Presentación

Los peligros de la Estadística[1]

JEAN-GUY PREVOST
Universidad du Québec à Montréal

El texto que sigue a continuación es el discurso inaugural pronunciado por Corrado Gini el 9 de octubre de 1939 con ocasión de la primera reunión científica de la *Societa Italiana di Statistica (SIS)*. Se trata, por tanto, de una alocución que marca un hito en la historia de la «escuela italiana de estadística», una corriente con la que Gini trata de reafirmar el carácter distintivo de la estadística italiana por contraposición a la orientación de la escuela anglo-sajona; en un momento además, en el que está a punto de producirse el *mainstream* de la disciplina. Globalmente, la escuela italiana de estadística que podemos identificar con Gini y la nebulosa de sus discípulos y colaboradores, se distingue por el interés que otorgan a la descripción de series estadísticas, con especial insistencia en la teoría de las medias, el estudio de la variabilidad y la concentración, y la relación entre variables, así como por la atención a lo que hoy se conoce como diseño de encuestas (*survey design)* aplicado a poblaciones. Por contraste, la corriente anglosajona, en la línea de Karl Pearson y luego de Roland Fisher, centró sus esfuerzos en la problemática del error, así como en cuestiones de muestreo y significación estadística. Ya en 1926, con ocasión de una visita a la London School of Economics, Gini había defendido los meritos de la singularidad italiana y había propuesto, de manera un poco provocativa «*statistics with the least mathematical means posible»,* tomando directamente a contrapié la definición propuesta por Fisher un año antes en su clásico «Statistical Methods for Research Workers»: «*The science of statistcs is essentially a branch of applied mathema-*

Nota editorial: Agradecemos a Jean-Guy Prévost su propuesta de traducir y publicar en la sección «texto clásico» el artículo de Corrado Gini: «I pericoli della statistica». Traducción de la presentación: José M. Arribas.

[1] Gini, C. «I pericoli della statistica», *Supplemento statistico ai Nuovi problemi di politica, storia ed economia*, vol. 5, no 2-3-4, 1939, p. 1-44 (en 1941, esta revista se convierte en «*Statistica»,* el nombre que conserva hasta hoy.) El texto procede de Gini, C. *Questioni fondamentali di probabilità e statistica*, vol. I, Rome, Istituto di statistica, 1968, p. 175-220, después en Gini, C. *Statistica e Induzione*, Bologne, CLUEB, 2001, p. 27-70 (con una traducción en inglés titulada «The Dangers of Statistics», p. 257-303).

tics and can be considered as mathematics applied to observed data»[2]. Gini volvería sobre esas características particulares de la estadística italiana en uno de sus últimos textos aparecidos en 1965 sin hacer ninguna concesión a la originalidad o al valor de las matemáticas aplicadas[3].

En 1939, el autor de *«I pericoli della statistica»,* era sin duda la figura dominante de la estadística y de las ciencias sociales italianas: su producción científica, que contaba entonces con más de 450 títulos, cubría un amplio espectro de disciplinas y problemas que van desde la metodología estadística a la demografía, la sociología, la economía, pasando por el estudio de los fenómenos migratorios o de las poblaciones primitivas[4]. Emprendedor intelectual de talla, Gini fue profesor en las universidades de Cagliari (1910-1913), de Padua (1913-1925), de Roma (1925-1955), donde puso en marcha la Escuela de Estadística (1927), de la Facultad de Ciencias estadísticas, demográficas y actuariales (1936); fue director y fundador del *Comitato Italiano per lo Studio dei Problemi della Popolazione (CISP)* y de las revistas *Metron,Genus* y *La Vita Economica Italiana*, así como impulsor del monumental *Trattato Elementare di Statistica.*

En el plano internacional trabajó a título de experto en numerosas ocasiones para la Sociedad de naciones (SDN) y participó activamente en los trabajos del Instituto Internacional de Estadística (IIS)[5]. La participación de Gini en la vida político-burocrática italiana del período de entre guerras es también importante y tiene diversas vueltas pues colabora como experto en el Gobierno antes y después de «la marcha sobre Roma». En 1925, Gini firma el manifiesto de los intelectuales fascistas iniciado por el filósofo y ministro de Educación Giovanni Gentile, y es designado como uno de los dieciocho expertos («solons») encargados de revisar la constitución italiana. Desde 1926, año de la fundación del Instituto central de Estadística (ISTAT), hasta 1932, Corrado Gini fue nombrado por Mussolini presidente del Instituto. En 1939, en un contexto de guerra en el que los campos rivales están ya trazados, aunque aún no se adivina la amplitud de la catástrofe, la oposición entre escuelas estadísticas nacionales se endureció

[2] Gini, C. «The Contributions of Italy to Modern Statistical Methods», *Journal of the Royal Statistical Society*, vol. 89 no 4, 1926, p. 707; Fisher, R. A. *Statistical Methods for Research Workers*, Londres, Oliver and Boyd, 1925, p. 1.

[3] Gini, C. «On the Characteristics of Italian Statistics», *The Journal of the Royal Statistical Society*, Series A (General), vol. 128, no 1, 1965, p. 89-109.

[4] Hay una bibliografía casi completa de las obras de Gini en V. Castellano, «Corrado Gini: à Memoir», *Metron*, vol. 24, no 4, 1965, p. 2-84. Esta bibliografía comporta 87 textos y notas de cursos y 827 «memorias, notas y artículos»

[5] Nota del traductor: Gini fue también un importante impulsor de la Estadística española. Sus conexiones con José Antonio Vandellós, director del Laboratorio de estadística Económica y Financiera de Barcelona y con Severino Aznar, facilitará la traducción de sus obras al español, sus viajes y conferencias por España, así como su participación en la creación de la Revista Internacional de Sociología donde publica con relativa frecuencia sobre demografía. Véase al respecto Arribas J.M. y Almazán A. «*La estadística española de posguerra (1939-1958)* en A.H.E.P. «Historia de la Probabilidad y la Estadística III», Delta Publicaciones, 2006.

repentinamente y la ruptura parcial de los intercambios científicos dio a los discursos sobre la autarquía un sentido muy concreto.

Pero no hay que interpretar solamente «*I pericoli della statistica*» bajo el prisma de un chauvinismo nacional estimulado por el contexto político e histórico, pues el texto se inserta en una trama de contribuciones sobre la inferencia y la inducción que combina reflexiones sobre la noción de probabilidad redactadas por Gini entre 1907 y 1911[6], trabajos de la comisión del Instituto Internacional de Estadísticas (IIS) encargada de preparar un informe sobre el «método representativo», la tentativa hecha por Gini y Luigi Galvani de determinar una muestra representativa por el método de la elección razonada, así como la célebre intervención de Jerzy Neyman ante la *Royal Statistical Sciety* en 1934[7], o la puesta a punto de Fisher de la noción de significación estadística y de los test asociados a ella[8]. Al mismo tiempo, «*I pericoli della statistica»,* constituye la primera tanda de una «revisión crítica de los fundamentos de la estadística» que dará lugar entre 1939 y 1946 a no menos de quince intervenciones en la SIS[9]. Un cierto número de trabajos italianos han subrayado la importancia del texto de 1939 en la historia de la disciplina (en Italia al menos) interpretándolo como una forma de «objetivación bayesiana», y subrayando la importancia dada por Gini a las probabilidades «a priori» (*priors)* en «el proceso lógico de determinación de inferencias»[10]. Las críticas adelantadas por Gini sobre las concepciones que él atribuye sobre todo a Fisher, pueden aproximarse, como señala I.Scardovi, a las formuladas en ese mismo momento desde la perspectiva bayesiana, por Harold Jeffreys en *Theory of Probability*[11].

[6] Estos textos han sido recogidos en su mayor parte en C. Gini, *Questioni fondamentali di probabilità e statistica*, vol. I, Rome, Istituto di statistica, 1968 et dans *Statistica e induzione, op. cit.*

[7] Jensen, A. «Report on the Representative Method in Statistics», *Bulletin de l'Institut international de statistique*, vol. 22, no 1, 1925, p. 359-380; Galvani, L. et C. Gini. *Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione* [1º dicembre 1921], *Annali di statistica*, Série 6, no 4, Rome, ISTAT, 1929; Neyman, J. «On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection», *Journal of the Royal Statistical Society*, vol. 97, no 4, 1934, p. 558-625.

[8] Fisher, R. A. *Statistical Methods for Research Workers*, *op. cit.*

[9] Una síntesis de los resultados de esta revisión es presentada por Gini en «Intorno alle basi logiche e alla portata gnoseologica del metodo statistico», *Statistica*, vol. 5-6, numero único, 1945-1946, p. 1-38.

[10] Frosini, B. V. «Objective Bayesian intervals : some remarks on Gini's approach», *Metron*, vol. 63, no 3, 2005, p. 435-450; Forcina, A. «Gini's Contributions to the Theory of Inference», *International Statistical Review*, vol. 50, no 1, p. 65-70. Para una visión general de las aportaciones de Gini en materia de inferencia estadística ver también A. Herzel et G. Leti, «Italian Contributions to Statistical Inference», *Metron*, vol. 35, no 1-2, 1977, p. 3-48.

[11] Scardovi, I. «Presentazione», in C. Gini. *Statistica e induzione, op. cit.*, p. xii. Para una visió en profundidad del debate entre Fisher y Jeffreys durante los años 1930, ver D. Howie. *Interpreting Probability. Controversies and Developments in the Early Twentieth Century*, Cambridge, Cambridge University Press, 2002. La obra de Jeffreys es citada favorablemente por Gini, aunque de forma breve en «I testi di significatività», comunicación presentada en 1943 y recogida en *Statistica e induzione, op. cit.*, p. 75-118 (la mencion de Jeffreys se encuentra en la nota 50.

El interés de este texto para una revista de metodología de las ciencias sociales se debe al hecho de que aparezca esbozada la crítica de los test de significación que será desarrollada por Gini en otros textos posteriores. El entusiasmo con el que sicólogos, economistas y sociólogos, o los medios de investigación médica, han acogido los test de significación, y en concreto la versión de Roland Fisher, constituye uno de las historias de éxito más remarcable de las ciencias sociales del siglo XX; aunque el éxito esté acompañado desde los primeros momentos de una crítica sostenida y, a veces, estridente[12]. A diferencia de la correspondencia que intercambiaron en la época Fisher y Gini, este texto no fue traducido a la lengua inglesa hasta 2001, y no parece que haya sido demasiado conocido entre los adversarios de los test[13]. Las críticas que contiene, anuncian o esbozan lo que será desarrollado con posterioridad al hilo del debate. Según Gini, el enfoque que funda la teoría de los test de significación descansa sobre un falso razonamiento. Se puede, en efecto, deducir de la probabilidad que el error «aleatorio»[14] excede el doble, el triple, o no importa qué umbral del error cuadrático medio de las probabilidades respectivas, según las cuales, este error sería «aleatorio», o por el contrario, «significativo». Ilustrando el fallo que él denuncia por medio de citas sacadas de un tratado de Yule y Kendall (*An Introduction to the Theory of Statistics,* 1937*)* y de un manual de Fisher (*Statistical Metrhods for Research Workers, 4 ed.,*1932), Gini describe en los términos siguientes la posición que le parece errónea:

> *«Prescindamos de las ya mencionadas objeciones y admitamos conocer la probabilidad Pa de que el error aleatorio de una constante exceda un cierto límite y por tanto la probabilidad 1- Pa de que quede entre dichos límites. Si pudiéramos decir que el verdadero valor de la constante queda, con cierta probabilidad, en un cierto entorno, ya sabríamos algo. Pero ¿podemos también decir que, si el error de la constante en cuestión excede tales límites esto supone una probabilidad Pa de que sea puramente aleatoria y por tanto una probabilidad 1 - Pa de que sea significativa?*
>
> *Es lo que generalmente se hace. Tenemos una probabilidad contra 22 de que la intensidad del error aleatorio exceda el doble, y una contra 370 de que exceda el triple del error cuadrático medio; por tanto se concluye: un error que exceda al doble del error cuadrático medio tiene una probabilidad de 1*

[12] La última de las contribuciones a este debate es la obra de Ziliak, S. T. and D. N. McCloskey. *The Cult of Statistical Significance. How the Standard Error Costs Us Jobs, Justice, and Lives*, Ann Arbor, The University of Michigan Press, 2008.

