ALGUNOS PROBLEMAS DE LA INVESTIGACION EMPIRICA

Enrique Carpena*

I. Investigación inducción y muestreo

En muchas concepciones científicas, aún hoy, subsiste la antigua idea griega de que el criterio de demarcación de lo científico consiste en la exigencia de la verdad para aquellas proposiciones que aspiran a integrar el campo de la ciencia. Esto significa, equivale a afirmar, la necesidad de verificación completa de los enunciados científicos.

Una proposición, de acuerdo al principio de verificabilidad completa, es verdadera, si, y sólo si, ha sido comprobada para todos los casos particulares a los que se refiere el cuantificador que incluye. Un criterio tan riguroso deja fuera del campo científico, todas aquellas proposiciones que incluyen un cuantificador universal del tipo "todos", y se refieren a conjuntos de fenómenos infinitos o casi infinitos. En el caso de la inducción, la aplicación de este criterio exige la inducción completa, y tanto en un contexto deductivo como inductivo, casi todas las leyes universales de la ciencia moderna, no pertenecerían al conocimiento científico de aplicarse una norma de cientificidad tan rigurosa. De la contradicción entre la práctica científica real y un criterio estrecho, surgen conceptos como el de "verificación incompleta", "inducción incompleta", y contemporáneamente, en las escuelas ligadas a lo que sin mucho rigor podemos denominar como neopositivismo o positivismo lógico, surge el criterio, aún más tolerante, de "verificación en principio". Lo cual implica, aceptar en el campo de la ciencia, proposiciones con fundamento teórico, de las que se puede indicar en "principio", eventuales procedimientos de verificación, aunque la verificación misma sea totalmente imposible en el momento de su formulación.

Tal liberalidad, en la actualidad, es práctica común en la comunidad

* Universidad Autónoma de Puebla.

científica, aunque subsiste la vieja discusión: la provisoriedad del conocimiento, su fragilidad manifestada en las formulaciones probabilísticas ¿es un problema ontológico? ¿Tiene que ver con la "verdadera" "esencia" naturaleza del ser? ¿O es un problema epistemológico derivado de nuestro deficiente sistema actual de describir y explicar el Universo, universo que se comporta en forma determinada y necesaria? La ciencia, para los modernos, ha dejado de ser aquel conjunto de proposiciones "verdaderas", para transformarse en un sistema de "hipótesis" probablemente verdaderas.

Si la descripción probabilística es una forma aproximada y deficiente, o es una forma que recoge el comportamiento real de los fenómenos, constituye, entonces, parte de una problemática filosófica que hoy admite múltiples respuestas. Respuestas fundamentadas en concepciones tan profundas y últimas, como aquellas que las diferentes filosofías tienen sobre el "ser", la "realidad", el "conocimiento", etc. El desarrollo de la explicación probabilística se basa en forma inmediata, en la llamada "inducción incompleta" y la inducción en las concepciones básicas del empirismo. En particular, del empirismo inglés, de Bacon en adelante, salvo excepciones con Hume, para quien la universalización inductiva no se justifica. Incluso en algunos modernos (Popper), la inducción no existe, al no existir un principio inductivo válido, capaz de legitimar este tipo de inferencia.

A pesar de un panorama poco claro, de problemas sin resolver, de extensas lagunas, los modelos probabilísticos de inferencia estadística, basados sobre la información de muestras aleatorias, se han desarrollado en forma rápida durante este siglo y su aplicación se ha extendido a todos los campos del conocimiento científico.

El ideal científico del siglo pasado estaba expresado por el modelo de las ciencias naturales que postulaban un conocimiento exacto universal.

Las ciencias sociales se planteaban como meta esa "naturalización" de sus teorías científicas como forma de superar sus descripciones tendenciales, algo inciertas y en general imprecisas. Paradójicamente el desarrollo científico, parece haber roto con estos ideales y las ciencias naturales han tendido a acercarse a una concepción tendencial e hipotética de los fenómenos, concepción más propia de las Ciencias Sociales que de los modelos newtonianos.

La problemática en torno al muestreo, ha sido desarrollada históricamente en un contexto inductivo, en general relacionada con los supuestos de la probabilidad, y en particular, vinculada con aquellas pruebas estadísticas que suponen independencia entre los sucesos, y que por lo tanto, exigen muestras al azar con reposición.

