

Estadística, orden natural y orden social

Juan Javier Sánchez Carrión

Universidad Complutense de Madrid. Facultad de Ciencias Políticas y Sociología
soso403@sis.ucm.es

Resumen

El artículo hace un recorrido por las diferentes técnicas estadísticas de uso más generalizado en la investigación social (recuento, media, desviación típica, distribución normal, regresión y muestreo), desarrolladas en un período que va del siglo XVII, cuando aparecen las obras de Graunt y de Petty, padres de la aritmética política y precursores de los actuales recuentos de la población, hasta finales del siglo XIX, principios del XX, cuando se desarrollan las técnicas que permiten estudiar las relaciones entre las variables (correlación y regresión), a cargo de Galton y de Pearson. Entre ambos momentos se encuentra la figura de Quetelet, a la que se le presta una especial dedicación, en tanto que precursor, con su «hombre medio», de la modelización estadística, y del uso generalizado y estandarizado de los censos. Además de tener un fin descriptivo, el estudio de los autores mencionados tiene por objeto mostrar las circunstancias, tanto de naturaleza heurística como sociopolítica, que se tuvieron que dar para que sus obras vieran la luz, al tiempo que se muestra el papel que tiene la estadística a la hora de naturalizar el orden social, como paso previo a su aceptación como orden político.

Palabras clave: sociología de la estadística, historia de la estadística, Graunt, Petty, Quetelet, Galton, Pearson.

Abstract. *Statistics, natural and social order*

This article talks about the different statistical techniques (counting, mean, standard deviation, normal distribution, regression), that have been developed from the XVII century, with Graunt and Petty, founders of Political Arithmetic and predecessors of present ways of counting the population, thru the end of the XIX century, when Galton and Pearson developed the models that allows to study relationships between variables (correlation and regression). Between both moments we can find Quetelet, someone very relevant as pioneer of statistical modelling, with his «average man», and father of periodical and standardized census. The article not only has a descriptive purpose, but also it pretends to show the circumstances, both of heuristic and soci-political nature, that made possible the invention of the different statistics, at the same time that I show its role in the naturalization of social order, as a previous condition for its acceptance as political order.

Key words: sociology of statistics, history of statistics, Graunt, Petty, Quetelet, Galton, Pearson.

Sumario

- | | |
|---|-----------------------------------|
| 1. Introducción | 4. La generalización: el muestreo |
| 2. Requisitos de la innovación | Bibliografía |
| 3. Estadística de recuento
y de modelización | |

Este trabajo parte del momento en que a partir del siglo XVII se desarrolla y se acepta la idea de que la naturaleza es inteligible y que está ordenada según ciertas leyes, que se pueden descubrir mediante el recurso al lenguaje matemático. El éxito de esta idea hizo que terminara exportándose al estudio de lo social, también supuestamente sometido a este mismo tipo de leyes susceptibles de ser expresadas numéricamente. Estas leyes naturales sustentarían la existencia de otras leyes, políticas, legitimadas por su adecuación a las primeras.

A lo largo de las páginas siguientes se ofrece un esquema de lo que yo entiendo que es el papel que la estadística cumple en este proceso de naturalización del orden social, al tiempo que se da cuenta de las circunstancias que se han tenido que dar para que esta labor naturalizadora haya sido posible. Más allá del contenido concreto de investigaciones particulares, mediante las que haciendo uso de distintas estadísticas se pueden estudiar temas tan diversos como, por ejemplo, el paro, la valoración de los líderes políticos o la relación entre el tabaco y el cáncer, está el hecho de que todas ellas tienen como común denominador el estar produciendo (intentando producir) hechos sociales, de carácter semejante a los que se acostumbra a llamar «naturales».

En este artículo se hace una revisión de las técnicas estadísticas que suelen aparecer en un libro introductorio de análisis estadístico de los datos, presentando a las personas que las desarrollaron o las aplicaron a lo social¹. Al mismo tiempo, se muestran las condiciones que las hicieron posibles y de qué manera contribuyen al proceso de naturalización de lo social. Cada una de estas técnicas (codificación y recuento, media, desviación típica, distribución normal, regresión, correlación y muestreo) las desarrollo en otros trabajos monográficos, unos acabados y otros en proceso de elaboración, en los que se tratan con más detalle de lo que se hace en este artículo.

