



Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Sociología

Documento de Cátedra 90

Tassi, P.: “De lo exhaustivo a lo parcial: un poco de historia sobre el desarrollo de los sondeos”

Tassi, P. (1988). De l'exhaustif au partiel: un peu d'histoire sur le développement des sondages. *Statistique et analyse des données*, 13 (2), 113-129

Traducción resumida del francés al español realizada por Martín Güelman (2013)

Cátedra de Metodología y Técnicas de la Investigación Social
Profesora Titular: Ruth Sautu

Este Documento de Cátedra forma parte de una serie que tiene como propósito contribuir a la formación de alumnos de la Carrera de Sociología de la Universidad de Buenos Aires en cuestiones vinculadas al diseño y realización de investigaciones científicas en el campo de las ciencias sociales. Su contenido complementa los textos de metodología de lectura obligatoria y optativa incluidos en cada uno de los tres cursos que se dictan en la Carrera.

Estos documentos son material de uso interno y no pueden ser incorporados a ediciones impresas ni reproducidos comercialmente.

La Cátedra solicita a los usuarios de estos Documentos que citen a sus autores indicando las referencias completas, es decir: autores, fecha, título, número y tipo de documento (traducción, resumen, elaboración propia, etc.). En los casos en que el usuario utilice sólo parte del documento, haciendo referencia a algunos de los autores/obras originales allí incluidos, consignar que fue tomado de nuestro Documento de Cátedra. Por ejemplo:

E. O. Wrigth (1985), *Classes*, London: Verso, citado en Documento de Cátedra II.1., Plotno, G., Lederman, F. & Krause, M. (2007) “Escalas Ocupacionales”.

Los orígenes

Generalmente se admite que la estadística abarca un conjunto de métodos que contribuyen al conocimiento de una población, tomando este último vocablo en sentido amplio, tanto si su uso proviene de la demografía como del recuento censal de individuos.

La aparición del término *estadística* es en sí relativamente reciente, sobre todo en comparación con la actividad correspondiente de recolección de datos que se remonta a la antigüedad.

En general, atribuimos la creación del término *estadística* a la escuela descriptiva alemana de Gottingue y, en particular, a Gottfried Achenwall (1746). El término proviene del latín *statisticum*: que tiene relación con el Estado.

Sin embargo, es más probable que Achenwall solo haya contribuido a imponer el empleo del vocablo puesto que hacia 1701 ya contábamos con una *Biblioteca Statistica* y en 1672 con un *Microscopium Statisticum*. Remontándonos aún más en el tiempo, el término *estadística* parece haber pertenecido al lenguaje administrativo francés del colbertismo.¹

Por el contrario, la aparición de la necesidad “estadística” de contar con datos numéricos precede a su denominación en varios milenios. En su origen, la necesidad provenía de los jefes de Estado deseosos de conocer los elementos bajo su poder: población, potencial militar, riquezas, etc. ¿Qué mejor conocimiento que el procedente de la observación sistemática y exhaustiva de todos los elementos de la sociedad? La idea de censo o de inventario aparece, entonces, de manera natural en la historia implicando además una impresión de precisión de la más alta calidad.

Los primeros censos parecen remontarse a la civilización sumeria de 5000 a 2000 años antes de nuestra era. El relevamiento de personas y bienes tuvo lugar regularmente en la Mesopotamia hacia el 3000 a.C. Egipto parece haber sido la primer nación en organizar sistemáticamente censos de población al menos desde el año 2900 a.C., así como en institucionalizar los censos fiscales (entre el 2700 y el 2500 a.C).

¹ Nota del traductor: Con el término *colbertismo* se designa a la doctrina económica elaborada por Jean-Baptiste Colbert en el siglo XVII.

China desempeña también un importante rol en la historia de los censos. Por ejemplo, el emperador Yao edita, hacia el año 2238 a.C, el relevamiento total de individuos y producciones agrícolas.

Los primeros intentos de extrapolación

A partir del siglo XIII de nuestra era, los datos devienen más numerosos en virtud de la proliferación de las categorías fiscales. Durante el siglo XIV comienzan a registrarse las actas de estado civil [*actes d' état civil*].

Pese a que durante todo este período se consolida la estadística administrativa, los censos resultan poco frecuentes.

La estadística retomará sus progresos fundamentales durante la segunda mitad del siglo XVII a partir de la necesidad que volvían a sentir los monarcas y sus consejeros de conocer y explicar los fenómenos económicos y sociales. Esta época coincide con el desarrollo de la escuela de aritmética política inglesa, precursora de la estadística inferencial. Esta escuela, cuyos fundadores fueron John Graunt (1620-1674) y William Petty (1623-1687), se guiaba por la preocupación de la cuantificación y la búsqueda de constantes de comportamiento que permitieran realizar estimaciones y previsiones: número de hijos por mujer, tiempo entre dos nacimientos para una misma madre, número de habitantes por vivienda y hogar, proporción de fallecimientos, etc.

