científicos estrictos. También me parece que el uso que he adoptado ayuda a pensar claramente acerca de los problemas complicados de las influencias multifactoriales que se ejercen sobre el comportamiento humano en la situación comunitaria. Porque a menos que concordemos en definiciones como ésta, estaremos a merced de obstáculos que impiden el pensamiento lógico, como los constituidos por las creencias ideológicas y con ello, bien podríamos hacer a un lado, por anticipado, todo esfuerzo para valorar los estudios sobre la sociedad humana.

Ordinariamente, si descamos generalizar resultados a partir de un grupo muestral, todos los sujetos deberemos elegirlos al azar. En algunos experimentos biológicos, en los que las distribuciones muestrales indican claramente la normalidad de la distribución del universo (como en el caso de muchos rasgos biológicos que se aproximan a la distribución probabilística normal, según ocurre en el caso de la estatura humana), la muestra se subdivide meramente (replicación) al azar, para el análisis de la variancia. Pero, generalizar a partir de tales submuestras, tan sólo porque los sujetos que se encuentran *dentro* de la muestra han sido subdivididos aleatoriamente, cuando ignoramos la distribución del rasgo en la situación comunitaria, me parece que proporciona una dudosa base para una firme generalización. Por otra parte, cuando sabemos que las distribuciones se aproximan a la normal, la aleatorización sólo dentro del experimento no puede comprometer la generalización científica a partir del grupo experimental que se extienda para abarcar la población o universo más amplios.

Los datos censales, basados en una enumeración completa de las poblaciones humanas, puede proporcionar información adecuada acerca de la distribución de rasgos tales como la edad, el sexo, la condición matrimonial, los agrupamientos de acuerdo con la nacionalidad, etcétera. Pero, muchos rasgos sociales importantes (tales como las actitudes, los papeles sociales, la participación grupal, algunas ocupaciones, los ahorros familiares, etcétera), tienen distribuciones que, en la mayoría de los casos, son desconocidas. Tomemos, por ejemplo, las actitudes; no conocemos la distribución de las actitudes públicas en comparación con

las actitudes privadas (reales) del mismo grupo de individuos humanos frente al mismo problema. Se han hecho unos cuantos estudios en esta materia, y éstos indican que las dos pueden diferir mucho entre sí. Ocupaciones aparentemente iguales, no son las mismas en diferentes culturas. Mientras que un rasgo físico como la estatura varía un tanto regularmente entre las razas, las actitudes de los individuos difieren entre las culturas, entre las clases sociales y entre las ocupaciones dentro de cada cultura.

Al lado de los pocos ejemplos de rasgos humanos que tienden a aproximarse a la distribución normal (como la estatura), hay otros rasgos (como el peso corporal) que son ligeramente asimétricos o se desvían ligeramente de la normal; otros (como las rentas urbanas) son más asimétricas, y aun hay otros rasgos (como las actitudes y la frecuencia en la participación grupal) que forman distribuciones jotiformes.

Si volvemos nuevamente a la aleatorización, debemos decir que hay, en general, dos clases de obstáculos para el estudio experimental; los hay filosóficos y los hay administrativos.

Muchos obstáculos filosóficos e ideológicos se producen porque se supone que las necesidades humanas son el primer criterio para el tratamiento que haya de darse con el fin de lograr un ajuste humano incrementado. Si se ha de ayudar, los más urgidos deben recibir consideración primaria. Las mores locales apoyan las expectativas humanas de que se proporcionará una ayuda de emergencia a los individuos más necesitados. Tanto la legislación social como las reglas administrativas incorporan procedimientos que parten de estas consideraciones y ponen en vigor su aplicación en la práctica.

El personal administrativo cargado con la responsabilidad de proceder de acuerdo con esto, mira de soslayo procedimientos tan impersonales y científicos como la selección aleatoria de los sujetos humanos que han de recibir el tratamiento que se necesite. Los administradores y funcionarios se muestran renuentes a elegir al azar a quienes han de recibir tratamiento, y la gente espera, naturalmente, un tratamiento personalizado que se brinde en primer término a los más necesitados. Los administradores estimulan entre sus subordinados las actitudes humanas que éstos deben mantener. De ahí que el esfuerzo cooperativo para participar en la investigación científica sobre sujetos humanos coloque en primer término el servicio y en segundo la investigación. En raras ocasiones se capta que la investigación que se planea adecuadamente puede diseñarse a modo de que no tenga que ir contra estos principios humanos de servicio. E intentare demostrarlo en seguida.