En este sentido, las conclusiones teóricas basadas en procedimientos muestrales, comparten todas las limitaciones del razonamiento inductivo, en particular, aquellos referidos a las dificultades de la fundamentación sintética —más allá de la lógica— del principio de inducción.

Cuando se utiliza una muestra para generalizar sus resultados al universo, se pone en práctica el camino más frecuente utilizado en la Sociología empírica en particular, y en general, en todas las corrientes que no comulgan con el empirismo radical, que exigiría la "verificación completa"; en otras palabras, el trabajo con censos.

Sin embargo, la pregunta que se le hace al empirismo, desde hace siglos, sigue vigente: ¿acaso la selección misma de cierta información no implica acercarse al objeto con una hipótesis? Si no fuera así, sería imposible encontrar un camino (seleccionar en la maraña de datos), y si es así, el comienzo parece ser deductivo.

Pero esa hipótesis inicial ¿cómo se obtuvo? La respuesta empirista sería: a partir de la experiencia.

La disputa entre el racionalismo y el empirismo, entre idealismo y materialismo, siempre se encuentra presente en estos problemas técnicos.

Así J. Galtung un sociólogo partidario del empirismo moderno, hace una distinción entre aquellos muestreos que permiten la inducción y aquellos que no permiten la generalización de los resultados al universo.

Para Galtung los enunciados científicos pueden dejar de ser "hipótesis" mediante una muestra probabilística. Al respecto dice: "En otros términos una teoría sustantiva puede ser puesta a prueba en tales muestras, es posible generalizar los descubrimientos; ellos mantienen su calidad de hipótesis mientras no se haya obtenido una muestra probabilística".[1]

La utilización de un esquema inductivo es una constante en los trabajos empíricos actuales en Sociología, y es por ello, que existe una marcada tendencia a la utilización de generalizaciones estadísticas, partiendo de datos muestrales. El utilizar la muestra con fines de generalizar los resultados obtenidos, puede responder a razones pragmáticas de construcción de hipótesis de relativamente bajo nivel de abstracción. En este caso, nos hallamos en lo que Popper denomina el "contexto de descubrimiento".

Pero si bien la teoría del muestreo ha sido desarrollada en un contexto inductivo, también cumple una función esencial en un contexto hipotético-deductivo, en el momento de la constrastación de las proposiciones básicas deductivas de los otros niveles de hipótesis. Estas pueden referirse a un universo de objetos, infinito o finito, pero no abarcable en su totalidad por razones económicas de tiempo, costo o acceso. Por lo tanto, será necesario seleccionar un subconjunto del universo a fin de contrastar las hipótesis sobre ese universo.

Este tipo de técnica es apta en aquellos contextos epistemológicos, donde no se exige la transformación de las hipótesis generales en un conjunto finito de proposiciones observacionales, tal como proponen algunos deductivistas al estilo de Moritz Schlick.[2]

II. Necesidad de la técnica del muestreo

Una muestra se puede obtener, tanto a partir de un universo infinito, como de un universo finito. En general, es un problema teórico-empírico específico, determinar cuál es el alcance de las proposiciones.

En el caso de universos infinitos, la necesidad de trabajar con muestras resulta obvia. Cuando el universo es finito, la muestra se obtiene por razo-

1 Galtung, J. *Metodología y técnicas de la investigación social.* Eudeba, Bs. As., 1966. Tomos 1, 2, 3.

2 Schlick, M. *El viraje de la filosofía,* en: A.J. Ayer (comp.). *El positivismo lógico,* Ed. FCE, México, 1965.

nes puramente económicas: el ideal sería trabajar con censos. Debe notarse que al trabajar con un censo, las hipótesis pueden ser verificadas en forma completa, ya que desaparece el problema de la justificación de la inferencia inductiva.

Los datos de la muestra —sea de un universo finito o infinito— pueden ser utilizados para la falsación de hipótesis sobre esos universos, ya que los datos reales, serán considerados a su vez, como muestras aleatorias de una población dada.

III. Representatividad de la muestra

La idea aproximada y más difundida de representatividad, consiste en imaginar a la muestra como una maqueta o como un mapa en escala del universo. Es decir, que para que una muestra sea representativa, todas las unidades deben estar representadas en la misma proporción que en el universo.