Las técnicas del *multiplicador* de la aritmética política serán utilizadas en detrimento de los censos, favoreciendo la aparición de encuestas parciales.

Paralelamente, el reemplazo de un conocimiento exhaustivo por una extrapolación fundada en el examen de una parte de la población será una actitud que comenzará a encontrar elementos de justificación con la aparición de los primeros resultados de probabilidades.

En Inglaterra, los métodos de la aritmética política prosiguieron con sus progresos. Las técnicas de extrapolación o *multiplicador* conocieron rápidamente un enorme suceso que contribuyó a poner un freno [*mettre un frein*] a los censos, juzgados por ciertos sectores como onerosos, poco precisos y “monstruosos” (Saint- Simon).

A comienzos del siglo XIX, parecían estar dadas las condiciones para el desarrollo de la metodología de los relevamientos parciales: la práctica,

esencialmente demográfica, había surgido; el cálculo de las probabilidades seguía su progresión y la voluntad de los Estados de poseer información numérica se confirmaba.

No obstante, deberán pasar más de 100 años para que el interés por los sondeos fuera reconocido y, más aún, para establecer una teoría matemática fundada en la aleatoriedad.

Un tiempo de cese: el siglo XIX

Contrariamente a lo esperado, el siglo XIX estuvo marcado por el retorno forzado de los censos, poco frecuentes durante el siglo precedente. En 1801 se realizan en simultáneo censos de población en Inglaterra, Dinamarca, Francia y Noruega. En muchos países los censos devienen regulares y periódicos.

Pese a que la teoría estadística conoció un desarrollo sin precedentes, el muestreo retrocedió frente a la exhaustividad.

Uno de los fundadores de la ciencia estadística tal como la conocemos en la actualidad es el belga Adolphe Quetelet (1796-1874). Quetelet es, ciertamente, el primero en concebir con gran claridad que la estadística podía estar fundada sobre el cálculo de probabilidades. Quetelet amplió el campo de aplicación de la estadística, de la economía y las ciencias sociales a la antropometría. Asimismo, organizó en 1853 en Bruselas el primer Congreso Internacional de Estadística.

Sin embargo, Quetelet lejos de preconizar los sondeos, se encontraba movilizado por la obsesión del relevamiento total; su preocupación constante era asegurar la exhaustividad de los datos recolectados. Aún así, aceptaba extender al conjunto de la población datos correspondientes a ciertos subgrupos (por ejemplo, la tasa de analfabetismo calculada entre los delincuentes).

Pese a la preponderancia de la enumeración total, las visiones parciales no desaparecieron por completo del bagaje de los estadígrafos.

El polo principal de utilización de los sondeos se sitúa en Estados Unidos con la emergencia de las encuestas de opinión pública, de la que hablaremos más adelante.

1895-1925: el debate sobre la representatividad

El concepto de representatividad de un resultado o muestra es muy reciente. Si bien Condorcet había hecho referencias alusivas repetidas veces, parece haber sido explícitamente utilizado por primera vez por Jean-Baptiste Say. Say afirma la imposibilidad de considerar representativa para el conjunto de una población una tasa de mortalidad calculada para sectores de altos ingresos, extrapolación que, como vimos, defendía Quetelet.

El debate sobre la representatividad tendrá como marco al Instituto Internacional de Estadística (IIS) creado en Londres en 1885.

El punto de partida del debate lo constituye la histórica exposición de Anders Nicolai Kiaer, director de la Oficina Central de Estadística del Reino de Noruega, en el congreso del IIS en 1895 en Berna, Suiza.

La idea de “una parte sobre el todo” es lanzada [*lancée*] a un público de estadígrafos. Es necesario remarcar que, a priori, ningún puente se había tendido entre las probabilidades y el muestro representativo.

Durante un congreso en Budapest en 1901, Kiaer expresa:

(...) el método representativo demanda un gran número de unidades de observación repartidas de modo tal que, en la medida de lo posible, los diferentes atributos se encuentren representados como lo están en el conjunto de la población.

Ladislau von Bortkiewicz, profesor en la Universidad de Berlín, sugiere recurrir al cálculo de probabilidades para probar estadísticamente la diferencia entre las distribuciones de la muestra y de la población total en las variables claves.

(...) se trata de determinar si la diferencia observada puede o no ser considerada un producto del azar, de causas fortuitas o accidentales.

Durante el mismo período, otras contribuciones fundamentales fueron aportadas por el inglés Arthur Bowley (1869-1957), quien desarrolló el muestro aleatorio y la estratificación y estableció, entre otras cosas, la ecuación de análisis de la varianza en universos estratificados.