En niveles inferiores al administrativo —tales como el que corresponde al personal de vigilancia— hay otros obstáculos que se oponen al esfuerzo cooperativo eficaz con los estudiantes de investigación. Los vigilantes recibieron una educación profesional que les capacitó para el servicio, y ésta incluyó en raras ocasiones un entrenamiento adecuado en aspectos importantes de la investigación (como las aplicaciones de la teoría de las probabilidades o, incluso, de la teoría sociológica). El énfasis principal ha consistido en el entrenamiento para la diestra realización de servicios para los necesitados. En el nivel ínfimo de los trabajadores profesionales (cuyo enfoque puede ser enteramente clínico) en raras ocasiones se entienden realmente los propósitos y los métodos de la investigación. Incluso pueden llegar a considerar esos trabajadores que la investigación representa una interferencia en sus prácticas, o una desviación innecesaria hacia fines arbitrarios, mecánicos e incluso misteriosos. Esta situación señala la necesidad de explicar previamente los propósitos de la investigación. Esos practicantes tampoco distinguen ordinariamente entre los dos significados del término “experimento” —la búsqueda, mediante ensayo y error, de técnicas de servicio mejorado (como las ilustradas en muchas de las llamadas “demostraciones” de la presunta superioridad de cierta técnica nueva) y el experimento científico. Tales demostraciones contrastan claramente con los experimentos científicamente vigilados y señoreados que se diseñan específicamente para valorar la efectividad de las nuevas prácticas, o técnicas.

Debería resultar de plena evidencia el que se requieren salvaguardas para preservar la integridad de los procedimientos de investigación (la aleatorización, por ejemplo). Estas salvaguardas deberían incorporarse en el diseño original de investigación y continuamente se deberían hacer esfuerzos para lograr que fueran comprendidas por los practicantes en todos los niveles de la organización, a modo de que las mismas técnicas prácticas pudieran convertirse en objeto de estudio y valoración objetiva.

Permítaseme considerar algunas de estas salvaguardas. En primer lugar, incluso una *selección aleatoria inicial* de sujetos humanos puede ser estropeada o puede hasta quedar inutilizada por una pérdida de casos que no llegue a explicarse. De ahí que la investigación original deba de incluir orientaciones para registrar en detalle las circunstancias de cada caso perdido (sea que se deba a muerte, a enfermedad seria, a desplazamiento, o a una negativa para proporcionar la información pedida en la entrevista). En segundo lugar, debe de adaptarse y seguirse un plan procesal, para reemplazar cualquier caso perdido y para hacerlo pronta-

mente. Este plan capacitará al director de investigaciones para seleccionar al azar casos nuevos, semejantes a los que han quedado fuera. En tercer lugar, *las negativas* en cuanto a dar información no deben de aceptarse como finales sino hasta que han fallado varios intentos de entrevista. Si las mismas persisten y los casos se pierden, tales casos deben de reemplazarse como se describe después.

La selección aleatoria de casos aceptados para recibir tratamiento debe de hacerse *sin violar* los principios humanos de servicio. Así, por ejemplo, si la organización tiene mayor número de solicitantes de sus servicios del que sus recursos permite aceptar para tratamiento, parece que una selección aleatoria de los casos que se tratarán podría ser un procedimiento administrativo de admisión mucho *más justo* que el que dejara la elección sujeta a la idiosincracia personal de cada practicante. De este modo, si una organización tiene un personal de magnitud dada, y el número de personas necesitadas, elegibles para tratamiento, es doble del que la organización puede aceptar, se pueden asignar números aleatorios a todos los solicitantes y *echar suertes*. Aquellos a quienes correspondan los números "afortunados", marcados de antemano al azar para su aceptación, recibirán servicio; aquellos a quienes no correspondan los números "afortunados" tendrán que ser excluidos del tratamiento hasta que los recursos de la organización puedan duplicarse, o hasta que los casos cerrados produzcan nuevas vacantes (5, pp. 194-5).