De todas las muestras posibles de obtener, a partir de una determinada población, hay una, y sólo una combinación que posea la misma función de densidad que el universo, y esta combinación es la única que lo representa exáctamente. Por lo tanto, en general todas las muestras pueden ser consideradas como desviaciones de la representatividad. La diferencia entre buena muestra y una mala muestra estará dada por el grado en que ésta se desvía de aquélla exáctamente representativa. La probabilidad de obtener la muestra, que posea la misma función de densidad que la población, es prácticamente cero, ya que esta probabilidad está dada por la razón entre "uno" (muestra con igual función de densidad del universo) y todas las muestras posibles a partir del universo, que resultan de la combinatoria $(\frac{m}{n})$, donde "m" es el tamaño del universo y "n" es el tamaño de la muestra.

De lo dicho se deduce que discutir sobre la representatividad de una muestra particular es discutir un pseudoproblema, salvo en el caso que se conozca el valor de los parámetros poblacionales y se pueda, de esta forma, realizar una comparación con sus estimadores muestrales. Pero aún así, el que una muestra represente adecuadamente ciertas variables conocidas del diseño, no garantiza su representatividad en torno a las demás variables.

Sí se puede garantizar en una muestra aleatoria, la probabilidad que ésta tiene de no alejarse de los parámetros poblacionales, más allá de cierto error. Aunque no se puede garantizar, que una muestra en particular, no se aleje de dichos parámetros sobrepasando los límites de error establecidos.

A partir de una muestra nunca se podrán hacer estimaciones tales como: "La media poblacional de ingreso es de $ 3,000.00", pero sí se pueden hacer afirmaciones del siguiente tipo: "Con un 90% de probabilidad, la media poblacional de ingreso se encuentra entre $ 2,800.00 y $ 3,200.00".

Esto significa, que de cada cien muestras que tomen del mismo universo, en noventa de ellas, no se cometerá un error de estimación mayor de $ 400.00. Pero también, significa que de cada cien muestras, en diez se cometerá un error mayor, y que sobre una muestra particular es imposible decidir, si se encuentra entre las noventa "correctas" o las diez "erróneas".

Al trabajar con un nivel de confianza tan alto como 90%, es bastante imposible que una muestra particular se aleje más allá del error establecido; sin embargo, existe una técnica sencilla que disminuye aún más las posibilidades de utilizar una muestra muy alejada de los límites deseados.

Esta técnica, consiste en dividir la muestra original en varias submuestras independientes, con el objeto de comparar los datos que éstos estiman. Si no hay variaciones importantes, habremos disminuído en forma significativa los riesgos implicados, como podrá verse, aplicando la regla de multiplicación de las probabilidades.

Como ejemplo práctico, podríamos mencionar, que si es necesaria una muestra de 1,000 casos, es preferible tomar cuatro muestras independientes de 250 casos, y calcular medias en cada una de ellas para las mismas variables. Si estas medias no se alejan mucho entre sí, se puede tratar las cuatro muestras como si fueran una sola; pero si alguna se aleja, será necesario excluirla o eliminar toda la muestra.

Esta prueba se puede realizar también *a posteriori*, analizando al azar, grupos de 250 casos cada uno, tarea que se puede encomendar directamente al computador.

IV. El tamaño de la muestra

Este, quizás, sea uno de los puntos que más preocupaciones concita en el momento del diseño de la investigación.

El tamaño de la muestra dependerá de cuatro órdenes de factores, uno de ellos el económico, quedará fuera de este análisis. Si desde el punto de vista presupuestal, es imposible cubrir las exigencias mínimas del diseño, será menester cambiarla o pensar en otro tipo de fuente de información.

Las otras variables que afectan el tamaño muestral son:

1) Probabilidad de error o nivel de confianza de la muestra
2) Amplitud máxima tolerable del error
3) Tipo de análisis que se proyecta realizar

La determinación de los valores de las tres variables que afectan el tamaño de la muestra, dependen de consideraciones teóricas y metodológicas ajenas al problema estadístico. De la índole del trabajo a realizar, dependerá si se trabaja con niveles altos o bajos de confiabilidad y error.