[13] Una parte de la correspondencia entre Fisher et Gini está disponible en http://digital.library.adelaide. edu.au /coll/special//fisher/corres/gini/index.html. Las relaciones entre ambos parece que fueron buenas: Fisher había publicado en *Metron* en varias ocasiones, se encontraron por primera vez en 1924 con ocasión del congreso internacional de matemática de Toronto y su correspondencia abraca tres decenios. Los dos están convencidos, en una época en la que la estadística conoce un éxito espectacular, que la estadística no debe desarrollarse en ruptura con sus aplicaciones prácticas.

[14] N.T. «accidental» en el texto.

> *sobre 22 de ser aleatorio y 21 contra 1 de ser significativo, y un error que exceda el triple del error cuadrático medio tiene una probabilidad de 1 sobre 370 de ser aleatorio y por tanto, 369 contra 1 de ser significativo*[15]*».*

Los partidarios de la teoría de los test de significación confunden, nos dice Gini, dos cosas bien distintas: la probabilidad de un resultado y la probabilidad de que ese resultado sea debida al azar más que a causas sistémicas, o, dicho de otro modo, la probabilidad de que el error de medida alcance un umbral dado, y «el grado de fiabilidad que se le debe atribuir a un resultado estadístico o a una diferencia entre dos resultados»[16]. En esta época aparece de la pluma de Fisher el argumento rechazado por Gini, según el cual, es la aplicación a veces negligente, incompetente o poco escrupulosa de los tests, lo que estaría en cuestión, más que la lógica misma[17]. De hecho, la interpretación que ofrece Gini en el pasaje citado anteriormente se aplica particularmente bien al uso mecánico que a menudo se ha hecho de estos test (¿no es su carácter mecánico lo que le ha valido a los test de significación el favor que han reencontrado en los practicantes de la investigación social y medica?). Pero fundamentalmente, lo que Gini pone en duda es la posibilidad de distinguir la parte de variación debida al efecto de errores «aleatorios» (accidentales) y la que es debida al efecto de causas «constantes» sobre la base de la intensidad de esos efectos. Ello le conduce a una posición radicalmente escéptica en cuanto a la posibilidad de una distribución previsible de datos que no respondan a condiciones estrictas y, en consecuencia, a la posibilidad de estimar en los umbrales definidos la representatividad de una muestra extraída al azar. Este escepticismo no le conduce, sin embargo, a una negación del principio de causalidad, sino a una insistencia sobre la imprevisibilidad de los fenómenos en razón de los límites de nuestros conocimientos[18].

En esta discusión en forma de soliloquio, hay un cierto nivel de confusión que se debe, en gran parte, a las dificultades objetivas que existen en ese momento para la circulación de ideas. La correspondencia entre Gini y Fisher se interrumpe en el momento en que Italia entra en la guerra mundial, y serán dos estadísticos alemanes, H. von Schelling y M.P. Geppert, los encargados de defender las tests de significación en la séptima reunión de la Sociedad Italiana de Estadística (SIS) celebrada en 1943[19]. Pero la confusión se debe también al hecho de que Gini no estaba al tanto de todos los elementos del debate, como

[15] C. Gini, «I pericoli della statistica», *loc. cit.,* p. 15.

[16] C. Gini, «I pericoli della statistica», *loc. cit*., p. 16.

[17] Cartas de Fisher à Gini, fechadas el 22 diciembre de 1939 y el 3 de mayo de 1940, citadas en «I testi di significatività», *loc. cit*., p. 312, n. 13 et p. 313, n. 14, 15, 16 et 17. La primera de estas cartas no figura en el archivo de Fisher digitalizado.

[18] C. Gini, «I pericoli della statistica», *loc. cit*., p. 28.

[19] Von Schelling, H. «Ma non può dopo tutto essere un caso?», *Atti della VI e VII Riunione della Società Italiana di Statistica*, Rome, 1945, p. 307-314; Geppert, M.P. «Sul valore dei cosidetti testi di significatività», *ibid*., p. 315-323.

pone de manifiesto el análisis de las fuentes hechas por Forcina[20]. Gini tiende a asimilar las concepciones y los procedimientos puestos al principio por Fisher junto a las propuestas posteriores de Neyman y Egon Pearson, sin embargo, estos protagonistas manifestarán con los años una clara conciencia de las diferencias entre la probabilidad fiduciaria (*fiducial probability)* de un lado, y los test de significación fundados sobre la aceptación o el rechazo de la hipótesis nula, o del enfoque del *nivel de confianza* (confidence level) y los test fundados sobre la comparación entre dos hipótesis[21]. Pero el contexto cognitivo —además de las restricciones debidas a la guerra— debió, sin duda, jugar algún papel en la incomprensión y divergencias entre los estadísticos. El enfoque del intervalo de confianza desarrollado por Neyman y Pearson es esencialmente una herramienta para la toma de decisiones particularmente adaptado al contexto de la producción industrial. Fisher lo abordará oponiendo una retórica fundada sobre el ideal de una ciencia desinteresada, pero sus propios test de significación, desarrollados en la granja experimental de Rothamsted, surgieron con un enfoque completamente práctico. La posición de Gini en esta época aparece todavía más alejada de toda preocupación instrumental: el estadístico italiano que ya había abandonado el INSTAT, hace mucho tiempo que deseaba imprimir a la SIS una orientación estrictamente científica y metodológica, por contraste con la posición adoptada por su rival la *Società Italiana di Demografia e Statistica* (SIDS) —más orientada hacia la evaluación de políticas públicas— creada en el mismo momento. Se podría pensar que el rigorismo y el escepticismo de los puntos de vista expresados por Gini en «*I pericoli della statistica»* y en los textos que le siguieron, se desprenden de la lógica propia de su interpretación del concepto de probabilidad, pero también habría que tener en cuenta lo mucho que deben a las demarcaciones existentes en el campo de la estadística italiana.

[20] Forcina, *loc. cit*., p. 67.

[21] Howie, *op. cit*., p. 177. Esta confusión será disipada al terminar la Guerra: ver Gini, C. «Intorno alle basi logiche e alla portata gnoseologica del metodo statistico», *loc. cit*., n. 22.

Los peligros de la estadística[1]

Corrado Gini[2]

Los inconvenientes y peligros a los que da lugar la investigación de fenómenos naturales y sociales sin la ayuda del método estadístico ya han sido repetidamente señalados y puede decirse que están satisfactoriamente documentados: la imposibilidad de inferir cierta regularidad de las masas de datos; las conclusiones erróneas a que se llega, ya sea por el hecho de que unos casos se presenten a nuestra observación con más frecuencia que otros o porque, aunque aparezcan ante nuestra observación con la misma frecuencia, llaman más nuestra atención y dejan una huella más profunda en nuestra memoria. Bajo la mejor de las hipótesis sólo se llega a conclusiones aproximadas, ya sea porque con la observación común sólo es posible considerar la regularidad cualitativamente, o por lo referente a la intensidad de los fenómenos singulares, que se pueden investigar cuantitativamente pero sin la precisión deseable. Es a partir de estos inconvenientes y peligros cuando se aprecia la utilidad, y a menudo la necesidad, del método estadístico[3].

Por otra parte ¿no tiene también el método estadístico sus inconvenientes y peligros?

«Con la estadística —se afirma ante todo— se demuestran las tesis más contradictorias»

«Con la estadística —se afirma también— se obtienen a veces conclusiones destinadas a ser falsadas por investigaciones posteriores, llevadas a cabo con el mismo método estadístico o con otros métodos.»

Los estadísticos no se han ocupado —que yo sepa— de tomar en consideración tales acusaciones. Y se han equivocado. No sirve meter la cabeza bajo el ala, como los avestruces, para ignorar una realidad desagradable. La realidad es que tales acusaciones están muy extendidas entre el gran público, y no se limitan al gran público, al contrario, a veces han afectado a científicos de valor indiscutible.

Tampoco convence decir que existen discrepancias en todas las ciencias y que precisamente en esa discusión aflora la verdad, y que igualmente en todas las

---

[1] Texto del discurso inaugural pronunciado por Corrado Gini el 9 de octubre de 1939 en la constitución de la *Societa Italiana di Statistica*. Editado en Gini, C. «I pericoli della statistica», *Supplemento statistico ai Nuovi problemi di politica, storia ed economia*, vol. 5, no 2-3-4, 1939, p. 1-44. Traducción de Alicia García-Redondo y Alejandro Almazán. N. T.

[2] Roma, R. Università, Istituto di Statística, septiembre 1939.

[3] Cfr. A propósito, los *Apuntes de estadística metodológica* citados en nuestras clases de las Universidades de Padua y de Roma a partir de 1914, se han traducido al español, tras revisión, bajo el título *Curso de Estadística*, Barcelona, Editorial Labor, S.A., 1935, pag. 10-14.

ciencias los resultados obtenidos son continuamente sujetos a revisión y corrección, donde surge, precisamente, el progreso científico. No convence, porque la acusación de contradicción atribuida a los resultados de la Estadística se basa en el hecho de que muchas veces las discrepancias a las que conduce no se corrigen, o al menos, no se corrigen con el método estadístico, y la acusación de insuficiencia y falsedad viene del hecho de que a menudo los resultados obtenidos por la Estadística no se encuentran integrados y pormenorizados, y además son anulados o contradichos por investigaciones posteriores.

En fin, basta con decir que no se puede echar la culpa al método de los errores de quien lo aplica. El método sería correcto y, en cambio, la aplicación incorrecta. Los métodos de investigación científica no son más que instrumentos, que tienen valor en cuanto se pueden aplicar prácticamente, y son comprendidos por el común de los mortales. Por otra parte, los resultados contradictorios o falsos, que se achacan a la Estadística, no siempre se han elaborado por personas inexpertas, sino a menudo por especialistas y a veces por personalidades científicas de primerísimo orden. De acuerdo, tales resultados discordantes o erróneos no deben hacernos perder de vista los muchos resultados verídicos y concordantes a los que se ha llegado con el método estadístico. Pero, si los primeros, según una opinión muy difundida entre los no estadísticos, conceden a los segundos cierta importancia, es necesario reconocer que el método estadístico es, incluso en manos de los estadísticos, un método peligroso. Por tanto, conviene buscar la forma de disminuir la peligrosidad, teniendo también en cuenta la capacidad normal de las personas que lo utilizan.