Ese procedimiento produciría un grupo experimental aleatorizado para la recepción del tratamiento. De los casos excluidos se podría seleccionar más fácilmente un grupo de contraste en forma aleatoria. Los casos de emergencia —de necesidad aguda— tendrían que ser aceptados, por supuesto, para un pronto tratamiento, incluso *antes* de iniciar el procedimiento descrito. En cualquier forma como ésta se podrían superar muchos obstáculos, insuperables hasta ahora, de selección aleatoria, y se podría seguir un procedimiento experimental acordado con los métodos-modelo de prueba y análisis.

El *modelo* de análisis matemático apropiado (1) es alguna forma de ecuación de regresión múltiple para la que tienen que estimarse los parámetros. En la ecuación de regresión, la variable dependiente (miembro de la izquierda) mide el resultado esperado de los "presuntos efectos" de las variables de tratamiento que aparecen en su miembro derecho. La variable de tratamiento puede estar bajo la vigilancia del investigador y puede ser aplicada por él en grado apropiado al logro del resultado que se espera. Sin embargo, usualmente, en los estudios de la comunidad en que se investiga el programa social de tratamiento dado por una agen-

cia comunitaria, es esta agencia la que controla el grado de tratamiento, y el investigador es un *observador pasivo* de los procedimientos de servicio y de sus presuntos efectos.

Las llamadas variables independientes, para nuestros fines, se pueden agrupar en cuatro clases. *Primera.* La variable de tratamiento, cuya fuerza puede ser "total" o "nula", como en la dicotomía del estudio experimental, en que el grupo experimental recibe tratamiento y el grupo de toque no lo recibe. Como demostraremos, los dos grupos deben de mantenerse como *estrictamente independientes* entre sí (5, pp. 194-5) si se quieren obtener resultados claramente delineados. *Segunda.* Un grupo de variables que es extraño a nuestro propósito o que opera como variable perjudicial y que, por tanto, tiende a enmascarar el resultado esperado, y que, en general, puede subdividirse en tres subgrupos de variables externas: variables conocidas, medidas en la investigación previa (edad, sexo, ingreso, por ejemplo) y variables cuyas distribuciones pueden ser conocidas o desconocidas. *Tercera.* Variables que se encuentran frecuentemente y que aun cuando se sabe que existen, no han sido medidas en forma adecuada (los papeles y su desempeño son ejemplo de éstas). *Cuarta.* Un grupo que existe o que, al menos, seríamos sensatos en postular que existe, constituido por variables desconocidas. Éstas pueden ser variables de trasfondo, variables antecedentes, variables intervinientes, variables latentes, variables aleatorias, etcétera, o variables a las que podemos designar con los términos que queramos o con aquellos con los que hayan sido designadas provisionalmente. Como hemos visto, el método científico por el que podemos vigilar y señorear estas variables independientes, es la aleatorización.

Cuando los obstáculos a la aleatorización no pueden superarse la alternativa que queda es el *pareamiento* (23) de los individuos en los dos grupos que han de compararse. El pareamiento no es en forma alguna un sustituto de la aleatorización. Es tan sólo un procedimiento alternativo y no uno preferido. Nos permite vigilar y señorear las variables independientes de nuestro experimento. Usualmente no proporciona una base científica para la inferencia y la generalización. El uso de las pruebas probabilitarias para determinar la significación o importancia estadística de los resultados, en tales situaciones, es enteramente *analógico*. El pareamiento es un procedimiento empírico y conduce solamente a resultados empíricos. Para variables independientes que no se miden o son desconocidas, los casos de pareamiento basados en factores conocidos que buscan vigilancia y señoría no bastan por lo general. En este contexto, por supuesto, usamos la palabra "medida" para incluir las va-

riables cualitativas. Para las variables independientes que son conocidas y/o medidas, el grupo experimental y el grupo de toque se igualan en lo que se refiere a esas medidas, para obtener la vigilancia y dominio necesarios. Esto puede hacerse, como se hizo notar anteriormente, mediante las distribuciones de frecuencia de las medidas o sobre la base del valor de sus parámetros (medias, desviaciones medias, etcétera).