1. Introducción

Producto de las lecturas e investigaciones llevadas a cabo y de las reflexiones que he ido haciendo a lo largo de los años que llevo como profesor de estadística y de técnicas de investigación social, he llegado a la conclusión de que la

1. Este artículo forma parte de un trabajo más amplio con el que trato de «ver» la parte hundida (invisible) del «iceberg» técnicas estadísticas de investigación, como complemento a la exposición que hago en mi *Manual de Análisis Estadístico de los Datos* (Alianza Editorial Manuales, 1999) de la parte emergida (visible) de dicho iceberg.

estadística es una herramienta cognitiva que cumple un papel básico en el proceso de presentar como natural un tipo de orden social, como paso previo a su aceptación como orden político. Todos los días escuchamos en los medios de comunicación que la tasa de paro es del, por ejemplo, 20%, que la valoración de los líderes políticos es de tantos puntos de media o que el tabaco y el cáncer están relacionados, tal como demuestran los estudios estadísticos. Todas estas conclusiones se escuchan no como opiniones o valoraciones sobre los temas a los que se refieren, sino como hechos verídicos, incuestionables —salvo que se sospeche que ha habido manipulación en la generación de estos resultados—. Si las conclusiones sobre el paro, la valoración de los líderes o la incidencia del tabaco en el cáncer no fueran numéricas, obtenidas mediante técnicas como, por ejemplo, la encuesta estadística, legitimada socialmente, perderían el atractivo que tienen, justamente atribuible al hecho de estar expresadas en números. Se diría que estos resultados, a los que podríamos catalogar como hechos sociales, son de la misma naturaleza que la existencia de la constante gravitatoria o que el Sol sale todas las mañanas a una hora que se puede fijar por el conocimiento que tenemos del movimiento de los planetas. Sin embargo, el que a nosotros nos parezca natural y que, además, aceptemos que el Instituto Nacional de Estadística publique tasas de paro o que otros institutos o investigadores estudien y publiquen sus conclusiones sobre casi cualquier tema imaginable, expresadas en medias o correlaciones, no significa que tales hechos (tasas, medias, correlaciones) no hayan tenido un principio en la historia, antes del cual hubieran sido, no sólo rechazados, sino siquiera imaginados. Justamente en este trabajo intento mostrar quiénes fueron las personas que introdujeron las principales técnicas estadísticas que hoy utilizamos masivamente y algunas de las circunstancias que se tuvieron que dar para que sus inventos fueran, primero, pensables, luego pensados y, finalmente, aceptados socialmente.

Escindido el mundo en naturaleza y sociedad, como requisito para poder estudiar la parte natural científicamente, y visto el éxito conseguido por dicho tipo de estudio, parece lógico pensar que hubiera gentes (entre otras las personas a las que se hace referencia en este artículo) que trataran de aplicar el nuevo método (la ciencia) al estudio de la sociedad, de manera que las conclusiones a las que se llegase al estudiarla tuvieran la misma naturaleza científica que las conclusiones elaboradas previamente sobre el orden natural. Dichas conclusiones sobre la sociedad, por su carácter de científicas, es decir supuestamente independientes de los valores de los investigadores y fiel reproducción de la realidad (el orden divino con el que el Creador hizo este mundo), tendrían que ser aceptadas por todo el mundo, como se aceptaban las conclusiones generadas por la ciencia sobre la naturaleza. Y, lo que es más importante, la aceptación de los resultados obtenidos haciendo uso del método científico se haría extensible al orden político que dijera supeditarse a los mismos —igual que hasta entonces había sido aceptado el orden político sustentado sobre la aceptación de la voluntad divina, expresada a través de las Sagradas Escrituras. Éste es el proyecto que se le presenta por delante a la estadística y del que son partícipes todos los pensadores que suelen aparecer entre los padres fundado-

res de esta disciplina. En todos los casos, empezando por las personas a las que yo voy a considerar como precursoras de esta nueva forma de mirar al mundo, Graunt y Petty, y siguiendo por Quetelet, Galton y Pearson, que son los restantes autores a los que dedico mi atención, los estadísticos manifiestan claramente su intención de aplicar el método científico, que estará más o menos desarrollado según la época de la que se trate, al estudio de la sociedad, primero como hilo conductor con el que adentrarse en el baconiano laberinto que es el mundo (Graunt y Petty) y después como método con el que descubrir las newtonianas leyes numéricas en las que dicho mundo está escrito (Quetelet, Galton y Pearson).

En las páginas que siguen, después de mostrar cuáles son los requisitos que se tienen que dar para que se produzca una invención, iré explicando de qué manera contribuye la estadística al proceso de naturalización de lo social del que vengo hablando, al tiempo que se muestra cómo se ha ido naturalizando su propio uso. Y para entender la naturaleza de los logros alcanzados por la estadística, podemos formularnos una serie de preguntas, relacionadas con las diferentes técnicas que la componen, que contempladas desde nuestra situación actual puede parecer que carezcan de sentido, justamente gracias al éxito que ha tenido la estadística en hacer natural algo que en su momento no tenía esta característica.