Por otra parte, la misma afirmación de que los métodos estadísticos que se emplean son, siempre y todos, correctos, y que los inconvenientes, que se lamentan en las conclusiones alcanzadas, dependen, siempre y todos, de quienes los aplican menos correctamente es, al cabo, gratuita y bien pudiera ser demasiado optimista. Merece, en cualquier caso, ser analizada críticamente.

De todos modos, si los resultados contradictorios o falsedades, a los cuales a veces llega la Estadística dependen de la deficiencia de los métodos o de la insuficiencia de quien los aplica, la naturaleza y la causa de tal deficiencia o insuficiencia merecen ser investigadas, reconocidas y señaladas para evitar incurrir de nuevo en ellas e informar, si fuera posible, del remedio.

* * *

En nuestra investigación, procederemos basándonos en ejemplos concretos, generalizando después sobre las conclusiones obtenidas.

Veamos un ejemplo de conclusiones contradictorias.

A y B son dos países. Se pregunta en cuál de los dos es más alta la mortalidad. La razón entre el número de muertos y la población es mayor en A que en B (supongamos que la diferencia relativa sea del 10%) y muchos estadísticos concluyen que la mortalidad es más elevada en A. Pero otros dicen: No, esa conclusión no es correcta: más aún, es errónea. En A hay muchos más niños que en B, y entre los niños, como es bien sabido, la mortalidad es particularmente elevada; conviene eliminar la influencia de la edad, con el método, por ejemplo, de la población tipo. Se observa ahora que en cambio la mortalidad es más elevada en B por una diferencia relativa —supongamos— del 8%. ¡Un momento! —objeta un tercer grupo de estadísticos— A es un país de marcada emigración y el número de hombres resulta más bajo que el de mujeres, lo que por el contrario no sucede en B; debemos eliminar también la influencia del sexo ya que es cierto que la mortalidad, al igual que otras circunstancias, es más alta en el sexo masculino. Eliminada también dicha influencia, los coeficientes de mortalidad resultan iguales en los dos países. Pero no es suficiente —observa un cuarto grupo: la desproporción de los sexos hace más raro el matrimonio y por tanto más alto el porcentaje de cónyuges en A; también debe eliminarse la influencia del estado civil. Hecha esta tercera eliminación, la mortalidad resulta de nuevo más elevada en B, pero ligeramente (diferencia relativa del 4%). Y ¿por qué no eliminar también la influencia de la profesión, que tanto influye en la mortalidad? Eliminado también este factor, la desventaja en la diferencia relativa de mortalidad del país B sube al 15%. Pero ¿por qué no eliminar también la influencia de la riqueza? Después de esta última eliminación, la mortalidad resulta de nuevo más alta en A, pero de forma apenas perceptible (diferencia relativa del 1%). Puede que el clima sea muy diferente en los dos países. Admitamos que se pueda eliminar también la influencia de este factor, y al final la mortalidad resulta casi igual en los dos países con una diferencia relativa del 2% a favor de B. Así, en una cuestión tan simple como la de saber si la mortalidad es más elevada en A o en B, hemos pasado, con los métodos estadísticos, a través de siete conclusiones más o menos distintas. La diferencia relativa a favor de B, resulta sucesivamente de +10%, –8%, ±0%, –4%, –15%, +1%, –2%. No puede decirse que se haya tratado de sucesivas aproximaciones hacía el resultado final; el valor más aproximado (±0%) al resultado final (–2%) se ha obtenido, de hecho, tras la segunda eliminación y el resultado más divergente (–15%) tras la cuarta eliminación. En realidad, los sucesivos valores obtenidos se disponen irregularmente en torno al resultado final.

¿Cuál es entonces la conclusión exacta?

En realidad, todas son en cierto modo exactas y todas pueden ser en cierto modo erróneas. La divergencia depende del hecho de que la expresión «mortalidad» se entiende de diferentes formas. Podemos, de hecho, querer medir, mediante la mortalidad, la resistencia vital de la población considerada en su composición efectiva por edad, sexo, estado civil, profesión, poder adquisitivo, y el entorno en el que vive; para algunos fines es precisamente lo que se necesita. O quizá podamos querer medir su resistencia vital prescindiendo de la circunstancia eventual para cada individuo, de la edad; y aún así puede servir para ciertos fines. O aún podríamos querer medir la resistencia vital prescindiendo también del sexo y del estado civil, lo que puede ser plenamente justificado cuando, por ejemplo, se quiere hacer una previsión de la resistencia vital de las generaciones futuras. Por otra parte, puede interesar comparar la resistencia vital de las dos poblaciones prescindiendo de cualquier influencia social, eliminando por tanto incluso la influencia de la profesión y del poder adquisitivo; y puede finalmente ayudar a los fines de la investigación eliminando también la influencia del distinto entorno físico, por ejemplo cuando interese averiguar cuál de las dos poblaciones será más adecuada para colonizar un nuevo territorio cuyo clima sea intermedio de los de los dos países considerados.

Definir con precisión el objeto de la investigación es una necesidad no sólo de las investigaciones estadísticas, sino de cualquier género de investigación científica, y por tanto se puede decir que inconvenientes parecidos se encuentran en todas las disciplinas. —Si, pero ¿se verifican en la misma medida? Debemos reconocer que muy difícilmente encontraremos, en la definición de un fenómeno no estadístico, la multiplicidad de significados que hemos visto que se podían atribuir a la palabra «mortalidad», y que análogamente se podrían atribuir a las palabras «natalidad», «nupcialidad», «morbilidad», etc. Esto es porque, que los fenómenos colectivos de los que se ocupa la Estadística, dependen de una multiplicidad de factores, y que, en según qué casos, puede interesar eliminar alguno, o un grupo, o eliminarlos todos menos uno, o finalmente, no eliminar ninguno. La posibilidad de conclusiones contradictorias, derivadas de conceptos contradictorios del fenómeno analizado, se presenta por lo tanto en todas las investigaciones científicas, pero en la investigación estadística presenta particular importancia.

* * *

Por otra parte, hemos supuesto hasta ahora que disponemos de todos los datos necesarios para eliminar la influencia de los factores que nos interesaba excluir. En la práctica, las cosas a menudo son bastante diferentes.

A veces depende de la naturaleza misma del fenómeno que los datos sean insuficientes a tal fin, como cuando se trata de eliminar la influencia del clima. Debemos, por tanto, disponer de los datos sobre la mortalidad del país A, en años en los que el clima ha sido, excepcionalmente, parecido al que hay normalmente en el país B; y viceversa, los datos sobre la mortalidad del país B en años en los que el clima ha sido, excepcionalmente, parecido al que hay normalmente en el país A. Pero el sólo hecho de no poder verificarse sino con excepciones, hace difícil disponer de una cantidad de datos suficiente, debe recordarse que tratándose de fenómenos estadísticos, no sólo es necesario contar con los datos, sino también que sean lo suficientemente numerosos para neutralizar las circunstancias perturbadoras de carácter aleatorio.

A fin de eliminar otros factores, la naturaleza puede ofrecer una experiencia suficientemente amplia, pero las conclusiones estadísticas pueden ser inadecuadas. ¿Cuál es el país que permite clasificar los muertos y vivos según la combinación de edad, sexo, estado civil, profesión y riqueza, y poder eliminar al mismo tiempo la influencia de todos estos factores en el coeficiente de mortalidad?

Que yo sepa, ninguno. El estadístico debe en tal caso contentarse con eliminar la influencia de algunos de los factores que querría excluir, llegando así a conclusiones aproximadas. Esta es la regla.

Otras veces se da el caso de que el material estadístico permite la eliminación de determinados factores, pero solo bajo ciertas hipótesis, validas solo de forma aproximada, por ejemplo bajo la hipótesis de la superposición de los efectos. Es la hipótesis en que se basa el método de correlaciones parciales, a las que el material estadístico permite recurrir en algunos casos, cuando el procedimiento de la población tipo o de los coeficientes tipo no se pueden aplicar.

Otras veces, en fin, el material estadístico no permite eliminar todos los factores γ, δ, ε, sino sólo los dos factores γ y δ, o bien δ y ε, o bien γ y ε. Es arbitrario eliminar esta o aquella pareja de factores. En tal circunstancia, hay quien prefiere eliminar una y quien prefiere la otra. La conclusión será siempre aproximada, pero la aproximación podrá interpretarse en distinto sentido según se haya eliminado una u otra pareja. En el ejemplo considerado antes, si se eliminan edad, sexo, estado civil y profesión, la mortalidad resulta más elevada en el país B que en el A; si se elimina edad, sexo, estado civil y riqueza, resulta al contrario, más elevada en el país A que en el B. Ahora supongamos que no se cuenta con los datos necesarios para eliminar simultáneamente, además de edad, sexo y estado civil, el poder adquisitivo y la profesión. Quien elimine poder adquisitivo, llegará a distinta conclusión que quien elimine la profesión. Ante la imposibilidad de eliminar todos los factores de perturbación, depende de la intuición estadística del investigador que se elimine uno u otro. Este caso se presenta en muchísimas, diría incluso en casi todas, las investigaciones estadísticas, y esta es la razón fundamental de que la intuición tenga tanta importancia en la Estadística aplicada.

Es cierto que existe la posibilidad de eliminar, uno tras otro, todos los grupos de factores eliminables: por ejemplo primero los dos factores γ y δ, después los dos factores δ y ε, y finalmente los factores γ y ε. Esto mostraría que los datos disponibles no permiten llegar a una conclusión firme, pero evitaría que una conclusión, tenida por firme sea rebatida posteriormente. Así, tales investigaciones requieren una laboriosidad en la realización de los cálculos que entra a menudo en conflicto con la disponibilidad de tiempo y de medios.

En conclusión, la insuficiencia de datos cuya naturaleza misma conlleva y, no menos a menudo, lo inadecuado de las aportaciones estadísticas, suelen impedir a la Estadística alcanzar conclusiones seguras y, debido a las limitaciones de tiempo y de medios, dejando un amplio campo a la intuición personal que lleva inevitablemente a conclusiones contradictorias por parte de distintos investigadores. Es lo que se verifica en todos los campos de la investigación, en los que el resultado depende de una compleja multiplicidad de factores imposible de ordenar objetivamente; tal es el caso de la medicina, de las bellas artes, de las artes militares, en todas —se puede decir— las artes. La importancia del papel que se deja jugar a la intuición subjetiva es lo que diferencia las artes de las ciencias. La Estadística aplicada es en buena parte aún, y siempre será en cierto grado, un arte en el cual la intuición, o si os gusta más, el genio del investigador, tiene y tendrá siempre un papel importante. Habrá quien asuma esto como una ventaja, otros como una desventaja frente a otras disciplinas. Esto nos explica, de todas formas, muchas de las divergencias y contradicciones de sus estudiosos. Quien, situándose en un punto de vista exclusivamente científico, juzga tales caracteres como una desventaja, debe, por otra parte, tener presente que los resultados, incluso a veces aproximados e inciertos, obtenidos con la Estadística, no podrían obtenerse generalmente con otros métodos. La Estadística es una ciencia marginal, que llega donde, con los datos y medios disponibles, las otras no llegan. A veces las otras ciencias llegan después, cuando los datos de la experiencia han aumentado y los medios técnicos se han desarrollado. En tales casos, la Estadística tiene sobre todo una función de vanguardia, que siempre es peligrosa pero preciosa. Así ha sido, por ejemplo, la función de la Estadística en el estudio de la variabilidad y la herencia de caracteres. En otros campos, las otras ciencias no llegan más allá. La Estadística explota entonces terrenos que de otra forma no se explotarían. A veces se trata —es cierto— de campos de menor rendimiento, por ejemplo, en la economía o en la biología, los resultados de la Estadística parecen de menor importancia frente a los que se pueden obtener con otros métodos. Pero otras veces es a la Estadística a quien deben el reverdecer de la construcción científica, volviendo a ser apreciada como ha ocurrido en múltiples ramas de las ciencias físicas, biológicas y sociales. Sin la Estadística, a menudo se habrían tenido que limitar a mantener sus conclusiones en el campo cualitativo o contentarse con aproximaciones insuficientes.