El tercer gran nombre del período es el de Alexander Ivanovich Tchuprow quien definió la afijación óptima en un muestreo estratificado, lo que entra en

contradicción con la idea corriente de representatividad según la cual la muestra es un “modelo reducido” [*modèle réduit*] de la población entera.

Luego de 1925 el punto de discusión no será ya si hacer el muestreo o no hacerlo, sino cómo extraer la muestra. Rápidamente, el debate involucrará a los partidarios del muestreo aleatorio [*random sampling*²] y del muestreo estratégico [*purposive selection*³].

Adolphe Jensen, jefe del departamento de estadística de Dinamarca, era consciente de ello:

(...) cuando el IIS discutía el tema veinte años atrás, la cuestión que generaba mayor interés tenía que ver con el reconocimiento del método. Hoy en día el debate es otro. Me atrevo a decir que en la actualidad difícilmente exista un estadígrafo que impugne la legitimidad del método representativo (...)

La introducción del modelo probabilístico se encuentra en curso. La etapa ulterior se iniciará en 1934.

No obstante, desde hacía varios lustros se venía desarrollando, en Estados Unidos, un dominio de utilización del sondeo: el conocimiento de la opinión pública.

Una evolución paralela: los orígenes de los sondeos de opinión en Estados Unidos

Los sondeos de opinión nacieron del deseo de obtener información, en la medida de lo posible cuantitativa, sobre el estado de ánimo [*état d' esprit*] de la población.

Los sondeos de opinión surgen en los Estados Unidos en ocasión de la cobertura de prensa de las elecciones presidenciales.

Desde 1824 los periódicos *Harrisbourg Pennsylvanian* y *Raleigh Star* venían realizando encuestas pre-electorales a través de la consulta individual a los electores. Seguidamente, otros periódicos como el *New York Herald*, el *Chicago American*, el *Columbus Dispatch* y, en particular, desde 1916, el *Literary Digest* retomarán este hábito apreciado entre el público.

² Nota del traductor: En inglés en el original.

³ Idem.

Las muestras recogidas por los periódicos no reposaban sobre ningún criterio de representatividad (el término recién aparecería 70 años más tarde). Sus propiedades residían esencialmente en el elevado número de unidades muestrales: 30 mil para el *New York Herald* en 1905, más de 2 millones para el *Literary Digest* en 1936.

Una fecha crucial en la historia del muestreo es el 3 de noviembre de 1936, día de publicación de los resultados de la elección presidencial en los Estados Unidos. Mientras que el *Literary Digest* había pronosticado la victoria de Alfred Landon, el ganador fue Franklin Delano Roosevelt. El triunfo de Roosevelt había sido predicho por tres “sondeos” independientes realizados por Archibald Crossley, Elmo Roper y George Gallup. Este último creó su propio instituto de opinión pública en 1935. Los tres estudios fueron realizados utilizando el método representativo de selección estratégica [*méthode représentative par choix judicieux*], seleccionando por cuotas a las unidades muestrales.

Pese a la amplitud de la muestra, el sondeo del *Literary Digest* se reveló sesgado.

A partir de 1937, numerosas publicaciones norteamericanas contaron con una sección de opinión, nutrida de encuestas con muestras representativas realizadas por organismos privados.

En 1938 aparecen los primeros institutos de estudios de opinión en Gran Bretaña y Francia. Poco después, numerosos países van a continuar con este movimiento hasta que, en 1947, tiene lugar en París el primer Congreso Internacional de Opinión Pública.

La situación en la Rusia pre-revolucionaria

En el siglo XIX, al mismo tiempo que surgían y se generalizaban en Estados Unidos los sondeos de opinión, el Imperio Ruso conocía un desarrollo considerable en los métodos fundados sobre una visión parcial de la población. En 1864 el poder central crea los *zemstvo*, gobiernos locales en el seno de los cuales existían departamentos estadísticos encargados de recoger, en particular, datos sobre el estado de la agricultura para facilitar la definición de las orientaciones de la política económica rural.

El padre espiritual de la “estadística zemstvo” fue Tchuprow (1842-1908), profesor de Economía Política y de Estadística en la Universidad de Moscú y miembro del IIS desde su fundación.

La existencia de institutos estadísticos en los *zemstvos* favoreció la aparición de numerosos estadígrafos matemáticos que contribuyeron fuertemente al desarrollo de los métodos.

Las técnicas de muestreo parecen haber sido empleadas tempranamente por los estadígrafos rusos en virtud de la imposibilidad de observar exhaustivamente los territorios de su competencia. En 1894, Tchuprow brinda en Moscú una conferencia sobre el “método monográfico”, el cual consiste en la selección de elementos típicos determinados a partir de un conocimiento exhaustivo de una población.

Se admite generalmente que los sondeos, en el sentido amplio del término, eran utilizados comúnmente en la estadística rusa independientemente de Kiaer.