Si por ejemplo, el estudio se refiere al grado de toxicidad de determinado medicamento, probablemente interese trabajar con mayor confiabilidad y menor error, que si se intenta determinar qué cualidades se esperan de un nuevo automóvil por parte de los consumidores para organizar la campaña publicitaria.

También se percibe intuitivamente, que al aumentar la confiabilidad y/o disminuir el error deseado aumentará el tamaño de la muestra y por lo tanto su costo.

a) Requerimientos estadísticos y tamaño de la muestra

Una vez decidido el error máximo que se acepta cometer y el nivel de confianza al que se trabajará, la determinación del tamaño muestral es sólo un problema de cálculo: en efecto, sabemos por experiencia que de una población dada se pueden obtener infinitas muestras, y que de acuerdo al Teorema Central del Límite, las medias muestrales tienden a distribuirse de acuerdo a una curva normal, cuando su número tiende a infinito.

Pero cuando la varianza de la población es una incógnita, y sólo disponemos de estimaciones muestrales, las medias muestrales tienden a distribuirse como la función "t" de Student:

$$t = \frac{\overline{x} - u}{s / \sqrt{N}}$$

donde

$\overline{x}$ = media muestral

u = media poblacional

s = desviación standar

N = tamaño muestral

t = valor de t que indica el nivel de confianza

Depejando N, tamaño de la muestra tenemos:

$$N = \frac{s^2 \cdot t^2}{(\overline{x} - u)^2}$$

s^2 = varianza, es el cuadrado de la desviación standar

Dentro de la ecuación de N hay dos incógnitas que es necesario despejar. El denominador "$\overline{x} - u$" representa el error, o sea la diferencia entre la media muestral y la poblacional. Su valor se fija obviamente en porcentaje de error a asumir, por ejemplo: 5% ó 10%; los valores de $\overline{x}$ y s, se pueden estimar a través de un pretest de pocos casos.

El valor de t se obtiene de la tabla de standarización de valores de t, así por ejemplo, para un nivel de confianza de 90%, el valor de t es igual a 1.64, para un nivel de confianza de 80%, será: 1.28 en pruebas de dos colas.

Si observamos la ecuación que define el valor del tamaño muestral, veremos que N depende proporcionalmente de la desviación standar y del nivel de confianza. A medida que aumenta t y/o s, tenderá a aumentar N, lo contrario sucede con el cuadrado del error que es inversamente proporcional a N, o sea, a medida que aumenta ($\overline{x} - u$) –suponiendo constante el numerador– disminuye N.

Debe notarse que a pesar de la creencia muy difundida acerca de que el tamaño de la muestra depende del tamaño de la población, es claro que el tamaño muestral depende de la varianza, como dato poblacional, y del tamaño de la población, por la "Ley de los grandes números", indirectamente.

Así por ejemplo, en una población de cien mil persona, donde se tuviese un ingreso único de $ 3,000.00 la varianza sería cero, y una muestra de un solo caso, sería suficiente como estimación. En cambio en una población de mil personas con veinte tipos diferentes de ingreso, la varianza sería mayor y N también mayor que en el caso anterior, a pesar de que la población es menor, y la variable a estimar, la misma.

Cuando se trabaja con varias variables relevantes a la vez, se debe calcular el tamaño muestral por cada variable –ya que cada una tendría una varianza diferente–, y tomar el N mayor.

También es necesario tener en cuenta, que la disminución del error y el aumento del nivel de confianza no es directamente proporcional al tamaño de la muestra sino a su raíz cuadrada, por cuanto:

$$\bar{x} - u = \frac{s \;.\; t}{\sqrt{N}} \qquad \text{y} \qquad t = \frac{(\bar{x} - u)\ \sqrt{N}}{s}$$

El siguiente ejemplo ilustra el cálculo del tamaño muestral N. Suponemos una población de un millón de habitantes sobre la cual se desea establecer su ingreso medio.

A fin de estimar la media y la varianza se realiza un pretest de 20 casos donde se establece que $\bar{x} = 3.000$ y $s = 1.000$.