* * *

Veamos otro ejemplo en el que la Estadística parece haber llevado a resultados contradictorios.

Clasificando en un mismo Estado las varias circunscripciones territoriales según los índices de prosperidad económica (renta, riqueza, ahorro), se comprueba que la natalidad resulta más baja en las zonas más prósperas. También la comparación entre Estados pobres y Estados ricos muestra generalmente una natalidad más baja en los Estados más ricos. Y la misma población presenta, a menudo, una natalidad más alta al principio de su historia que en los estadios sucesivos, cuando su organización económica se ha desarrollado y ha crecido la riqueza acumulada. En fin, en muchos países, las clases sociales elevadas presentan la natalidad más baja del total de la población. Se llega a la conclusión de que el bienestar económico constituye un factor desfavorable a la reproducción.

La comparación entre el número de nacidos en meses y años sucesivos tiene, por otra parte, demostrado que en los periodos de crisis económica la natalidad se contrae, y en los de prosperidad se expande. En una misma clase social, las familias con medios económicos mayores resultan, en general, más numerosas. Esto ha hecho llegar a otros investigadores a la conclusión opuesta de que el bienestar económico favorece la reproducción.

En fin, comparando el número de hijos de las familias de contribuyentes clasificados según la renta, resulta que tal número crece con el crecimiento de la renta, hasta un cierto punto, y después decrece, lo que ha sugerido la conclusión intermedia, que el bienestar económico hasta un cierto límite promueve y después frena la reproducción.

Divergencias parecidas dan muy buen juego a los detractores de la Estadística. Un análisis cuidadoso muestra sin embargo que, en realidad, éstas son en parte aparentes.

No es noticia para los estudiosos de los hechos biológicos la constatación de que las reacciones del organismo humano ante la acción momentánea de un factor pueden resultar opuestas a las que provocaría una acción prolongada. Así pasa con las reacciones provocadas por los estimulantes. Esto puede explicar, al menos hasta cierto punto, como, ante una variación de las condiciones económicas, la natalidad reaccione de forma distinta cuando se trata de oscilaciones mensuales o anuales o, a veces, modificaciones evolutivas a largo plazo o de diferencias permanentes que se verifican según el territorio o la clase social.

Por tanto, es común la observación de que, para muchísimos fenómenos físicos, biológicos y sociales, los efectos no resultan proporcionales a su intensidad y a veces, según el nivel inicial o según la intensidad de las variaciones, difieren también en la tendencia. Por lo que concierne a la temperatura, la alimentación, el ejercicio físico, no tenemos experiencia cotidiana. El principio de la superposición de los efectos, a menudo utilizado en las ciencias físicas y que, en Estadística está en la base —como recordaba— del método de las correlaciones parciales y, más en general, del método de los residuos, es generalmente admisible, como primera aproximación, pero sólo en breves intervalos de la variable. Por tanto, no hay que sorprenderse si las variaciones de la reproducción humana varían a veces en intensidad y tendencia al variar el bienestar económico. Podría, por ejemplo, explicar cómo, mientras se sube de clase social la natalidad disminuye, pero en las clases bajas el número de hijos resulta menos elevado en los estratos económicos ínfimos.

También en estos casos las contradicciones surgen del hecho de que, en realidad, las expresiones «bienestar económico» y «mejora económica» son expresiones vagas, que necesitan ser precisadas, bien en la duración del bienestar, bien en su nivel inicial y la intensidad de sus variaciones. Inconvenientes parecidos pueden obviamente verificarse, y se verifican, en todos los campos de la investigación científica y no son una peculiaridad de la Estadística.

* * *

Por otra parte, sin embargo, las consideraciones expuestas no explican completamente las llamativas contradicciones. Evidentemente, hay en éstas otro factor.

Cuando comparamos datos relativos a meses, años, circunscripciones territoriales, estados evolutivos, categoría de las rentas, clases sociales y, en la misma clase social, grados jerárquicos, que diferencian en lo que concierne al bienestar económico, se refieren a modalidades que en realidad no se diferencian sólo por este carácter. Meses y años de crisis se diferencian a menudo de meses y años de prosperidad, también por lo que respecta a las condiciones climáticas, y, generalmente por lo que respecta a la mortalidad. Los contribuyentes de las varias categorías de renta, al igual que los pertenecientes a los distintos grados jerárquicos de una misma clase social, se diferencian en lo que respecta a la edad. Circunscripciones territoriales, ricas y pobres, se diferencian generalmente en lo que respecta a la densidad de la población, el grado de urbanismo, la ocupación profesional. Las distintas clases sociales y la población que se encuentran en un estado evolutivo distinto se diferencian en múltiples aspectos psicológicos, por ejemplo en lo que se refiere al conocimiento de los medios anticonceptivos. ¿Hasta qué punto, las relaciones observadas entre variaciones de la natalidad y variaciones del bienestar económico según los meses, años, circunscripción territorial, clases sociales, grados jerárquicos, categoría de renta, estados evolutivos, dependen de las variaciones del bienestar económico, y hasta qué punto, al contrario, de otros factores diferenciales que quizá acentúan, quizá atenúan o neutralizan, quizá dominan la influencia de las variaciones del bienestar económico?

Dificultades similares se derivan del hecho de que los fenómenos que se examinan dependen de un complejo de variables numerosas y difícilmente aislables globalmente; si en ningún campo de la investigación científica están ausentes, son más frecuentes, más graves y menos fácilmente eliminables del campo de la investigación estadística, particularmente cuando se refiere a los fenómenos colectivos sociales o biológicos relativos a la especie humana, que no se pueden someter a experimentación.

Por lo tanto, es necesario reconocer que la Estadística, por la misma naturaleza de los fenómenos que trata, está más expuesta a incurrir en contradicciones de lo que lo están otras disciplinas. Se impone una particular cautela en su aplicación, pero no se invalida la exactitud del método.

* * *

Ya he indicado que frente a las contradicciones que se le reprochan a la Estadística y los desmentidos que a veces han provocado sus conclusiones, nosotros no podemos eludir el juzgar si los métodos utilizados generalmente por los estadísticos son ciertamente correctos. Y aquí entro en un campo más delicado, ya que se trata de una amenaza a la confianza en el conjunto de procedimientos a los que están ligados nombres entre los más autorizados de nuestra disciplina y que son considerados, por la generalidad de los estadísticos, como la última y más segura expresión del progreso científico.

Son dudas largamente maduradas. A este propósito llamamos vuestra atención sobre la siguiente circunstancia. Mientras la generalidad de los estadísticos no calculan la constante característica de una serie estadística, (media, proporción, índice de variabilidad, índice de concordancia o de conexión, etc.) sin determinar los errores probables, el error cuadrático medio o la probabilidad de que su error alcance cierta intensidad, y creería incluso que podría ser acusado de ignorar los procedimientos más seguros de la metodología estadística si no lo hiciese, es una minoría la que no calcula tal error más que en casos particulares y acompañado de restricciones y reservas. Lo más curioso es que en esta minoría, están al menos algunos especialistas que han dedicado profundos estudios a la dispersión. Me sitúo incondicionalmente entre estos descreídos. Entre los colegas que mantienen una actitud análoga, querría recordar al llorado Lucien March, del que sé que tenía un punto de vista análogo. Es una pena que Lucien March haya desaparecido sin haber expuesto, por lo que me consta, las motivaciones detalladas. Pero lo comprendo: se trata de argumentos tan sutiles en los cuales en realidad es difícil darle al propio pensamiento una forma definitiva y, en tales condiciones, asumir la responsabilidad de cuestionar lo que generalmente se ha considerado como fundamento de la credibilidad de los resultados estadísticos. Nosotros mismos hemos pasado por ese trance antes de manifestar estas dudas, surgidas al inicio de mis estudios y consolidadas progresivamente; pero, al final, creo que si no lo hiciera faltaría a mi deber de maestro hacia los discípulos y colegas que me honran con su confianza y a veces recurren a mí en busca de directrices para sus investigaciones.

* * *

Un primer elemento de incertidumbre en el cálculo del error que se puede atribuir, con una cierta probabilidad, a una constante estadística, o de la probabilidad de que ésta se vea afectada por un cierto error, es reconocido generalmente: proviene del hecho de que en dicho cálculo se sustituye el valor exacto —desconocido— de la constante, por el valor aproximado —más o menos exacto— que se conoce.

Desde un punto de vista estrictamente lógico, el círculo vicioso es evidente, se admite que no es afectada por el error una cantidad para la que se quiere determinar la probabilidad de cierto error; pero desde un punto de vista práctico, se observa que la sustitución no tiene importancia cuando el número de casos es elevado, y por otra parte se ha comprobado, que para las principales constantes, es susceptible de tenerse en cuenta aún cuando el número de casos es pequeño. Creo que a este propósito habría que hacer alguna observación.

Observación que creo importante, ya hecha otras veces a propósito de la inversión del teorema de Bernoulli[4] admitida generalmente, es la de que al sustituir el valor exacto por el valor observado de la constante, se supone que éste puede diferir de aquél, indiferentemente en un sentido u otro, y por tanto indepen-

[4] Cfr. Nuestras observaciones de las pag. 382-386 del artículo «Sobre las vacunaciones anti-tíficas en el ejército italiano durante la guerra», realizado en colaboración con el Dr. L de Berardinis y publicado en «Metron», Vol. III, n-3-4, 1-II-1924

dientemente del conocimiento del valor observado de la constante, pues no se tiene ninguna noción de su valor efectivo. Cuando esto no ocurra —y generalmente no ocurre— sólo se podrá determinar un límite, según los casos inferior o superior, de la probabilidad de que el valor observado de la constante sea afectado por un error aleatorio.

* * *

En muchas aplicaciones, más allá de sustituir el valor verdadero por el valor observado de la constante, se sustituye también el valor teórico por el valor empírico de su índice de variabilidad.