1. La primera pregunta se puede formular de la siguiente manera: ¿qué ha tenido que pasar para que hoy en día parezca completamente natural el que se hagan todo tipo de recuentos y registros, al tiempo que se aceptan como hechos verídicos sus resultados, cuando siglos atrás esta actividad contadora y registradora de la población tenía un carácter excepcional, sin que el producto (numérico) del recuento tuviese mayor peso que un buen razonamiento?
2. ¿Qué circunstancias se han tenido que dar para que la gente acepte ser resumida (modelizada/simplificada/negada) por un número abstracto (la media), que incluso puede que no coincida con la opinión, valoración, nivel de conocimientos, etc. de ninguna persona en particular? Por ejemplo, ¿por qué en vez de resumir con el estadístico media, no recurrir a la opinión del más viejo del grupo, como se ha hecho durante largo tiempo de la historia, cuando se quiere conocer la opinión del colectivo de gente estudiado? ¿Cómo se ha llegado a que el primer tipo de resumen, estadístico, esté legitimado mientras que el segundo aparezca como injustificado?
3. ¿Qué tiene la *variabilidad* (medida por la desviación típica o por la varianza) para que se convierta en objeto de explicación (uso del coeficiente de determinación como medida de la variabilidad explicada), habida cuenta que cada persona es diferente —como se dice coloquialmente, «es hija de su padre y de su madre»—? Por ejemplo, ¿por qué no situar como problema de investigación la explicación de la *homogeneidad*, en un momento en el que, tal como ocurre hoy en día, para mucha gente éste es el verdadero problema de nuestra civilización?

4. ¿Cómo es posible que de la constatación de que dos cosas están relacionadas estadísticamente (correlación y regresión) se pueda derivar que una puede explicar la otra, operativizando así un problema como el de la causalidad, siempre presente y nunca antes resuelto de esta manera operativa en la historia del pensamiento filosófico?
5. Y si a la inmensa mayoría de las personas no le preguntan, ¿cómo es que pueden aceptar como opiniones de *toda* la población —entre la que se encuentran ellas mismas— las del reducido grupo (muestra) que ha sido entrevistado en, por ejemplo, una encuesta, algo que nunca habría sido aceptado por personas de hace menos de un siglo?

Todas las preguntas anteriores, que como decía más arriba puede que para mucha gente carezcan de interés por referirse a hechos cotidianos y asumidos como naturales (los recuentos, el uso de la media, etc.), han adquirido este estatus debido a que la estadística, siguiendo el proceso histórico al que parcialmente se hace referencia en este artículo, «ha entrado en nuestras vidas», naturalizando los procedimientos —y, consecuentemente, los resultados obtenidos con los mismos— contenidos en las preguntas previamente formuladas (sobre los recuentos, las medias, las desviaciones típicas, etc.). Las páginas siguientes tratan de ofrecer claves para entender cómo se ha llegado a la situación actual.

2. Requisitos de la innovación

Que la estadística, tal como ahora la conocemos, se utilice masivamente y pueda cumplir la función naturalizadora que yo le atribuyo, algo que parece normal mirando el fenómeno desde nuestra posición actual, no quiere decir que para las gentes que contribuyeron a su desarrollo el empeño fuera sencillo. Para poder inventar algo, en este caso la estadística y sus estadísticos (recuentos, medias, desviaciones típicas, correlaciones, etc.), y que además lo inventado tenga trascendencia social, es necesario que se cumplan un par de requisitos, uno de naturaleza cognitiva (que el invento sea potencialmente inventable) y otro de naturaleza social (que aquello que se invente sea aceptado masivamente por la sociedad).

Como explica Lukacs en su *Historia y conciencia de clase*, existe un límite a lo que la conciencia (el entendimiento) de las personas puede pensar, determinado por cuales sean sus condiciones existenciales; superado dicho límite, el pensamiento (la conciencia) no es posible. Por ejemplo, no era pensable una encuesta por muestreo dirigida a los súbditos de Carlos II de Inglaterra, coetáneo de Graunt y de Petty, con la que se intentara ver la aceptación de la monarquía que se acababa de restaurar en ese país, cuando, en primer lugar, los súbditos no eran equivalentes, requisito previo para poder combinar sus opiniones a la hora de calcular, por ejemplo, una opinión media; además, ausente la idea de representatividad política, de la que se deriva la de representatividad estadística utilizada en el muestreo, nadie que no hubiera sido entrevistado habría aceptado los resultados obtenidos en la encuesta como representativos de todos los

habitantes del país; y, por último, por si fuera poco, dado que la legitimidad del rey no venía de los súbditos sino de Dios, a éste sería a quien, de llegarse a hacer una encuesta, habría que entrevistar. El uso de esta herramienta cognitiva (la encuesta estadística por muestreo) solo sería posible después de que la mayoría de los habitantes del país adquiriesen la categoría homogeneizadora de ciudadano (portador de derechos —civiles, sociales y políticos— y de obligaciones —fiscales y militares—), base para que pudieran ser combinados (en el cálculo de la media) e intercambiados (en el muestreo), y de que de manera generalizada se aceptase que la soberanía del gobernante reside precisamente en esos ciudadanos, que por esta misma condición son los que han de ser consultados a la hora de tomar decisiones que les afectan.