El pareo para el logro de la vigilancia y señorío necesarios, hará que se *pierdan* aquellos casos que no puedan parearse. Estos casos perdidos son, por lo general, los valores extremos de las medidas. Esto hace, a su vez, que se pierda información, y hace que disminuya el tamaño de los grupos experimental y de toque residuales. Por otra parte, incluso la pérdida de información acerca de los casos no pareables queda compensada en cierto grado por el conocimiento más íntimo y detallado de cada uno de los casos retenidos, que es adquirido por el investigador conforme maneja la información caso por caso en el proceso de pareamiento. En algunas ocasiones, la pérdida de casos puede compensarse mediante la adición de nuevos casos elegidos aleatoriamente, para que sirvan de substitutos de los casos perdidos. Supongamos que las pérdidas sean relativamente considerables ¿qué pasará entonces con la confiabilidad de las inferencias? En esto, hay una compensación que consiste en parear las muestras no aleatorias con base en una variable que se sepa que está *correlacionada* con la variable-efecto, *redefiniendo en seguida el universo* a modo de que quede constituido por una sub-población estratificada con base en este factor de pareamiento. Esta forma de proceder permite, en ocasiones, una generalización a partir de las muestras, hacia la sub-población, en términos de la dirección de cambio de la variable-efecto, *sin que prediga* en forma confiable, la magnitud del mismo. Este procedimiento puede ser una alternativa frente al pleno rechazo del experimento en su totalidad, y puede permitir que se hagan inferencias calificadas. Pero, nuevamente, estas inferencias calificadas son solamente generalizaciones empíricas y no científicas. En consecuencia, el uso analógico de las pruebas probabilitarias en este contexto, equivale a decir: "Si estas muestras hubieran sido aleatorias y de esta clase y tamaño, los valores podrían haber sido tales y cuales".

Una vigilancia y señorío de variables no medidas, en este contexto de pareamiento, pueden lograrse también, hasta cierto grado al menos, siempre que estudios previos hayan demostrado la existencia de una correlación entre los factores conocidos medidos y las variables no medidas del experimento. Así, por ejemplo, supongamos que tenemos pareado nuestro grupo experimental con un grupo de toque con base en la

distribución de clases ocupacionales; entonces variables no medidas tan importantes como las actitudes ocupacionales pueden vigilarse y señorearse parcialmente puesto que sabemos que los escalones superiores de la escala ocupacional poseen sistemas de actitudes que difieren de los sistemas de actitud de las clases ocupacionales inferiores. En forma parecida, la vigilancia y señoría mediante pareamiento de las clases de ingreso produce alguna vigilancia y señoría correlativos sobre el sistema de actitudes, pues los sistemas de actitudes están correlacionados con las diferentes clases de ingreso. Esto resulta de un conocimiento empírico.

En ocasiones se da el caso de que se encuentren *pequeñas diferencias* (5, pp. 37-8, 104-7, 209-10 y 22) entre las variables de un experimento. Esto es, frecuentemente, lo que ocurre cuando dos grupos muestrales se parean con base en una o más variables, y las varias medidas del efecto muestran ligeras diferencias entre los dos grupos. Aquí, necesitamos un medio de sumar estas pequeñas diferencias para descubrir si hay patrones omnicomprendivos y combinados que puedan ser importantes o significativos. Cuando tales resultados ocurren en muestras aleatorias, la fórmula de R. A. Fisher (12) sobre las probabilidades combinadas pueden producir un valor probabilitario omnicomprendivo que sea estadísticamente importante a pesar de que los valores-p separados no sean, por sí mismos, significativos desde el ángulo estadístico. Para las variables cualitativas que se muestran en las clases de las columnas y renglones de una tabla o cuadro de contingencia, el carácter aditivo del cuadro de χ^2 proporciona una prueba semejante. Por último, el álgebra moderna de las desigualdades proporciona frecuentemente, un procedimiento practicable de análisis (2; 3; 5, pp. 212-22, 16). Pero, en todos esos ejemplos, debemos estar seguros de cuál es la base sobre la que se levantan las inferencias a las que llegamos, sea que nuestro uso de la probabilidad esté basado, científicamente, en muestras aleatorias o que tengamos tan sólo una base puramente analógica para nuestras conclusiones.