Esta sustitución no es necesaria para la media de magnitudes intensivas, que indican la frecuencia de los fenómenos, ya que, si $\boldsymbol{p}$ es la frecuencia media en cuestión, el valor de la desviación cuadrática media se calcula, como se sabe, mediante la fórmula

$$\sqrt{\frac{p(1-p)}{n}}$$

pero es generalmente necesario, para las medias de magnitudes extensivas (como la estatura, la renta, etc.), que expresan la cantidad de los fenómenos. De estas magnitudes, realmente se conoce sólo la desviación cuadrática media empírica, la cual incluye a menudo una componente sistemática, que generalmente no se elimina, y a menudo sería imposible eliminar. Esto tiende a dar una medida errónea por exceso de la probabilidad de que a la media le afecte un error mayor en un límite dado.

Se ha determinado, por ejemplo, la estatura media de los $\boldsymbol{n}$ reclutas de un país y se desea determinar la probabilidad de que ésta sea afectada de un error aleatorio de intensidad superior a $\boldsymbol{x}$. Esto se basa una propiedad de la curva de errores aleatorios, hay una relación constante entre la probabilidad de hallar un error de intensidad superior a $\boldsymbol{x}$ y el error cuadrático medio, y como error cuadrático medio se asume generalmente la diferencia cuadrática media σ' de la estatura obtenida de los $\boldsymbol{n}$ reclutas. Esta desviación cuadrática media no representa el error cuadrático medio aleatorio, ya que las estaturas observadas, además de por causas aleatorias, pueden diferir, y efectivamente difieren, por causas sistemáticas, como la clase social, la profesión, renta, factores hereditarios, etc. En realidad, deberemos basarnos en la desviación cuadrática media σ de la estatura de los individuos que no estuvieran sujetos a parecidas diferencias sistemáticas, desviación cuadrática media que constituiría el componente aleatorio de σ'. Sólo basándonos en σ podremos determinar la probabilidad, que buscamos, con la cual la estatura media de los ν individuos observados difiere, en una cantidad superior a $\boldsymbol{x}$, de la estatura media que se obtendría de un número infinito de individuos, los cuales se distanciarían de los individuos observados por meras diferencias aleatorios. Basándonos ahora en σ' obtenemos una probabilidad que también tiene sentido, pero un sentido diferente. Obtenemos la probabilidad con la cual la estatura media de los ν individuos observados difiere, por una cantidad superior a $\boldsymbol{x}$, de la estatura media que se obtendría de un número infinito de individuos, y que diferirá de los individuos observados, no sólo por diferencias aleatorios, sino también por diferencias sistemáticas de la misma importancia que aquéllas por las cuales se diferencian entre ellos los $\boldsymbol{n}$ individuos observados.

Para darse cuenta de la diversidad de los resultados, a los que se llega basándose en σ' o en σ conviene recurrir a un ejemplo relativo a una magnitud intensiva, en el cual σ' se ha determinado empíricamente y σ se puede determinar teóricamente.

Tomemos, por ejemplo, los datos de sexo, relativos a las 10.690 familias que han declarado el duodécimo hijo en Sajonia en el decenio 1876-85. La proporción de nacimiento de varones[5] resulta $\boldsymbol{p}_{12} = 0{,}51683$. Se quiere saber cuál es la probabilidad P_a de que $\boldsymbol{p}_{12}$ esté afectado por un error aleatorio igual o superior a ±0,001.

La desviación cuadrática media efectiva de los 10.690 informes de varones nacidos en las familias consideradas resulta σ' = 0,1523; su componente aleatorio, basado en la fórmula

$$\sqrt{\frac{p-(1-p)}{n}}$$

resulta σ = 0,1443. Basándonos en σ, el error cuadrático medio de $\boldsymbol{p}_{12}$ resulta

$$= \frac{0{,}1443}{\sqrt{10.690}} = 0{,}001395$$

de donde se obtiene un valor de P_a = 0,473, que puede considerarse valor correcto. Basándonos en σ', el error cuadrático medio de $\boldsymbol{p}_{12}$ resultaría

$$= \frac{0{,}1523}{\sqrt{10.690}} = 0{,}001473$$

de donde se comprueba P_a' = 0,496, que en realidad muestra la probabilidad de que el valor observado de $\boldsymbol{p}_{12}$ difiera por lo menos ±0,001 del valor que se habría

[5] «El sexo desde el punto de vista estadístico» —Biblioteca del Metron. Instituto de Estadística de la R. Universidad de Roma, 1908, pag. 384, y «Consideraciones sobre la Probabilidad a posteriori y aplicaciones al informe sobre sexos en los nacimientos humanos» en «Estudios económico-jurídicos» publicados a cargo de la Facultad de Jurisprudencia de la R. Universidad de Cagliari, Año III Volumen III, 1911, pag 111.

verificado, si en lugar de las 10.690 familias observadas se hubiera observado un número infinitamente grande de familias que, diferirían de la observada respecto a la tendencia a producir los dos sexos, no solo aleatoriamente, sino también sistemáticamente en la misma medida en que las observadas difieren entre ellas.

En este caso, la diferencia entre P_a y P_a' es poco relevante, porque la componente sistemática de la variabilidad de la proporción de los sexos en las familias es pequeña, sin embargo, la diferencia resulta muy notable en otros fenómenos, en los que la componente sistemática es fuerte. Consideremos, por ejemplo, la mortalidad infantil anual, estudiada por Lexis[6]. En Bélgica, en los 19 años de 1847-65 el coeficiente medio de mortalidad masculina anual resultaba ***p***=0,1711; para la desviación probable efectiva de los 19 coeficientes anuales de mortalidad, Lexis obtenía un valor = 0,008315, al cual contraponía el valor de la desviación probable teórica = 0,00095. A esto corresponde una desviación cuadrática media efectiva σ' = 0,00141.
Basándonos en σ, el error cuadrático medio de $\boldsymbol{p}_{12}$ resultaría

$$= \frac{0,00141}{\sqrt{19}} = 0,000323$$

de donde se obtiene la probabilidad P_a = 0,002 de que el valor verdadero de ***p*** no esté comprendido entre los límites 0,1721 y 0,1701. Basándose en σ', el error cuadrático medio sería entonces

$$= \frac{0,01236}{\sqrt{19}} = 0,00284$$

de donde se obtendría una probabilidad P_a' = 0,724 (362 veces mayor que P_a) de que ***p*** no estuviera comprendido entre estos límites. En realidad P_a' señala la probabilidad de que el valor de ***p*** no esté dentro de los límites para un número infinito de nacidos, cuya probabilidad de muerte difiera sólo aleatoriamente de la probabilidad de muerte de los nacidos en los 19 años considerados, mientras P_a señala la posibilidad de que el valor de ***p*** no esté dentro de los límites citados para un número infinito de nacidos cuya probabilidad de muerte difiere, además de aleatoriamente, también sistemáticamente, de la de los nacidos en el periodo 1847-65, como se diferencian también sistemáticamente entre ellos los coeficientes anuales de mortalidad que se han verificado en dicho periodo.

En conclusión, basándose en la desviación cuadrática media empírica σ', no se puede calcular sino el límite superior de la probabilidad de que el error aleatorio de la desviación alcance o exceda cierto límite. Esto es ciertamente mejor

[6] W. Lexis, Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft, Freiburg i.B., Wagner, 1877, pags. 79-81

que nada; ya que el límite superior puede estar tan alejado del verdadero valor que la utilidad práctica de su determinación resulta muy escasa.

* * *

Prescindamos de las ya mencionadas objeciones y admitamos conocer la probabilidad P_a de que el error aleatorio de una constante exceda un cierto límite y por tanto la probabilidad 1- P_a de que quede entre dichos límites. Si pudiéramos decir que el verdadero valor de la constante queda, con cierta probabilidad, en un cierto entorno, ya sabríamos algo. Pero ¿podemos también decir que, si el error de la constante en cuestión excede tales límites esto supone que hay una probabilidad P_a de que sea puramente aleatoria y por tanto una probabilidad 1 - P_a de que sea significativa?

Es lo que generalmente se hace. Tenemos una probabilidad contra 22 de que la intensidad del error aleatorio exceda el doble, y una contra 370 de que exceda el triple del error cuadrático medio; por tanto se concluye: un error que exceda al doble del error cuadrático medio tiene una probabilidad de 1 sobre 22 de ser aleatorio y 21 contra 1 de ser significativo, y un error que exceda el triple del error cuadrático medio tiene una probabilidad de 1 sobre 370 de ser aleatorio y por tanto, 369 contra 1 de ser significativo[7].

Aquí está el fundamento de toda la teoría de los mencionados «test de significación» dirigidos a medir el grado de credibilidad que se debe atribuir a un resultado estadístico o a la diferencia entre dos resultados estadísticos. Es la teoría la que ha marcado, en el pensamiento de la mayor parte de los estadísticos modernos, el remate del edificio de la metodología estadística.

Ahora, o me equivoco —pero más bien me parece que no— o en tal procedimiento se comete un grosero error lógico, cambiando lo que en Estadística se llama proporción de derivación por lo que llamamos proporción de composición.

Un razonamiento análogo en otro campo llevará a poner luz sobre la falacia del procedimiento. Supongamos que hemos establecido que, los tramposos que

[7] Citaré, por todos, dos tratados merecidamente muy difundidos: los de G. U. Yule y de R. A. Fisher. Se lee en el primero: «...the great majority of [the] members [of the binomial distribution] lie within a range 3 σ on each side of the mean, i. e. of $\pm 3\sqrt{npq}$ on each side of the value np. If the distribution is exactly normal, 0,9973 of the curve lies within this range. We can therefore say that, if a particular sample gives a value of p outside this range, the deviation from the expected value is most unlikely to have arisen from fluctuations of simple sampling. If n is large the chances are about three in a thousand that it arose in that way» (An introduction to the Theory os Statistics, por G.U. Yule y M. G. Kendall, 11.ª edición, Londres, Griffin, 1937, pag. 352.) Y más en general, Fisher escribe: «....Twice the standard deviation is exceeded only about once in 22 trials, thrice the standard deviation only once in 370 trials, while… to exceeded the standard deviation sixfold would need nearly a thousand million trials It is convenient to take [the] point [for which P=0.05 or 1 in 20] as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant. Using this criterion, we would be led to follow up a false indication only once in 22 trials, even if the statistics are the only guide available» (Statistical Methods for Research Workers, Cuarta Edición, Oliver and Boyd, Edimburgo, 1932, 1932, pag 45)

circulan por Roma son el 1,80% en una fecha concreta, estimado, entre determinado público local, por ejemplo de Aragno. Usted va a esa hora a Aragno y a la primera persona que se tropiece, dígale en nombre del cálculo de probabilidades: hay un 1,80% de probabilidad de que usted sea un tramposo. Lo absurdo de tal afirmación es evidente. Para calcular la probabilidad de que una persona encontrada por casualidad en Aragno, a esa hora, sea un tramposo, deberían ustedes conocer no sólo la probabilidad (que hemos supuesto del 1,80%) de que a esa hora se encuentre un tramposo en Aragno, sino también la probabilidad (que podríamos suponer, por ejemplo, del 2%) de que a esa hora se encuentre una persona que no sea un tramposo, y, finalmente, el cociente entre la frecuencia de tramposos y la población de Roma (que podemos suponer de 1 sobre 10.000). La probabilidad de que a esa hora en Aragno una persona con la que te tropiezas sea un tramposo, se podría entonces determinar y sería de 80 contra 20.080 y no de 80 contra 20. ¡Una diferencia notable, como veis! Si no se conocen, por una parte, la probabilidad de que una persona que no sea un tramposo se encuentre a esa hora en Aragno, y por otra, la proporción de tramposos en Roma, la probabilidad de que un tramposo se encuentre en Aragno no nos dice nada sobre la probabilidad de que una persona que se encuentra en Aragno sea un tramposo.