Al mismo tiempo, el que algo sea pensable es condición necesaria pero no suficiente para que ese algo (por ejemplo, un invento) tenga éxito social. Para garantizar el éxito de un invento es necesario que los intereses que subyacen al invento coincidan con los intereses de aquellos grupos sociales que lo pueden introducir masivamente en la sociedad. Por ejemplo, preocupado por el excesivo carácter modelizador/resumidor de la estadística al uso, basada, en el caso de las distribuciones de frecuencia univariadas, en la trilogía de los estadísticos de tendencia (media), dispersión (desviación típica) y forma (por ejemplo, coeficiente de asimetría), Tukey desarrolló hace algunos años un enfoque alternativo, el Análisis Exploratorio de los Datos (EDA), con el que se prestaba atención no sólo a los individuos que se ajustan al modelo, sino también a los desviados (los no ajustados). Por ejemplo, piénsese en un gráfico como el *box-plot*, con la cantidad de información que contiene sobre los distintos valores de la distribución, y compárese con la, a mi modo de ver, escasísima información contenida en la trilogía de los estadísticos al uso. Por ejemplo, a mis alumnos suelo compararles esta trilogía con la trilogía de las medidas que permiten describir a las candidatas a Miss: ambas ofrecen la información necesaria, desde el punto de vista de los intereses de las personas que la van a utilizar, para caracterizar a las personas, tanto a las que forman parte de la distribución (quizá millones) como a las aspirantes a reina de belleza, pero la información ofrecida por estos estadísticos es insuficiente para otras consideraciones alternativas, tanto sobre lo que sea una población como sobre lo que sea la belleza femenina. En una sociedad en la que dominan los intereses homogeneizadores (la media como referente), y donde la desviación (medida por la desviación típica) sólo se tiene en cuenta para ser explicada (uso de la regresión), como paso previo a su control (eliminación), el hecho de que alguien desarrolle una estadística que vaya un poco más allá de lo estrictamente necesario para satisfacer estos objetivos resulta gratuito, razón por la cual según mi opinión el EDA no ha tenido el éxito que se le pudo augurar en el momento de su aparición.

3. Estadísticas de recuento y de modelización

Con el par de ideas que acabo de introducir previamente, papel naturalizador de la estadística y existencia de requisitos necesarios para su desarrollo, es con el

que pienso las técnicas estadísticas, siguiendo su desarrollo histórico y la lógica que subyace a su aparición. Comprender el proceso de constitución de la estadística es entender la forma como esta disciplina ha contribuido a la naturalización del orden social, mediante el recurso a la construcción de hechos sociales (porcentajes, medias, correlaciones, etc.), tan sólidos como los llamados «naturales». A la hora de organizar esta exposición entiendo que hay una estadística de recuento y codificación, que fue la que primero se desarrolló históricamente, y otra estadística de modelización, que sólo fue posible a partir de que la estadística anterior hizo acopio de suficiente información como para buscar regularidades detrás de la gran maraña de números producidos. Y dentro de la modelización distingo una modelización de la homogeneidad, organizada en torno a la media y la distribución normal, y otra de la variabilidad, sustentada sobre una medida de variabilidad (la varianza) y otra que permita explicarla (la regresión y el coeficiente de determinación). Éste es el esquema que seguiré para dar cuenta del papel que, según mi opinión, cumple la estadística en la sociedad.

El recuento y la codificación del mundo²

La que yo llamo «estadística de recuento y codificación» es la estadística que se empieza a desarrollar a partir del siglo XVII, tomando como referencia la *aritmética política* inglesa y la *estadística descriptiva* alemana. Se trata de unos recuentos en los que, en primer lugar, se empieza a contar de manera distinta a como se venía haciendo hasta ese momento (cuenta el Estado, además de la Iglesia; se cuentan habitantes y no almas o vecinos, y se cuentan todo tipo de atributos de las personas, no solo aquéllos relacionados con su condición de soldados o contribuyentes), como consecuencia lógica de la aparición de monarquías absolutas que adquieren competencias sobre cada vez mayor número de aspectos de la vida de los individuos. Además de estas diferencias con el tipo de recuentos precedentes, el recuento que practican los aritméticos políticos aparece como *método* y fuente de legitimidad de una manera distinta de argumentar, no escolástica, basada en la utilización, como decía Petty, de *número, peso y medida*, en lugar de «utilizar sólo palabras comparativas y superlativas, y argumentos intelectuales». Desde el siglo XVII se asiste a un proceso mediante el cual progresivamente se amplían los objetivos fiscales (conocer el número de contribuyentes para repartir los impuestos) y militares (conocer el número de soldados potenciales del reino) que hasta entonces había tenido el recuento de la población, añadiendo otros de tipo político, económico y social, acordes con las nuevas funciones que van adquiriendo los Estados y las políticas, primero económicas y andando el tiempo también sociales, que promueven. Y como forma de legitimar estas nuevas monarquías y sus políticas, la ciencia (máxima expresión de la razón iluminista) desempeñará el papel que hasta

2. Este apartado aparece más desarrollado en un trabajo mío titulado «Reflexiones sociológicas sobre el recuento de la población (a partir del siglo XVII)». Mimeografiado.

entonces había desempeñado la interpretación de los textos sagrados: conocer la voluntad divina (la Verdad), ahora expresada en forma de leyes, escritas en los nuevos libros de la naturaleza y de la sociedad.