Aun tras haber superado las dificultades para medir los cambios resultantes de alguna influencia natural o algún tratamiento, podemos encontrar que ha habido algún mejoramiento en los casos no expuestos a esas influencias o tratamiento. Entonces ¿cómo podremos separar y evaluar los presuntos efectos del tratamiento de por sí? El problema se puede explorar, y en ocasiones se puede resolver mediante métodos de vigilancia y señoría grupales. El argumento es como sigue: si la ganancia media de una medida del efecto, para el grupo de toque (o no tratado) es menor que la ganancia media de la misma variable en el grupo experimental,

debe de analizarse la diferencia en el cambio, para determinar —si es posible— si es una burda medida de tratamiento diferencial. En seguida enfrentamos el problema de explicar *por qué razón* mejoró también el grupo de toque. Aquí recurrimos a la inferencia hipotética y postulamos lo siguiente: en la situación comunitaria natural, existe, probablemente, un cierto número de fuerzas de recuperación. Éstas proceden, frecuentemente, de la participación en los grupos, de experiencias que tienden a construir y a integrar personalidades humanas sin el beneficio del tratamiento (5, pp. 90-3). Hay fuerzas naturales recuperativas en la participación grupal que satisface las *necesidades* básicas de los seres humanos (necesidad de obtener respuesta, necesidad de sentirse necesitado por alguien, necesidad de “pertener” a algún grupo, necesidad de recibir reconocimiento de parte de los demás, necesidad de seguridad). Estas necesidades, frecuentemente, se satisfacen sin programas organizados de tratamiento, puesto que grupos tales como la familia y los grupos amistosos tienden a satisfacerlas en el contexto de la comunidad natural. Es sólo cuando tales fuerzas dejan de operar en forma efectiva cuando se recurre al tratamiento organizado.

Es obvio que la solución de este importante problema de método —el consistente en separar los presuntos efectos del tratamiento de los efectos de estas influencias naturales de recuperación que postulamos— no podrá lograrse a menos que el tratamiento *se reduzca estrictamente* al grupo experimental. Si el tratamiento salpica a algunos de los miembros del grupo de toque, los dos *no son independientes* (5, pp. 195-6) y la agudeza del contraste esperado se difumará, con lo cual sacrificaremos la validez misma de la prueba que nos propusimos aplicar. Esta separación continua de los dos grupos durante el experimento, es *esencial* para cualquier prueba objetiva de que no sólo ha habido ganancia diferencial sino de que ésta ha sido debida al tratamiento. Cuando —conforme con mucha frecuencia— la información acerca de las técnicas de tratamiento, así como cierto entrenamiento o ciertos vislumbres del tratamiento que se ha de someter a prueba se deslizan de los entrevistadores al grupo de toque, pueden perderse todas las ventajas de la prueba experimental. Estas fugas del tratamiento, pueden producirse cuando el administrador de una organización cooperativa de servicio, urgido de introducir desde luego en la práctica las nuevas técnicas supuestamente superiores, monta lo que llama una “demostración” de las ventajas del nuevo tratamiento, y lo hace *en el curso* de la prueba experimental del tratamiento. Dentro de un buen procedimiento de investigación, esto es lógicamente inadecuado, puesto que presupone, desde el principio, la superioridad del nue-

vo tratamiento y ese es, precisamente, el punto que la investigación estaba destinada a probar originalmente. De este modo, se comete el error lógico de la "petición de principio" y todo el programa de investigación queda invalidado antes de concluirse, con lo que se vuelve inútil.

A pesar de que los estudios de los grupos de toque que no se basa firmemente en una selección aleatoria de casos sufren severas limitaciones cuando se les considera desde un ángulo estrictamente científico, pueden señalarse ciertos puntos específicos que hacen que tales estudios puedan ser útiles. Éstos tienen una cierta utilidad provisional, que deriva del hecho de que proponen estimaciones; en cuanto tales, pueden sugerir hipótesis fructíferas para que se realicen experimentos más estrictamente científicos. Asimismo, cuando un estudio experimental se repite en grupos similares de sujetos, bajo condiciones parecidas, los resultados pueden ser corroborativos. En tal caso se fortalece nuestra confianza en la inferencia provisional. Esto es también lo que ocurre cuando los resultados obtenidos confirman las expectativas que pueden derivarse de una teoría sociológica cuidadosamente formulada, o son consistentes con ella.