Si se trabaja con un error máximo del 3%, de modo que $\bar{x} - u = 90$, y un nivel de confianza del 90%, debe buscarse en la tabla de distribución de la t de Student el valor correspondiente al 90% y que es 1,64 para la prueba de dos colas. Aplicando la ecuación anterior donde:

$$N = \frac{s^2 \;.\; t^2}{(\bar{x} - u)^2} \qquad \text{y reeemplazando tenemos}$$

$$N = \frac{1.000^2 \;.\; 1{,}64^2}{90^2} = 283 \text{ casos}$$

Este resultado indica que el tamaño de la muestra para un 905 de certeza y 3% de error, es de 283 casos. El procedimiento debe observarse para todas las variables del estudio y asumir el N mayor.

b) Análisis de correlación y tamaño de $\bar{x}$ la muestra

Una vez establecido el tamaño muestral en función de los parámetros de confiabilidad señalados anteriormente, el investigador debe preguntarse si la cantidad de casos calculados será suficiente para la etapa de procesamiento de datos. Sólo se puede dar respuesta a este interrogante sabiendo de antemano, la cantidad de variables y categorías de variable que se analizarán simultáneamente.

Trabajaremos con ejemplos para dejar claro este punto, de suma importancia.

Si el análisis que se proyecta es univariado, las variables tienen un máximo de once categorías, y requerimos un promedio de diez casos por categoría (cifra arbitraria), el total de casos requeridos sería de 110.

Supongamos que se analiza la variable educación y sólo interesa conocer la distribución de la población por grado de instrucción formal que se distribuye así:

1) Sin educación
2) Primaria incompleta
3) Primaria completa
4) Secundaria incompleta
5) Secundaria completa
6) Preparatoria incompleta
7) Preparatoria completa
8) Universitaria incompleta
9) Universitaria completa
10) Posgrado incompleto
11) Posgrado completo

Si se quiere relacionar la educación con otra variable como ingreso para detectar cómo influye en el grado de educación alcanzado, se obtendrá un cuadro del siguiente tipo:

Educación / Ingreso	Bajo	Medio	Alto
Sin educación			
Primaria incompleta			
Primaria completa			
Secundaria incompleta			
Secundaria completa			
Preparatoria incompleta			
Preparatoria completa			
Universitaria incompleta			
Universitaria completa			
Posgrado incompleto			
Posgrado completo			

En este cuadro, las necesidades de casos se triplican, el número de celdas a llenar será de 11 x 3 = 33, y la cantidad de casos 330, con lo cual se superará el cálculo realizado en función de la confiabilidad, que es el ejemplo del punto anterior, y que arrojó una muestra de 283.

Al introducir una tercera variable, por ejemplo, el sexo del entrevistado, tendremos el cuadro anterior duplicado del siguiente modo:

HOMBRES				MUJERES			
Educa-ción / Ingreso	Bajo	Medio	Alto	Educa-ción / Ingreso	Bajo	Medio	Alto
Sin educación				Sin educación			
Primaria incompleta				Primaria incompleta			
Primaria completa				Primaria completa			
Secundaria incompleta				Secundaria incompleta			
Secundaria completa				Secundaria completa			
Preparatoria incompleta				Preparatoria incompleta			
Preparatoria completa				Preparatoria completa			
Universitaria incompleta				Universitaria incompleta			
Universitaria completa				Universitaria completa			
Posgrado incompleto				Posgrado incompleto			
Posgrado completo				Posgrado completo			

Con lo cual el requerimiento de casos habrá ascendido a 660, superando así ampliamente los requerimientos de confiabilidad, que deberán ser recalcados en función de este nuevo dato. De nada nos serviría una muestra pequeña, si los requerimientos del análisis son mayores.

Generalizando este ejemplo, la expresión matemática que indica el tamaño de la muestra en este nivel es:

$N = C^v \cdot F$ donde: N = tamaño muestral

C = categoría de la variable

v = número de variable a correlacionar simultáneamente.

F_0 = frecuencia media deseada para cada celda.

Los datos que se utilizarán para calcular N, corresponderán al cuadro más complejo; y el N para la investigación, será el más alto calculado por cualquiera de los procedimientos señalados.

V. Los tipos de muestras

a) Muestras aleatorias y no aleatorias

La diferencia entre muestras aleatorias y no aleatorias, se refiere al modo en que éstas son obtenidas, y no a las muestras mismas.