En definitiva, se trataba de una estadística, la del siglo XVII en Inglaterra, aunque con nombre de aritmética política, que ya buscaba el amparo de la ciencia (Petty era un admirador de Bacon, además de socio fundador de la *Royal Society for Improving of Natural Science*) con el fin de dar mayor autoridad al conocimiento que se produce con su uso. Llegado el siglo XIX, esta estadística contadora alcanzará su madurez, al lograr institucionalizar los censos periódicos, estandarizados a nivel internacional, que hoy en día conocemos. Precisamente sería Quetelet, autor al que me referiré inmediatamente, quien desempeñaría un papel fundamental en este proceso de institucionalización administrativa de la estadística, al estar legitimado para este proyecto por su altura científica, en tanto que «descubridor» de la primera regularidad (ley) estadística de tipo social, el *hombre medio* —el mercado de Adam Smith, con su mano invisible, sería otra gran ley, ésta de naturaleza económica, que precedería a la de Quetelet, en tanto que realización en el mundo económico del mismísimo orden Divino—. Resultado de estos recuentos, progresivamente legitimados por su carácter científico, será la producción masiva de hechos sociales (tasas de paro, inflación, crecimiento, pobreza, fecundidad), objetos medidos estadísticamente que, como dice Desrosières, «constituyen los puntos de apoyo para describir las situaciones económicas, denunciar injusticias sociales, justificar acciones políticas»; en definitiva, para el regular funcionamiento de la vida en sociedad. Justamente será esta producción de hechos sociales, de carácter natural, una de las formas como la estadística contribuirá a la naturalización del orden social —otra, tal como se irá viendo a continuación, será la modelización de dicho orden.

Relacionado con el recuento aparecen dos temas muy importantes, íntimamente relacionados, como son la elaboración de las «casillas» en las que se va a meter a la gente contada y su posterior clasificación dentro de las mismas. Estoy hablando del problema de la codificación, como actividad consistente en cortar el continuo con el que, de no ser por esta actividad, se presentaría el mundo ante nuestros ojos, para que una vez que el mundo ha sido cortado se pueda proceder a la clasificación de los individuos en los cortes efectuados (hombre-mujer, parado-no parado, depresivo-no depresivo, etc.). Sólo después de que ha tenido lugar la actividad cualitativa de cortar y clasificar se puede pensar en la actividad cuantitativa de contar. Se puede pensar, desde una perspectiva positivista, que las variables y las categorías (así vamos a llamar al producto de los cortes) pre-existen al momento de la clasificación de los individuos (el mundo ya nació cortado en hombres y mujeres, parados y no parados, depresivos y no depresivos, etc.), y que esta clasificación es automática y aproblemática (no hay ninguna duda de lo que sea un hombre o una mujer, un parado frente a un no parado o un depresivo frente a un no depresivo), o, por el contrario, también se puede pensar, ahora desde un enfoque constructivista, que las categorías no pre-existen al momento de la investigación

y que, además, la clasificación (medición) de la gente entraña más dificultad de la que podría parecer a primera vista. En este segundo caso resulta de interés tratar de dar sentido al proceso que ha seguido la construcción de las categorías estadísticas que se utilizan en cada época y la forma como se mide (clasifica), para entender tanto las naturalezas de la «jaula» en la que se termina metiendo a los contados como la del trabajo del «enjaulador». El recurso al análisis de la evolución que han tenido las clasificaciones que aparecen en los distintos censos españoles disponibles, partiendo del de los Millones, de 1591, podría servir para entender una posible lógica subyacente a los criterios de clasificación que en ellos se utilizan. Una reflexión sobre el problema de la medida, mediante la cual se trate de dilucidar el carácter, dado o construido, de los fenómenos sociales, completaría el tratamiento sobre la codificación.

La modelización de la homogeneidad: la media y el «hombre medio» de Quetelet³

Dentro de la estadística modelizadora procedo a distinguir entre la modelización de la homogeneidad, mediante el recurso a la media y la distribución normal, los dos primeros modelos estadísticos, y la modelización de la variabilidad, que tiene lugar haciendo uso de la varianza y su explicación mediante modelos como la regresión. Una vez que los individuos (no fuegos, vecinos o almas) son contados, pronto se observan regularidades en los recuentos (el propio Graunt, cofundador de la aritmética política en el siglo XVII, ya es un pionero de la constatación de regularidades en el campo de los fenómenos demográficos, incluido el suicidio). Con el tiempo, cuando cognitiva y socialmente resultó posible, Quetelet modelizaría estas regularidades recurriendo a la media y a la distribución normal, construyendo una idea que tuvo un enorme interés a lo largo de todo el siglo XIX, su hombre medio —con posterioridad la idea de *hombre medio*, producto, según Quetelet, de una tendencia (*penchant*) social, dejó de ser utilizada, pero no así la *media*, que ha pasado a ser la principal medida, también de tendencia, pero ahora no social, sino, algo más aséptico, *central*.