Como indicamos anteriormente, en este mismo artículo, el diseño *ex post facto* comienza con la observación de alguna diferencia presente en un momento t , e intenta remontarse hasta una fecha anterior ($t-1$) a través de la serie de acontecimientos sucesivos que pueden explicar las diferencias observadas en el presente (t). Como indiqué en 1955, podemos obtener una muestra probabilitaria de casos en un momento dado (t), pero no podemos aleatorizar el grupo anterior del que la muestra, en un momento elegido (t), es un residuo; residuo que ha quedado tras una pérdida de casos por muerte y movilidad entre el momento t y el $t-1$. ¿Cómo se puede obtener, entonces, una función muestral probabilitaria para estimar la parte que los acontecimientos azarosos han desempeñado en el intervalo cubierto por este proceso. Una aproximación a la resolución de este problema, en los estudios *ex post facto* se puede lograr mediante la adaptación de conceptos como los usados en los procesos estocásticos y en las cadenas de Markov. Una breve afirmación hecha por Stouffer (14) resume la materia de este modo: "los procesos estocásticos son cambios sistemáticos en la probabilidad de un acontecimiento, a través del tiempo. Una persona o una sociedad, en el momento t_1 , es una función probabilitaria de su comportamiento en el momento t_1-1 y aun en periodos previos. Si en un momento ignoramos todos los periodos anteriores, excepto el inmediato precedente, y suponemos simplemente que hay alguna secuela estructural tal que el comportamiento en

t, sea una función probabilitaria del estado de los asuntos y del comportamiento en $t-1$, tendremos lo que se conoce como un proceso de Markov, un modelo de creciente importancia en la psicología y en las ciencias sociales”.

Bush y Mosteller (4) dan por supuesto que tenemos un conjunto de alternativas A_j que han de identificarse como clases de respuestas, y un conjunto de valores probabilitarios P_j que responden a esas alternativas. En el caso más simple, hay dos alternativas A_1 y A_2 con probabilidades p y $1-p$ de que ocurran respectivamente. Estas probabilidades corresponden a la realización de los acontecimientos. Estos acontecimientos, en el sistema matemático, deben identificarse con varias clases de acontecimientos del mundo real, que influyen en el comportamiento del organismo. Estos acontecimientos, en el diseño *ex post facto* no están vigilados y señoreados, ordinariamente, por el experimentador, excepto si es posible el proceso de pareamiento. Entre los factores a que esto concierne, hay diferentes grados de dependencia, en el sentido de la cantidad de información que los miembros previos de la secuela proporcionan con respecto a los miembros ulteriores. Mann y Whitney (19), usando pequeñas muestras de hasta 8 casos, y Marshall (20), quien trabajó con muestras más amplias, desarrollaron procedimientos no paramétricos para probar las hipótesis de que una de las variables aleatorias es estocásticamente mayor que la otra. Todas estas técnicas parecerían adaptables a un estudio *ex post facto*, cuidadosamente diseñado, y la introducción de funciones de probabilidad proporciona al investigador instrumentos de análisis más poderosos que los que se han usado hasta ahora.

Para cerrar este artículo, llamaré nuevamente la atención del lector hacia el hecho de que muchos aspectos relevantes del método experimental no pueden discutirse aquí a causa de la limitación de espacio. Algunos de esos aspectos adicionales se presentan en las referencias de pie de página y en las fuentes que a continuación citamos.

REFERENCIAS

Para la Bibliografía, véase la obra de E. Greenwood, citada

(1) Arrow, K. J., “Mathematical Models in the Social Sciences”, capítulo VIII, pp. 129-54, en *The Policy Sciences: Recent Developments in Scope and Method*. Stanford University, California, 1951.

(2) Berkeley, E. C., “The Algebra of States and Events” *Scientific Monthly*, 78: 232-41, abril, 1954.