El rasgo fundamental de una muestra aleatoria, es que todas las unidades tengan igual probabilidad de ser seleccionadas, o en su defecto, tengan probabilidades conocidas para poder corregirlas a probabilidades iguales a través de algún sistema de ponderación.

Si la probabilidad se desconoce, no será posible utilizar la estadística inferencial, uno de cuyos supuestos básicos es la independencia entre los sucesos que conforman la muestra. En otras palabras, por independencia se entiende que la selección de un suceso cualesquiera no afecte la probabilidad de selección de los demás de selección.

La mayor parte de las pruebas estadísticas (test de significación, intervalos de confianza, etc.) suponen el muestreo aleatorio. Es el único tipo de muestra que permite hacer aseveraciones tales como: "La probabilidad de que el ingreso poblacional oscile entre 10,000 y 12,000 pesos es del 95%".

b) Muestra simple al azar

Es el tipo más sencillo de muestra aleatoria a partir del cual, y por introducción de normas adicionales, se construyen los otros tipos de muestras aleatorias.

La muestra simple al azar debe cumplir dos requisitos:

1o. Cada una de las unidades debe poseer la misma probabilidad de ser seleccionada.
2o. Cada combinación posible de sucesos debe poseer la misma probabilidad de ser seleccionada.

Con respecto al primer requisito, no es general para todas las muestras, ya que aunque las probabilidades no sean iguales, de ser conocidas, será posible ajustar el modelo mediante algún mecanismo de ponderación. Para cumplir con el requisito de igualdad de la probabilidad en la selección, se puede apelar a diversos expedientes, a saber:

1o. Trabajar con reposición. De esta manera si bien logramos independencia estadística, nos arriesgamos a seleccionar dos veces la misma unidad, riesgo que disminuye con el aumento del tamaño del universo.
2o. Adoptar el supuesto de que el universo es infinito. Este supuesto no es muy "fuerte" cuando: a) se trabaja con universos grandes, y dado que: b) el nivel rudimentario de algunas técnicas en ciencias sociales lleva a una acumulación considerable de errores (piénsese en el grado de imprecisión que implica la medición de una variable a través de una entrevista).
3o. No hacer el supuesto de infinitud del universo, ni utilizar reposición. En este caso, después de cada extracción, la probabilidad de las demás unidades aumenta; pero si la probabilidad de todas las unidades restantes se mantiene igual, independientemente de las unidades seleccionadas en las extracciones anteriores, tenemos independencia entre una extracción y otra; salvo en el hecho de que ningún individuo podrá ser seleccionado más de una vez, con lo cual se viola el supuesto de independencia, aunque tal violación no es grave, sí se tienen en cuenta las consideraciones que se hicieran con respecto al supuesto de infinitud.

Si la muestra fuera muy grande en relación con el universo (30%) se pueden aplicar factores estadísticos de corrección.

c) Procedimientos de selección en muestra al azar

La técnica más frecuente para obtener una muestra al azar, es:

1o. Realizar un listado de unidades del universo
2o. Numerar cada una de las unidades del listado
3o. Realizar la selección por cualquier procedimiento mecánico. En general lo más rápido y apropiado, es la utilización de una tabla de números al azar. El punto más difícil de este tipo de muestra, reside en obtener un listado del universo; incluso muchas veces se debe redefinir el universo para adecuarlo a la lista que se posee o puede obtenerse.

d) Muestra sistemática

A partir del listado del universo, y tomando un número de lista seleccionado al azar, se selecciona cada unidad. El n en general es arbitrario pero conviene fijarlo mediante la razón entre el tamaño del universo y el tamaño muestral para asegurarnos el recorrido total del listado, y variar cada tanto, el número de partida y la razón, para evitar el sesgo.

Si el listado ha sido al azar, este tipo de muestra es similar a la muestra aleatoria simple, pero posee una serie de ventajas prácticas, por ejemplo, la de no tener que utilizar mecanismos de azar para seleccionar las unidades.

e) Muestra estratificada proporcional

Este tipo de muestra es posible sólo cuando exista alguna variable que permita estratificar el Universo (sexo, edad, educación, clase social, etc.) de tal manera que cada individuo caiga en un solo estrato.