El hombre medio, al que se llegaba mediante el cálculo de la media y de la distribución normal —Quetelet solamente estaba interesado en las medias de distribuciones que tuvieran una forma normal, única circunstancia en la que entendía que estaba justificado el uso de la media—, que eran técnicas estadísticas por primera vez aplicadas a fenómenos sociales, digo que el hombre medio sería el modelo o síntesis del resto de los individuos singulares; y las desviaciones de estos individuos respecto al hombre medio Quetelet las con-

3. Se puede ver una ampliación a este tema en un par de trabajos míos: «Quetelet y la sociología», al que acompaña una traducción de una parte del libro de Quetelet *Physique sociale*, y en «Sociología, orden social y modelización estadística: Quetelet y el hombre medio». El primero de los artículos, junto a la traducción, aparecen en el número 87 de la *Revista Española de Investigaciones Sociológicas*; el segundo, en el número 3 de la revista *Empírea*, editada por la UNED.

sideraría como errores, producto de un accidente social (atribuidos, según el autor, al tipo de causas *accidentales* y *variables* que operan en la sociedad). Tal como diría Quetelet, al igual que ocurre cuando un tirador dispara a una diana, que la mayoría de los disparos se concentran en torno al punto central, distribuyéndose el resto por su alrededor según una curva de forma normal, así pasa en la sociedad cuando nos fijamos en las características de los individuos: la mayoría se concentran en torno a un valor central, mientras que el resto se reparte a su alrededor, siguiendo esa misma forma de distribución estadística (la distribución normal). En el caso del tirador hay una explicación a este fenómeno de concentración-dispersión. La existencia de una diana y la intención del tirador de dar en el blanco explicarían la concentración, mientras que el viento, la defectuosidad de la carabina o la impericia de quien la utiliza darían cuenta de la dispersión de los tiros. Tratándose de los individuos, también Quetelet tendría una explicación a este fenómeno de la concentración-dispersión: hay causas *constantes*, de tipo social, que llevan a que los hombres tiendan a ser iguales, «hombres medios», solo contrarrestadas por la existencia de otras causas, *variables* y *accidentales*, que se interponen en esa *tendencia* (*penchant*), y que son las responsables de la existencia de los hombres diferentes que se observan en la sociedad. Es decir, igual que una persona venida de otro mundo (y que por lo tanto no supiera nada de dianas y de tiradores), podría entender la concentración de los disparos cuando conociera la existencia de este juego de tiro y sus reglas, así también nosotros podemos explicarnos, gracias a la observación de Quetelet, la supuesta concentración de los individuos en torno a un valor central (el hombre medio), atribuida a la existencia de unas causas constantes que producirían esta regularidad. Y esas causas, que Quetelet empieza a identificar y medir, abriendo el camino a los sociólogos venideros, son tan naturales como las causas que puedan producir el movimiento de los planetas o la caída de los cuerpos, todas ellas (causas naturales y sociales) productoras de hechos inamovibles (naturales o sociales), que tan sólo pueden ser captados por los científicos, que los presentarán a la sociedad en forma de leyes.

En el caso de Quetelet, el uso del cálculo de probabilidades y de la ley de los grandes números, de aplicación exitosa en las ciencias naturales (en particular en la astronomía), servirían para legitimar las generalizaciones que hizo a partir de su observación de que hay regularidades y que ciertas distribuciones tienen una forma aproximadamente normal: si la altura media de las personas se repite año tras año (hay regularidad), con una distribución de las alturas de tipo normal (sometida a ley estadística), y además también hay regularidades en el número de delitos, suicidios o bodas (fenómenos estudiados por Quetelet), ello quiere decir que los fenómenos sociales (morales, diría nuestro autor) están sometidos a leyes, inmutables, establecidas al igual que las de la naturaleza por el Creador, a las que los hombres se han de ajustar y en las que solamente el científico, dado que es capaz de descubrirlas, puede intervenir —o facilitarles la información al gobernante y al negociante para que sean ellos los que intervengan.