- (3) Birkhoff, G. & MacLane, S., *A Survey of Modern Algebra*, New York, MacMillan, 1953.
- (4) Bush, R. R. & Mosteller, F., *Stochastic Models for Learning*, Wiley, New York, 1955.
- (5) Chapin, F. Stuart, *Experimental Designs in Sociological Research*. Harpers, New York, edición revisada, 1955.
- (6) Cochran y Cox describen 150 diseños experimentales del tipo que nosotros hemos designado con la expresión "diseños proyectados". Véase Cochran, W. B. & Cox, Gertrude M., *Experimental Designs*. Wiley, New York, 1950.
- (7) Comte, Auguste, *Cours de Philosophie Positive*, Paris, 1864, II, 13-4.
- (8) Dixon, W. H. & Massey, F. J. Jr., *Introduction to Statistical Analysis*, McGraw-Hill, New York, 1951. Capítulo 9. "La Menia: Estimación y Pruebas de Hipótesis", pp. 94-118; y capítulo 17, "Estadística No-Paramétrica", pp. 247-63.
- (9) Dodd, S. C., *A Controlled Experiment in Rural Hygiene in Syria*, American University Series, N° 7, Beirut, 1934.
- (10) Fisher, R. A. & Mackenzie, W. A., "Studies in Crop Variation, II. The Manual Respose of Different Potato Varieties", *Journal of Agricultural Science*, 13:311-20, 1923.
- (11) Fisher, R. A., *Design of Experiment*. Hafner, New York, varias ediciones desde 1935. Véase también Edwards, A. L., "Experiments: Their Planning and Execution", capítulo 7, pp. 259-88, en G. Lindzey's, *Handbook of Social Psychology*. Cambridge, 1954. Capítulo 8, del volumen 1, por Bush & Mosteller pp. 289-334.
- (12) Fisher, R. A., *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh, 5ª edición, 1934, p. 98.
- (13) Friedman, Milton, "The Use of Ranks to Avoid the Assumption of Normality". *Journal of American Statistical Association*. 32:200:675-701, diciembre de 1937. También Chapin, *op. cit.*, pp. 202-22.
- (14) Gittler, J. B., *Review of Sociology: Analysis of Decade. 1945-1955*. Wiley, New York, 1957, capítulo 2, "Métodos Cualitativos", por S. A. Stouffer.
- (15) Greenwood, Ernest, *Experimental Sociology*. King's Crown, New York, 1945.
- (16) Haberman, S., "Distributions of Kendall's Tau based on Partially Ordered Systems", *Biometrika*, 42, partes 3 y 4, diciembre de 1955, pp. 419-24.
- (17) Johnson, P. O., *Statistical Methods in Research*, Prentice Hall, New York, 1949, capítulo v, "Procedimientos ordinarios de Prueba de las Hipótesis Estadísticas" pp. 60-103. También el artículo de Johnson sobre "Modern Statistical Science and its Function in Educational and Psychological Research" *The Scientific Monthly*, 62:6:385-96, junio de 1951.
- (18) König, René, *Beobachtung und Experiment in der Sozialforschung: Praktische Sozialforschung*, II. Koln, 1956, pp. 221-58.
- (19) Mann, H. F. & Whitney, D. R., "On a Test of Whether One of two Random Variables is Stochastically Larger than the Other". *Annals of Mathematical Statistics*, 18, 1947, 50-60.
- (20) Marshall, A. W., "A Large Sample Test of the Hypothesis that One of two Random Variables is Stochastically Larger than the Other", *Journal of the American Statistical Association*, 46:255:366-74, septiembre de 1951.
- (21) Una crítica fue proporcionada por Simey, J. S., en "Social Investigation: Past Achievements and Present Difficulties", *British Journal of Sociology*, VIII, 2:121-9, junio de 1957.

(22) Stouffer, S. A., "Quantitative Methods", capítulos 2, en Gittler's *Review of Sociology*. Op. cit., p. 45. De las pequeñas diferencias, dice Stouffer, "...hay un uso creciente de las estadísticas de orden simple... Si se tienen diez cuadros que muestran diferencias en la misma dirección, la probabilidad de obtener tal resultado con una hipótesis cero de falta de diferencia, es tan pequeña (1 en 1024) que no pueden invocarse pruebas elaboradas. La consistencia entre numerosas submuestras, encontrada mediante clasificaciones cruzadas, se usa en forma creciente como criterio de significación o importancia estadística".

(23) Thorndike ha señalado que cuando las variables cuantitativas se usan en el pareamiento es deseable considerar la regresión respecto a la media y hacer los ajustes apropiados; véase Thorndike, R. L., "Regression Fallacies in Matched Group Experiments", *Psychometrika*, 7:85-102, 1942. En los estudios empíricos de pareamiento citados por Chapin, *op. cit.*, en que se usan frecuentemente variables cualitativas, que no se distribuyen normalmente (ocupación, sexo, raza, etcétera), tales ajustes estadísticos refinados no pueden hacerse ordinariamente.