Idéntico es el razonamiento para la probabilidad π_α de que un error de cierta intensidad (o superior a cierta intensidad) sea aleatorio. Para determinarlo es necesario que conozcamos no sólo la probabilidad P_a, de que un error aleatorio tenga cierta intensidad en aquella fecha (o sea superior a la intensidad considerada), sino también que conozcamos la probabilidad $\boldsymbol{p_s}$, de que un error no sea aleatorio, y al mismo tiempo, la probabilidad P_s, de que un error, que no sea puramente aleatorio, presente la intensidad considerada (o sea superior a la intensidad considerada). En ese caso, la probabilidad de que el error de existente de tal intensidad (o una intensidad superior al límite considerado) sea aleatorio, se obtiene con la fórmula

$$\pi_a = \frac{P_a}{P_a + fP_a}$$

Donde

$$f = \frac{p_s}{1 - p_s}$$

Pero si no contamos con esos conocimientos —y, en la práctica, salvo casos particulares, no los tenemos— es absurdo pretender calcular π_a, y es un grave error cambiar P_a, razón de derivación, por π_a razón de composición. El hecho es que generalmente eso se hace con el aval de las enseñanzas, concordes y repetidas, de los más importantes estudiosos del Cálculo de probabilidades y la Estadística, sin confirmar, por una parte, la posibilidad de que también las mentes más selectas caigan en groseros malentendidos y, por otra, el terrible servilismo de la mente humana al aplicar sin crítica las enseñanzas recibidas.

Hay que decir que la mente humana se divide a veces en compartimentos estancos, ya que no es creíble naturalmente que todos aquellos que han introducido o aconsejado o aplicado tales métodos ignorasen la diferencia que hay entre los cocientes de derivación y los de composición. ¡Quién sabe cuántos habrán llamado la atención, en sus escritos o sus enseñanzas, poniéndonos en guardia, por ejemplo, contra los errores que se cometerían deduciendo la probabilidad de muerte de un niño de la frecuencia con que los niños figuran entre los muertos! ¡Y pocas páginas o pocas lecciones después, si no es en la misma página o en la misma lección, ellos habrán cometido, en el sentido inverso, el mismo error!

Vendrá bien ilustrar la diferencia entre P_a y π_a, mediante un ejemplo concreto que intentaremos construir de forma que eliminemos la dificultad descrita en los párrafos precedentes.

Supongamos que elegimos, en una gran finca, dos parcelas de terreno y debemos determinar el rendimiento medio de cada uno durante n años, diferentes para las dos parcelas. Entre el rendimiento medio de las dos parcelas se da cierta diferencia. Se plantea el problema de determinar la probabilidad de que tal diferencia sea significativa y obedezca a una fertilidad diferente del terreno y no dependa, al contrario, de factores variables de año en año, que respecto a la fertilidad del terreno, se podrían considerar como aleatorios.

Supongamos que, sobre la base del conocimiento de los rendimientos de cada parcela en años singulares, sea posible determinar la variabilidad debida a los factores aleatorios respecto a la fertilidad del terreno y calcular la probabilidad P_a de que la diferencia aleatoria entre los rendimientos medios en ν años de las dos parcelas muestre la intensidad observada. Sea $P_a = 0{,}9$. Este valor de P_a nos dice que sería de esperar que 9 veces sobre 10 la diferencia entre el rendimiento medio de n años, relativos a la misma parcela de terreno (que se supone mantenida en las mismas condiciones de fertilidad), igualase o superase la intensidad en cuestión.

La cantidad $P_a = 0{,}9$ expresa por tanto la relación probable de frecuencia entre dos clases de diferencias (unas inferiores y las otras superiores a un límite dado), correspondientes a la misma parcela y aleatorias, respecto a la fertilidad del terreno. Nadie acomete la acción de expresar la probable relación de frecuencia entre dos categorías de diferencias (unas dependen de la fertilidad del terreno y las otras son aleatorias respecto la fertilidad), correspondientes a dos parcelas diferentes, elegidas al azar y ambas superiores al límite considerado. Se pretende de hecho que, en paridad con $P_{a,}$ esta segunda relación debe resultar diferente según que sea más o menos probable que las dos parcelas escogidas al azar tengan diferente fertilidad, o sea, según que sea más o menos alto el valor de p_s, y por tanto el de $f = \frac{p_s}{1 - p_s}$. Se pretende que, en paridad con P_a, y f, la segunda relación debe resultar distinta, según que sea más o menos probable que la diferencia de fertilidad, eventualmente existente entre las dos parcelas consideradas, iguales o superiores a un límite dado, según que sea más o menos alto el valor de P_a. La cantidad $\boldsymbol{p_a}$, que expresa la segunda relación, no depende por tanto sólo de P_a, sino también de $\boldsymbol{f}$ y de P_s.

Pero en tales casos, ¿tampoco podremos decir nada sobre la mayor o menor probabilidad de que cierto error o cierta diferencia sean aleatorios o por el contrario significativos?

Si no conocemos la frecuencia de tramposos en Roma y la probabilidad de que una persona que no sea tramposo se encuentre a esa hora en Aragno, todo lo que podríamos decir (basándonos en el conocimiento de que los tramposos frecuentan los locales públicos más que los no tramposos) es que hay mayor probabilidad de encontrar un tramposo en una persona que te encuentras por casualidad en Aragno que en otra persona que esté fuera de Aragno.

Si no conocemos, como generalmente no se conoce, la probabilidad de que un error sea puramente aleatorio o, por el contrario, sea en parte sistemático, y al mismo tiempo, la curva de probabilidad de los errores en parte sistemáticos, lo único que podríamos decir (basándonos en la idea de que, a falta de compensación entre componentes aleatorios y componentes sistemáticos del error, los errores puramente aleatorios son en general menos intensos que los que tienen también un componente sistemático) es que los errores más pequeños serán puramente aleatorios con más frecuencia que los errores mayores. En el ejemplo anterior de las dos parcelas, podemos decir que cuanto mayor es la diferencia entre sus rendimientos medios y altos, es más probable que la fertilidad de ambos terrenos sea diferente. No es serio: muchos encontrarán que para llegar a tan obvia conclusión no era necesario poner en marcha el cálculo de probabilidades. De cualquier modo, está claro que no es necesario para conocer el valor de $P_{a.}$ Pero yo, en realidad, no veo qué más se podría decir. Esta misma conclusión, no es aceptable sino con ciertas restricciones, de las que hablaremos más adelante.

Pero ¿la teoría de la dispersión —se preguntará,— no dice nada sobre esta cuestión? No —se responde—, no nos dice nada. Recoge, no una sola constante o índice o dato estadístico, sino una serie entera de constantes o índices o datos, la teoría de la dispersión nos enseña a calcular si estas cantidades se diferencian entre sí más o menos de lo que deberían diferenciarse por efecto de los puros errores aleatorios y por tanto tiende a separar, en la dispersión en conjunto, la parte aleatoria y la parte sistemática; pero no nos dice cuál es la probabilidad de que una diferencia sea también sistemática, ni cuál es la distribución de la intensidad de las diferencias que son también sistemáticas, así como no permite deducir cuál es la probabilidad de que una de las diferencias encontradas tenga carácter puramente aleatorio.

En el ejemplo citado, será ciertamente interesante conocer, para un número adecuado de años, el rendimiento de todas las parcelas de la propiedad, y comparar el índice de dispersión efectivo con el que habría tenido lugar si la fertilidad del terreno fuera la misma en todas las parcelas. Esa comparación nos dará la medida de la heterogeneidad de la finca, en lo que se refiere a la fertilidad del terreno. Pero una cierta diferencia entre el índice de dispersión efectivo y el teórico, podrá corresponder a situaciones muy distintas. Podrá ser que la diferencia se deba a la circunstancia de que la mitad de las parcelas sean un poco más fértiles y la otra mitad un poco menos, y por el contrario, podrá ser también que dependa del hecho de que las parcelas tengan en su gran mayoría la misma fertilidad, menos algunas que son casi estériles. Está claro que la probabilidad $\boldsymbol{p_s}$, de que dos parcelas elegidas a propósito no tengan la misma fertilidad, sería muy distinta en los dos casos; y también está claro que de igual modo sería diferente la probabilidad P_a, de que dos parcelas elegidas al azar no tengan la misma fertilidad, su rendimiento medio para algunos años se diferenciará más de una cantidad determinada.

La teoría de la dispersión está fundamentada y es fructífera. Pero no hay que pedirle más de lo que puede dar.

* * *

Volvamos a la usual determinación del error cuadrático medio (o del error probable) de las constantes estadísticas para alertar contra otro peligro al que se expone. No sólo no está en situación de afirmar qué probabilidad tiene un cierto error de ser aleatorio y cuál de ser significativo, sino que corre el riesgo de tomar por significativos errores que son puramente aleatorios.

Supongamos haber determinado la media de una cincuentena de caracteres de dos poblaciones, o bien la media de un carácter en 50 circunscripciones territoriales de un Estado. Encontramos que la diferencia entre las medias obtenida para dos caracteres A y B, en las dos poblaciones, o bien la diferencia entre la media general del Estado y la media de los caracteres en las dos circunscripciones α y β exceden el doble de la desviación cuadrática media, y por tanto concluimos, según la que viene considerándose como buena norma del método estadístico, que se trata de diferencias que se pueden considerar significativas. El antropólogo, el demógrafo o el sociólogo, según que se trate de caracteres antropológicos, demográficos o sociales, tendrán acceso a investigar las causas sistemáticas de esas diferencias en cuestiones raciales, económicas, profesionales, o de otro género. Es de esperar 2 veces sobre 50 (un poco más de 2 veces, contando con que la distribución de las medias sea normal) que las diferencias superen el error cuadrático medio por puro efecto del azar. Si consideramos, no las medias de los caracteres singulares o de las circunscripciones singulares tomados aisladamente, como se hace en la acostumbrada determinación del error probable de las constantes, sino la serie completa de dichas medias, como se hace en la teoría de la dispersión, deberemos concluir correctamente que las diferencias encontradas se pueden atribuir al efecto del azar y sería perseguir a un fantasma ir detrás de las causas sistemáticas de tales diferencias.

«En el fondo —se dirán— esta es una verdad obvia. Cuando se actúa basándose en una probabilidad, se corre inevitablemente un riesgo. Si tomamos como aleatoria una diferencia que sabemos que se verifica aleatoriamente solo 48 veces de 50, sabemos ya que nos exponemos a equivocarnos 2 veces de cada 50». Esto es exacto; pero son precisamente las verdades obvias las que a veces se olvidan más fácilmente. Por otra parte, si en la vida, que es acción, está justificado asumir riesgos, el no asumirlos en la ciencia conduciría a la inmovilidad. En la ciencia, que es la búsqueda de la verdad, puede parecer discutible que sea preferible una conclusión afectada de riesgo a una prudente reserva.