La modelización de la variabilidad y su explicación: Galton y Pearson, la herencia y la evolución de las especies

Tuvieron que pasar algunos años antes de que el hombre medio de Quetelet, modelo estático de todo lo que era bello y bueno, según palabras del propio autor, terminara por ser considerado un hombre mediocre, como corresponde a alguien que tiene los valores medios de la distribución en una sociedad en la que Darwin acaba de decir que las especies evolucionan (el modelo de orden social se hace dinámico) y que sólo los más fuertes sobreviven (con la lucha por la supervivencia, frente a la selección natural, como motor del cambio). A Galton, primo del famoso científico evolucionista, le correspondería el mérito (o demérito, si es que se tiene una idea lamarckiana de la evolución, que acentúa la cooperación y la adaptación frente a la competencia darwinista) de plantearse el hecho de que si hay débiles y fuertes, y para la selección de las especies, incluida la humana, es importante esta diferencia, ¿cómo considerar un accidente, según era el pensamiento de Quetelet, la variabilidad de los individuos para centrarse en el hombre medio? Justamente la variabilidad (motor de la evolución, el nuevo modelo teórico de la época) y no la tendencia (motor del hombre medio, ahora ya un viejo modelo, heredero quizá del canto a la igualdad de la Revolución Francesa) era lo que había que estudiar. Y para ello Galton primero construyó el recorrido intercuartílico y posteriormente utilizó la desviación típica, como medidas de dicho concepto de heterogeneidad.

En el trabajo de Galton se trataba simplemente de hacer una traslación del modelo que se utilizaba para interpretar el mundo natural al campo de lo social. Una vez que la desviación es lo importante, y que lo mediocre pasa a ser considerado como un lastre para la evolución de las especies (en el caso de Galton para el desarrollo del Imperio británico), sólo faltaban dos cosas, íntimamente unidas: desarrollar la *teoría* sobre la herencia que permitiese mejorar la raza, que se consideraba la «causa constante» que explicaba la transmisión tanto de características físicas como sociales, y crear las herramientas cognitivas (desarrollar el *método*) necesarias para medir (la varianza o la desviación típica) y explicar (la correlación y la regresión) la variabilidad observada en las personas. Con el primero de los objetivos Galton crearía la eugenesia, primero como ciencia y luego como asociación para la mejora de la raza humana, que tendría gran repercusión en el mundo occidental hasta acabada la Segunda Guerra Mundial, cuando cayó en desgracia debido al uso radical hecho por los alemanes de esta idea eugenésica.

Como resultado de esta batalla de Galton y Pearson por la introducción de la eugenesia han quedado ideas y técnicas que han marcado definitivamente los desarrollos, tanto de la estadística como de la investigación social que la utiliza. Por un lado, y por oposición a Quetelet, del trabajo de Galton y Pearson ha quedado la idea de que lo importante es la variabilidad, de tal manera que sin variables no habría estadística (las constantes no tienen cabida en la matriz de datos, que es la base del análisis estadístico); por otro lado, si la variabilidad es lo importante, habrá que explicarla, como paso previo a su control. Y,

como forma operativa de explicación, superando discusiones anteriores de tipo filosófico sobre qué es la causalidad (¿si Aristóteles, con sus causas eficientes, materiales, etc. levantara la cabeza!), Pearson fundamentó la idea de contingencia (dos cosas están relacionadas cuando una ocurre al tiempo que la otra), que es susceptible de ser medida utilizando modelos estadísticos como la correlación y la regresión, desarrollados por Galton y Pearson —con el tiempo, y para evitar el problema de las correlaciones espurias, que aparecería poco después de empezar a utilizar estas técnicas estadísticas, un discípulo suyo, Fisher, desarrolló los diseños experimentales, que es la técnica que compite en la actualidad con la correlación a la hora de establecer causalidades—. El salto que dieron Galton y Pearson en la investigación fue enorme, ya que sentaron las bases de lo que hoy son las técnicas estadísticas de análisis de los datos: del recuento de la población (Graunt y Petty), que hizo posible la modelización estadística de las regularidades observadas (Quetelet y la media), pero sin que se supieran explicar y cuantificar estadísticamente (apelación más o menos genérica de este autor a la influencia de causas constantes y variables), digo que del trabajo de aritméticos políticos y de Quetelet, con el de Galton y Pearson se pasó a la explicación estadística de la nueva modelización (la variabilidad), mediante el recurso a la idea de varianza explicada (medida mediante el coeficiente de determinación). Debido a estos desarrollos, cuando en la actualidad se hacen investigaciones no sólo se habla de la influencia que tiene una variable sobre otra, sino que además se puede medir la intensidad de dicha influencia, expresándola en tantos por ciento de variabilidad explicada.