Posteriormente se obtiene una muestra sistemática o al azar de cada estrato proporcional al tamaño del mismo.

f) Muestra estratificada no proporcional

En este caso se varía la proporción muestral respecto al universo. Este procedimiento es particularmente útil, cuando existen limitaciones para aumentar el n muestral ya que permite reducir el número de estratos no relevantes y aumentar el n en aquellos estratos que poseen una proporción muy baja en el universo, pero que son muy importantes desde el punto de vista teórico. Convendrá, en general, disminuir en la muestra la porporción de casos de aquellos estratos donde la varianza disminuye. Por ejemplo: en un estudio de Comportamiento Político se comprobó que la varianza de respuestas de amas de casa era casi nula, representando el 45% del universo. Por lo tanto, su representación en la muestra se disminuyó al 10%.

Al medir "clase social" la clase alta, tal como había sido definida, abarcaba el 2% de la población y en una muestra de 500 casos habría probablemente unos 10 casos en esa categoría. Por lo tanto, se aumentó su proporción en la muestra en un 20%. Con este tipo de manipulaciones en ningún caso se afecta la premisa básica del muestreo aleatorrio: las probabilidades siguen siendo conocidas.

g) Muestreo por Cluster

Este tipo de muestra sumamente útil en el trabajo sociológico, consiste en estratificar el Universo de acuerdo a un criterio puramente geográfico en zonas lo más pequeñas y heterogéneas posibles. Luego se selecciona al azar una serie de estos estratos, denominados Clauster se realiza el tipo de muestra que se desee, para obtener las unidades de estudio. Este tipo de muestra se puede realizar en tantos escalones como sea necesario.

La muestra así diseñada, permite una sustancial reducción de la dispersión y, por tanto, una reducción considerable de costos.

h) Muestras no aleatorias

No existen técnicas estadísticas desarrolladas para este tipo de muestras, lo cual impide una precisión adecuada de los datos y la estimación del error que se está cometiendo, por lo que no es aconsejable su utilización ni para test de hipótesis, ni para generalizaciones. Es útil su utilización en el caso de estudios exploratorios y de pretest, en los cuales no se desea dejar al azar la selección de las unidades, tanto porque no existen criterios para estratificar adecuadamente el universo, como porque se desea asegurar el máximo de heterogeneidad con un N muy pequeño.

BIBLIOGRAFIA

Ayer, A.J. *El positivismo lógico.* Ed. FCE, 1965.
Blalock, H.M. *Estadística Social.* Ed. FCE, México, 1965. Cap. X.
Braitwhaite. *La explicación científica.* Ed. Tecnos, Madrid, 1962. Caps. V, VI, VII.
VI, VII.
Cochran, W.G. *Sampling Methods.* Ed. J. Wiley, N.Y., 1953.
Galtung, J. *Metodología y técnicas de la investigación social.* Ed. Audeba, Bs. As., 1966. Tomo I: 2.3
Hyman, H. *Surbey Design and Analysis.* Ed. J. Wiley, N.Y., 1955.
Kerlinger, F. *Investigación de la conducta.* Ed. Galache, México, 1975. Caps. 7 y 8.
Kish, L. *Muestreo de encuestas.* Ed. Trillas, 1975.
Mc Carthy, P.J. *Introduction to Statistical Reasoning.* Ed. Mc Graw Hill, N.Y., 1957. Cap. X.
Popper, K. *La lógica de la explicación científica.* Ed. Tecnos, Madrid, 1962. Cap. X.
Sukhatme, P.V. *Teoría de encuestas por muestreo.* Ed. FCE, México, 1956. Caps. I, II, III, IX.

CUADRO D. *Distribución de* t

df	Nivel de significación para la prueba de una sola cola					
	.10	.05	.025	.01	.005	.0005
	Nivel de significación para la prueba de dos colas					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.985
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

FUENTE: El cuadro D es una abreviación del cuadro III de *Statistical Tables for Biological, Agricultural and Medican Research* (ed. 1948), de R.A. Fischer y F. Yates, publicada por Oliver & Boyd, Ltd., Edimurgo y Londres, con la autorización de los autores y editores.

crítica jurídica
Revista Latinoamericana de Política, Filosofía y Derecho

contenido 2

ANALISIS