Un avance metodológico esencial surge en 1910, en paralelo a los trabajos de Bowley en Inglaterra. Tchuprow estudia el muestreo aleatorio y desarrolla su utilización para medir la precisión de los estadísticos calculados. Asimismo, hace referencia al muestreo por conglomerados y al muestreo estratificado, con y sin reposición.

Las dificultades vinculadas con la distancia, el idioma y, lógicamente, los eventos ligados con la Revolución de Octubre, impidieron durante mucho tiempo apreciar la importancia de la escuela rusa en la “estadística parcial⁴”. El desarrollo de la metodología y de la utilización de los sondeos en la Rusia zarista fue considerable. Numerosos estadígrafos rusos de comienzos del siglo XX pueden ser considerados como los precursores e iniciadores del método, en la misma medida que Kaier, Bowley y Neyman.

Los inicios del desarrollo contemporáneo

Luego de 1925 la discusión sobre el método representativo ya no giraba en torno a sus premisas, sino sobre sus modalidades de aplicación.

⁴ Nota del traductor: Por *estadística parcial*, el autor refiere a aquellos estudios que no relevan a todas las unidades de una población sino que se basan en algún tipo de muestreo.

De igual modo que Bowley y Tchuprow, Jerzy Neyman (1894-1981) puede ser considerado uno de los fundadores de la teoría estadística de los sondeos.

El aporte de Neyman será considerable ya que, en primer lugar y a diferencia de Bowley que los situaba en un mismo plano, tomará partido por el muestreo aleatorio en desmedro del muestreo intencional. Neyman muestra que las dos técnicas de muestreo no son equivalentes y que conviene utilizar el muestreo aleatorio dado que la selección deliberada de las unidades en el muestreo intencional reduce la confianza general.

Neyman consideraba representativas aquellas metodologías de sondeo que producían estadísticos convergentes con los parámetros poblacionales y permitían calcular la precisión de los resultados.

Más allá de esta cuestión fundamental de elección entre métodos de muestreo, Neyman contribuyó con numerosos desarrollos ya vislumbrados por Bowley: el muestreo aleatorio simple con y sin reposición y el muestreo estratificado. Asimismo, contribuyó con el desarrollo del muestreo por conglomerados, hizo avanzar significativamente la teoría de la estimación por sondeos y estudió la afijación óptima de un muestreo por estratos⁵. Estableciendo la superioridad de las técnicas de sondeo aleatorio en universos estratificados, Neyman orientó una gran parte de la investigación en esta dirección.

A partir de ese momento, las bases de la estadística de los sondeos se encuentran sentadas. Surgen dos grandes vías de desarrollo: una ligada a los problemas de la práctica (forma de extraer la muestra, estimación, varianza, no-respuestas, etc.) y otra referente al modelo probabilístico subyacente a la teoría de los sondeos. Los avances de la primera vía permitieron la construcción de la metodología fundamental de los sondeos.

Los desarrollos actuales

Desde mediados de los años cincuenta, las investigaciones metodológicas se fundan crecientemente en las relaciones existentes entre los sondeos y el modelo de la estadística inferencial.

⁵ Nota del traductor: La afijación es la distribución del tamaño muestral global entre los estratos diferenciados. La afijación óptima es aquella que no solo considera el peso relativo del estrato en la población, sino también su variabilidad o heterogeneidad respecto de la variable considerada en la estratificación. A aquellos estratos de mayor heterogeneidad y peso poblacional les corresponderá un tamaño muestral superior.

Un primer eje de investigación concierne a la aplicación en los sondeos de conceptos estadísticos introducidos a partir de la teoría de la estimación de Ronald Fisher, tales como la admisibilidad, la exhaustividad y la verosimilitud.

A modo de conclusión, vale resaltar que la lectura de las grandes obras o artículos que marcaron la historia de la estadística a fines del siglo XIX y principios del XX mostró que generalmente fueron omitidas en las investigaciones las condiciones de obtención de datos. Ello resulta perfectamente comprensible en virtud de las principales etapas de desarrollo de los sondeos.

Sin embargo, los procedimientos estadísticos fueron utilizados habitualmente como si las propiedades de la muestra fueran semejantes a las de la población. En otros términos, se procedió siempre como si la extracción de todas las muestras brindara la misma probabilidad de estimar los parámetros poblacionales.

¿No sería necesario modificar los métodos habituales para tener en cuenta el procedimiento aleatorio de recolección de datos y la complejidad creciente de los planes de sondeo?

Recientemente, han comenzado a surgir reflexiones sobre la validez de los procedimientos estadísticos cuando los datos provienen de sondeos complejos. Esta tendencia a reexaminar la validez de las herramientas de la estadística matemática en el modo de obtención de los datos, que promete futuros desarrollos para la ciencia estadística en su conjunto, debe acentuarse rápidamente.