4. La generalización: el muestreo

Por último, y después de ver en el apartado anterior el recuento y la modelización, digamos que no todos los individuos a los que hay que gobernar van a poder ser contados y, posteriormente, modelizados de la manera regular que exigen las labores de gobierno, sea de ciudadanos o de consumidores. A lo largo de la historia, cuando el número de personas era muy grande, como ocurría al tratar de contar la población de países, ciudades, etc., el tratamiento de todas ellas resultaba casi imposible, haciendo que esta operación sólo se pudiera efectuar muy de tarde en tarde, como atestigua la falta de regularidad con la que se realizaron los censos hasta finales del siglo XIX. Con la llegada del siglo XX, la equivalencia entre los hombres, convertidos todos en ciudadanos, que había permitido la combinación de sus comportamientos para producir un comportamiento medio (la media), bien fuera modélico o mediocre, digo que la equivalencia, además de la combinación de los individuos (sus opiniones) iba a permitir hacerlos intercambiables, de manera que unos pudieran actuar en representación de los otros —lo mismo que ocurre en la representación política, que es la base de posibilidad de la representación estadística, aunque los procedimientos de selección de los representantes sean distintos. Tal intercambio fue técnicamente posible con la aplicación de las técnicas de muestreo, de uso generalizado una vez que se inició el segundo cuarto

de nuestro siglo y los procesos electorales se hubieron generalizado, con la participación potencial de casi todos los individuos de un país, en las que unos miembros de la población, intercambiables (cualquiera vale con tal de que haya sido seleccionado al azar), actúan en representación de todos los demás. A partir de ese momento dejarán de ser necesarios, tanto la utilización de multiplicadores con los que estimar el total de la población a partir del grupo de personas que han sido contadas (procedimiento utilizado, por ejemplo, por los aritméticos políticos), como las encuestas monográficas, a lo Le Play, dirigidas a una población que se pudiera considerar como tipo: a condición de que la selección se haga de la manera oportuna, dando a todos los individuos de la población igual (o conocida) probabilidad de ser elegidos, y de que, una vez elegidos, los individuos seleccionados participen en la investigación (no haya no respuesta), bastará estudiar un grupo reducido de personas (una muestra) para generalizar (inferir) las conclusiones obtenidas a la población de la cual se ha extraído. La inferencia que se haga utilizando este método no evitará el error, pero, si se han cumplido los requisitos necesarios del muestreo, al menos dejará al investigador la posibilidad de medirlo.

A diferencia de lo que hacían Quetelet, Galton o Pearson, el cálculo de probabilidades ya no solo permitirá descubrir regularidades a las que se calificará de leyes, pudiendo modelizar así los fenómenos sociales al igual que se venía haciendo con los llamados «naturales», sino que además permitirá hacer todo tipo de investigaciones de colectivos amplísimos, utilizando la aplicación de dicho cálculo (de probabilidades) al muestreo. Este procedimiento estadístico, que tiene un valor excepcional para producir hechos sociales, diríamos en tono coloquial que «como churros», no estará exento de limitaciones, propias de los supuestos heurísticos (por ejemplo, ¿existe un valor verdadero por relación al cual se puedan medir tanto el error fijo, o sesgo, como el error variable de muestreo, que son los referentes estadísticos utilizados para ver la bondad de la muestra?) y sociales (¿es posible trasladar los supuestos del muestreo, válidos para poblaciones de canicas o de tiradas de dados, a poblaciones humanas, habida cuenta de, por ejemplo, la falta de disponibilidad a participar en las investigaciones de una parte importante de la población?) que lo sustentan, a las que habrá que referirse al tratar de este tema en un trabajo específico⁴.

Bibliografía

- GALTON, Francis (1889). *Natural Inheritance*. Londres: McMillan. Edición parcial de sus obras, en castellano, a cargo de ÁLVAREZ PELÁEZ, Raquel (1988). *Francis Galton. Herencia y eugenesia*. Madrid: Alianza Universidad.
- GRAUNT, John (1661). *Observations upon the Bills of Mortality*. Edición de Charles Henry Hull (1964). *The Economic Writings of Sir William Petty. Together with the*

4. En mi libro *La bondad de la encuesta: el caso de la no respuesta* se hace una reflexión sobre los problemas que se plantean, con particular referencia a la no respuesta, al utilizar las técnicas de muestreo en la investigación social (Alianza Editorial Materiales, 2000).

Observations upon the Bills of Mortality, more probably by Captain John Graunt. Nueva York: August M. Kelley Bookseller.

PEARSON, Karl (1892). *The Grammar of Science*. Londres: Scott. Reedición de 1957, sobre la tercera edición de Pearson, de 1911. Nueva York: Meridian Books, Inc.

PETTY, William (1676). *Political Arithmetic*. Edición de Charles Henry Hull (1964). *The Economic Writings of Sir William Petty. Together with the Observations upon the Bills of Mortality, more probably by Captain John Graunt*. Nueva York: August M. Kelley Bookseller.

QUETELET, Adolphe (1835). *Sur l'homme et le developpement de ses facultés ou essai de physique sociale*. París: Bachelier, Imprimeur-Libraire.

— (1869). *Physique sociale ou essai sur le developpement des facultés de l'homme*. Bruselas: C. Buqcart, Libraire editeur. Reedición anotada de 1997, por Éric Vilquin y Jean-Paul Sanderson. Bruselas: Académie Royale de Belgique.