¡Cuántas diferencias en los resultados estadísticos, tomados como significativas, fueron contradichas por investigaciones posteriores! La causa puede atribuirse muchas veces al riesgo inevitablemente emparejado a la aplicación de los errores probables a las simples constantes estadísticas.

* * *

El peligro de que, aplicando a los casos sencillos el cálculo de la probabilidad de un error, seamos, sin fundamento, inducidos a negar el carácter aleatorio de la diferencia, cuando ésta supera un cierto múltiplo del error cuadrático medio, o, en general, un valor que cabría de esperar sólo excepcionalmente, es mucho más grave cuando el ejemplo que se considera no se ha elegido al azar, sino dependiendo de su propio carácter excepcional. Es lo que se hace generalmente cuando se estudia la herencia en base a la casuística. Por ejemplo, buscamos las familias en las cuales la herencia de cierta enfermedad se muestra de forma más marcada, presentándose la enfermedad además de en los padres, en todos los hijos, y después decimos «No podéis negar que esta enfermedad es hereditaria. Considerad, que en el conjunto de la población, dicha enfermedad se presenta pongamos en el 1% de los casos; la probabilidad de que se presente en todos los $\boldsymbol{n}$ hijos de una familia, como se desprende de nuestra investigación es por tanto de $(1/100)^n$; en una familia, por ejemplo, de cuatro hijos, apostamos 100 millones contra 1 que no se trata del azar.

Aplicamos este criterio a fin de decidir si un juego se desarrolla con honestidad. Pasando el otro día por delante de un puesto de lotería, noté que, de uno de los bombos habían salido, uno tras otro, en la segunda y tercera extracción, dos números sucesivos; creo que eran el 63 y el 64. Veamos cómo, en casos similares, razonaría uno de los tantos recolectores de casuística que disfrutan al someterlos a la prueba del cálculo de probabilidades. «Había —dirán— una probabilidad de 1/89 de que, en la segunda extracción, saliera el 63 y de 1/88 de que en la tercera saliese el 64; por tanto una probabilidad de $1/89 \times 1/88$ (se puede decir cerca de 1/7800) de que los dos números salieran seguidos. Se podría por tanto apostar 7800 contra 1 a que no se trata de un efecto del azar. Pero antes de sospechar de nadie, veamos los resultados de otras extracciones ocurridas en el pasado.» Busca, busca, y al final encuentras. ¿Qué encuentras? Encuentras que en un año, en un bombo, en una extracción, los números que han salido son 1, 11, 21, 31, 41. «Aquí la estafa es evidente» —piensan— es para apostar 1 contra 5.000 millones a que no se trata del efecto del azar, porque sólo hay una posibilidad sobre 5.000 millones de que ese resultado tenga lugar». Nuestro probabilista no duda más, y va a reclamar al Director General de Loterías para que lleve a los tribunales al responsable de las extracciones, ordene una investigación severa, tome las medidas para asegurarse la honestidad del juego. El Director General de Loterías, perplejo, llama a otro probabilista, un super-probabilista, si se puede llamar así. «De ningún modo —responde éste— ¿por qué?» La secuencia 63, 64 tiene —ciertamente— una probabilidad ocurrir sobre 7800 y la secuencia 1, 11, 21, 31, 41 tiene una probabilidad sobre 5.000 millones, asumiendo que los 90 números tienen la misma probabilidad de salir. ¿Pero por qué? Esto significa que a veces puede ocurrir y que por tanto es natural que a veces ocurra. Me diréis que en todos los bombos del Reino, desde que existe el juego de la Lotería, no se han realizado aún 5.000 millones de extracciones. Pero tengan presente que cualquier secuencia de dos números diferentes, no solo la de los números 63, 64, tiene una probabilidad sobre 7800 y que análogamente cualquier secuencia de cinco números diferentes, no sólo la secuencia 1, 11, 21, 31, 41 tiene una probabilidad sobre 5.000 millones de verificarse. Alguna secuencia concreta debe verificarse. No deben fijar su atención en una extracción especial y con una secuencia determinada; deben considerar todas las extracciones realizadas en todas las secuencias posibles. Si descubren que la distribución de todas las frecuencias posibles difiere sistemáticamente de la que haría esperar el cálculo de probabilidades, volved y hablaremos». El probabilista agacha la cabeza y reconoce su error.

Así, deberán reconocer su equivocación todos los probabilistas que eligen, entre las fichas clínicas o en las genealogías, ciertos casos característicos de la herencia de caracteres. Han pretendido después calcular probabilidades con lo que podrían descartar que se tratase de una coincidencia casual.

Sería un indicio de que no se trata de coincidencia casual, sino que el carácter realmente se hereda, que la combinación característica considerada se verificara con más frecuencia que la que cabría esperar, según el cálculo de probabilidades, por puro efecto del azar. Cuántas veces habremos oído hacer esta observación: Mirad esa pareja: 5 hijos y los 5 varones; y la otra, 6 hijos, todos chicas. ¿Qué tendrán en el cuerpo para no producir más que hijos del mismo sexo? En realidad, muchas veces no tienen nada especial: la uniformidad del sexo de sus hijos es el efecto del azar. Otras veces, en cambio, no es así: hay, en algunas parejas, una tendencia más o menos marcada, a producir un sexo. En cuántos casos, no lo podemos decir, y tampoco podemos decir con cuánta probabilidad la uniformidad de sexo de una prole sea efecto del azar más que de una tendencia sistemática; pero podemos decir plausiblemente que del 10 al 15% de la desviación cuadrática media, que se verifica en la composición por sexo de las familias, depende de la tendencia sistemática de las parejas a producir varones o hembras; el resto depende en cambio del azar[8].

* * *

[8] Cfr. *El sexo desde el punto de vista estadístico*. Op. Cit., Cap. X : La variabilidad individual en la tendencia a producir los dos sexos. Pag. 371 y siguientes.

Cuando se actúa, no con caracteres discretos (como los números de la Lotería o las combinaciones de sexos en proles aisladas), sino con caracteres continuos o que deben prácticamente tratarse como tales (como la estatura o el números de individuos de cada sexo en los nacimientos de una población total) la determinación de la probabilidad de que en un caso aislado se presente cierto error aleatorio supone otra dificultad y expone a otro peligro al que difícilmente podemos sustraernos. La probabilidad ${}_xP_a$ de que la determinación del carácter sea afectada por un error aleatorio de intensidad $\boldsymbol{x}$, es, en tal caso, infinitesimal. Se puede, sin embargo, calcular la probabilidad $\geq {}_xP_a$ de que se presente un error de intensidad igual o superior a $\boldsymbol{x}$. Queriendo y no pudiendo calcular ${}_xP_a$, se calcula en su lugar $\geq {}_xP_a$, que es manifiestamente otra cosa. En el ejemplo de las dos parcelas, yo mismo he hecho esa sustitución —probablemente sin que ninguno de vosotros lo haya advertido—, la sustitución es algo acostumbrado. A menudo además, como se ha dicho, la probabilidad $\geq {}_xP_a$, se puede determinar sólo aproximadamente, sustituyendo el valor real de la constante por su valor observado y el valor teórico de su error cuadrático medio, por el respectivo valor empírico que incluye una componente sistemática. Generalmente, en fin, se sostiene haber calculado la probabilidad de tal modo que el error hallado $\boldsymbol{x}$ sea aleatorio. Así se cometen cuatro errores al tiempo: a) sustitución de $\boldsymbol{p_a}$ por ${}_xP_a$; b) sustitución de ${}_xP_a$ por $\geq {}_xP_a$,; c) determinación inexacta de$\geq {}_xP_a$,; d) aplicación de $\geq {}_xP_a$, a un caso individual, a menudo elegido a la vista de su carácter excepcional.

He explicado antes el equívoco del que nace el primero y las deficiencias de los datos disponibles que suelen hacer inevitable el tercer error y hace un momento he llamado la atención sobre las circunstancias de las que surge el segundo; pero es sobre todo en el cuarto en el que pretendo insistir como conclusión de los párrafos precedentes.

La conclusión es que el Cálculo de probabilidades no se puede aplicar correctamente en casos aislados, sino sólo en conjuntos de casos. Se lanzó al Cálculo de probabilidades un mensaje que quizá quiso ser una crítica feroz. «El Cálculo de probabilidades —ha dicho Adolfo Thiers— es el arte de precisar lo que se ignora». Que yo sepa, la crítica, frecuentemente difundida, no ha encontrado respuesta. Retomémosla y respondamos que el Cálculo de probabilidades enseña a precisar en la masa lo que se ignora en los casos aislados. Esto se puede hacer y se hace con ayuda del Cálculo de probabilidades. En otros casos, esto se hace, quizá con un grado de precisión menor, sin su ayuda. Con ayuda del Cálculo de probabilidades, el físico, ignorando el movimiento de las moléculas, determina las propiedades resultantes de los cuerpos, y el demógrafo, ignorando el sexo de los *nascituros* concretos, determina la probabilidad de que en el conjunto de nacidos, los varones excedan en cierta medida a las hembras, y el asegurador, ignorando la duración de la vida de personas aseguradas concretas, prevé la duración media. Sin el apoyo del Cálculo de probabilidades, el Ministerio de Finanzas, ignorando la actividad económica de los contribuyentes concretos, prevé el impuesto sobre las rentas, y el Jefe de Gobierno, ignorando los sentimientos individuales de cada ciudadano, prevé la orientación política de la nación. Pero si el Cálculo de probabilidades pretende enseñar a la Estadística, o la Estadística elude aprender del Cálculo de probabilidades, el arte de precisar para los casos aislados lo que para los casos aislados se ignora, Estadística y Cálculo de probabilidades se engañan y no pueden resultar sino fuente de errores y desilusiones.

* * *

«Después de todo, todo lo sólido permanece —diréis— mantendremos la aplicación del Cálculo de probabilidades a la masa de casos mediante la teoría de la dispersión».

Pero las reservas no se habrán despejado. Hay otro punto sobre el que conviene llamar la atención. Y es que cada conjunto de casos puede a su vez considerarse como un caso aislado de orden superior. Una proporción estadística o una media representa un caso aislado entre la serie de proporciones o de medias a los que se aplica la teoría de la dispersión; pero la serie a su vez se convierte en caso aislado por la aplicación de los métodos «χ^2» ó «z».

Comparando la distribución efectiva de una serie de proporciones estadísticas y de medias con la respectiva distribución teórica, se podrá decir si las proporciones y las medias de la serie presentan una dispersión superior a la normal, normal o inferior a la normal, pero no se podrá determinar mediante el método χ^2 ó z, cuál es la probabilidad de que la divergencia entre la distribución efectiva y la teórica supere en una serie el doble del error cuadrático medio, u otro límite convencional, se podrá concluir que las proporciones y las medias de la serie en cuestión difieren entre sí sistemáticamente, lo que, por otra parte, sería manifiesto si, habiendo aplicado χ^2 ó z a muchas series, se hubiera encontrado que los valores obtenidos para χ^2 o para z, no exceden al doble del error cuadrático medio, u otro límite convenido, con más frecuencia de la que cabría esperar por efecto del azar.

Por otra parte, obtenidos así toda una serie de valores de χ^2 o de z, esta serie viene a ser a su vez un caso aislado a fin de determinar si su dispersión es menor o superior a la teórica; se podrá establecer si su dispersión resulta, en el caso considerado, superior a la normal, normal o inferior a la normal, pero no se podrá determinar la probabilidad con que, la divergencia entre la distribución efectiva hallada y la teórica, debe atribuirse al azar.

Y aquí me veo en la obligación de aclarar una aparente contradicción, por la cual se podría creer que estoy acabado. Ya que, por una parte, he afirmado que la teoría de la dispersión es correcta y puede llevarnos a determinar la componente sistemática de la dispersión y, por otra parte, he mostrado cómo no estamos en situación de establecer con qué probabilidad la dispersión hallada pueda considerarse significativa o aleatoria.

Esa cuestión no es más que un caso particular de aquélla que consideramos bajo enunciado general para todas las constantes. De hecho, aparte de la dificultad de determinar la probabilidad de que, por efecto del azar, se presente cierta dispersión cuando ésta sea medida mediante un índice que varía con continuidad, y los peligros de aplicar tal probabilidad a un caso aislado, debemos tener presente que no basta conocer tal probabilidad para conocer la probabilidad de que la dispersión hallada pueda considerarse como aleatoria. Haría falta conocer también —y generalmente no lo sabemos— la probabilidad de que la serie de datos considerada sea perturbada por causas sistemáticas, y la probabilidad de que, siéndolo, la dispersión observada se verifique por efecto de dichas causas también sistemáticas.

En general, podemos decir que la Estadística enseña a determinar muchos índices; medias, proporciones, índices de variabilidad, de concentración, de dispersión, de transvariación, de desemejanza, de conexión, de concordancia, etc., todos relativos a conjuntos de casos. Ahora bien, en cuanto el conjunto de casos sobre los que se opera interesen, no por sí mismos, sino como muestra de conjuntos mayores, cada índice debe tomarse como aproximado. Todos los índices están, ciertamente, afectados por errores aleatorios, que para ciertos fines interesaría eliminar; puede darse en muchos casos la eventualidad de que sean afectados también por errores sistemáticos. Aparte del caso en que tal eventualidad resulte excluida *a priori* o *a posteriori*, en cuyo caso no hay cuestión, estamos ante la imposibilidad, salvo conocimientos particulares derivados de otras fuentes, de determinar la probabilidad de que la diferencia hallada entre dos índices, o entre uno de ellos y su valor teórico, sea puramente aleatorio. A menudo —como ya he dicho— todo lo que podemos afirmar es que, cuanto más pequeña es la diferencia, es más probable que ésta sea aleatoria, y cuanto más grande, es más probable que en ella se incluya una componente sistemática.

También para poder afirmar esto, es necesario que la comparación se haga entre diferencias relativas a datos de la misma serie o series relativas al mismo fenómeno y de cada forma comparable desde el punto de vista adoptado. Así, en una serie de coeficientes anuales de mortalidad, estaremos autorizados a decir que las ligeras desviaciones de la media que se verifican algunos años tienen carácter aleatorio más probablemente que las fuertes desviaciones que se verifican en otros; pero si en una serie de coeficientes anuales de mortalidad, un año dado la desviación permanece entre los límites del error cuadrático medio teórico de la serie, y en una serie razones entre sexos en los de nacimientos, la desviación, un año dado, excede por el contrario, el límite del error cuadrático medio teórico de la serie, sería arbitrario afirmar que la primera desviación tiene mayor probabilidad que la segunda de ser puramente aleatoria. Volvamos, para ilustrar este punto, a nuestra fórmula

$$\pi_a = \frac{P_a}{P_a + fP_s}$$

Cuando se comparan datos relativos a la misma serie, o series relativas al mismo fenómeno, es de resaltar que el valor de ***f*** permanece constante y que, con

ISSN: 1139-5737

el aumento del error, decrece P_a y, con menor intensidad, decrece P_s. Por tanto, podemos decir que π_a decrece al decrecer de P_s, y con el crecimiento del error, sin, por otra parte, poder decir nada, ni sobre la proporción entre la disminución de π_a y el aumento del error, y aún menos sobre la magnitud absoluta de π_a. Cuando, por el contrario, se trata de datos relativos a fenómenos distintos, podremos decir que P_a resultará más bajo para el dato relativo —pongamos— al fenómeno para el cual el error resulta muy leve, que para el relativo al fenómeno β. Pero, por otra parte, ***f*** puede ser, y normalmente es, distinto para los dos fenómenos; no se descarta que, si P_a resulta más bajo para el dato relativo a , también P_s resulta por ésto más bajo en la misma proporción o en proporción mayor. De manera que no se puede excluir que π_a resulte igual o incluso más bajo para el dato relativo a β que para el relativo a .

Y ¿podremos decir algo de la variación de π_a al variar el número de observaciones?

P_a, como hemos visto, decrece al crecer el número de observaciones; precisamente la intensidad del error que tiene una cierta probabilidad P_a de ser aleatorio, varía en razón inversa a la raíz cuadrada del número de observaciones. Pero, ¿cómo varían con el aumento de observaciones, ***f*** y P_s? No se puede hacer, a este propósito, una afirmación de carácter general. En algunos casos, la componente sistemática del error es constante, y por tanto independiente del número de observaciones. Pero, en otros, ésta decrece al crecer el número de observaciones, como cuando depende de la inexperiencia inicial del investigador, o del sujeto en observación, que se corrige sobre el terreno. En otros más, ésta aumenta, como cuando depende del progresivo cansancio del investigador, o del sujeto en observación, o incluso del hecho de que el fenómeno observado (por ejemplo, estatura de un individuo, o posición de un cuerpo) varía con el tiempo. De forma similar, ***f*** puede ser constante o, por el contrario, aumentar o disminuir.

En consecuencia, ninguna afirmación de carácter general está autorizada sobre la variación de π_a con la variación del número de observaciones, mientras que se pueden hacer afirmaciones fundadas, más o menos seguras, para casos particulares basándose en el conocimiento de la estructura del fenómeno y de de las condiciones en las que se desarrolla la investigación o el experimento. Por ejemplo, si tales conocimientos llevan a admitir que ***f*** y P_s permanecen constantes al variar el número de observaciones, se podrá decir que π_a, con el aumento de dicho número, disminuye y disminuye en razón menos que proporcional al aumentar la raíz cuadrada de tal número. No será posible precisar —a menos que no se conozcan los valores de f y P_s- la ley que rige la disminución de π_a cuando aumentan el número de observaciones.

Estas observaciones sabrán amargas a los muchísimos que han fundamentado los resultados de sus investigaciones sobre los llamados «test de significatividad», y han trabajado en ellos a cincel. Pero en realidad no pueden ser del agrado de ningún estudioso de la Estadística, porque llevan a concluir que no estamos en situación de medir la previsibilidad de un resultado estadístico o de una diferencia entre dos resultados estadísticos, en cuanto aquel o éste se consideren representativos del resultado o de los resultados que se deberían obtener en un conjunto de casos mayor, salvo el caso de resultados que se sabe están afectados únicamente por errores aleatorios.

Resaltar los resultados obtenidos en un conjunto necesariamente limitado de casos, de los resultados que se saben obtenidos de un conjunto de casos mayor, o, al límite, infinito, no es verdaderamente indispensable en todas las investigaciones estadísticas y para todos los fines —como pretenden algunos erróneamente-; pero en algunas investigaciones y encuestas es verdaderamente indispensable y en otras, si no indispensable, es muy útil, con el fin de obtener la precisión deseada en las conclusiones. Me doy cuenta por tanto de la gravedad de dicha conclusión; pero el fin de la ciencia es, no obtener conclusiones agradables, sino conclusiones fundadas.

* * *

Pero aquí tenemos a nuestro probabilista que vuelve a ponerse nervioso. ¿Qué querrá? Recordaréis que debería haber agachado la cabeza ante las observaciones del super-probabilista, y aunque sea un trago amargo, reconocer su error. Pero, tenaz como todo buen científico, se lo ha vuelto a pensar y, antes de darse por vencido, ha querido analizar la dispersión de las extracciones de la lotería en las que sospechaba que podría haber trampas. Cálculos laboriosos, hechos y rehechos, le han llevado a la conclusión de que la dispersión es supranormal. Las aplicaciones de χ^2 y z demuestran que, por efecto del azar, tal dispersión se verificaría sólo una vez sobre mil. Entonces vuelve a la carga con el Director General de la Lotería, y esta vez con el apoyo del super-probabilista. «Ved —exclama triunfalmente, poniéndole al Director General bajo la nariz el cuaderno con sus cálculos— tengo razón, no podría verificarse ni una vez sobre mil, por efecto del azar, la dispersión que se verifica en vuestras extracciones. Podéis apostar mil contra uno que las extracciones no están ligadas a la suerte; hay estafa, os lo digo» Y el super-probabilista interviene de paso: «Señor Director General, le aseguro: he verificado de uno en uno estos cálculos y no encuentro nada más que decir. A la luz del Cálculo de probabilidades, él tiene razón. Que sea una estafa, como él dice, o que las extracciones se hagan de forma que se favorezca inconscientemente a determinados números o combinaciones, no lo sé; pero hay que reconocer que las extracciones no dependen puramente de la suerte». El Director General les escucha, reflexiona un momento, después se incorpora y secamente les dice: «Sé cómo se desarrollan las extracciones y puedo garantizar que se producen por pura suerte» y envía a 48 probabilistas y super-probabilistas sus cálculos y todo el Cálculo de probabilidades que conducía a conclusiones parecidas.

¿Quién tenía razón? Al menos desde el punto de vista lógico, tenía razón el Director General. ¿Por qué? Porque es verdad que el Cálculo de probabilidades nos dice que tenemos una probabilidad sobre mil de que por causas aleatorios se verifique la dispersión observada, pero no nos lo dice la probabilidad P_s de que, verificándose las perturbaciones sistemáticas sospechadas, dieran lugar a la dispersión en cuestión (todo lo que sabemos es que P_s está comprendido entre 0 y 1) y tampoco nos dice la frecuencia f de tales perturbaciones sistemáticas frente a las oscilaciones puramente aleatorias. En resumen, no se puede formular ningún juicio sobre la probabilidad de que la dispersión observada tenga carácter aleatorio sin conocer el sistema de extracción. Sí se puede excluir, como hace el Director General, basándose en el conocimiento del sistema de extracción, cree poder seguramente descartar que las perturbaciones sistemáticas se verifiquen de manera que si f = 0, la probabilidad π_α de que la dispersión observada sea aleatoria, no es de 1/1000 sino de

$$\frac{1/1000}{1/1000+0P_s}$$

es decir, es = 1.

Es un caso particular de la divergencia entre π_a y P_a, que ya habíamos aclarado de forma general. Pero es un caso particularmente significativo, en cuanto que ilustra otro aspecto de las aplicaciones del Cálculo de probabilidades que no convendría olvidar.

Entonces ¿cuál es la moraleja de la fábula?. La moraleja de la fábula —o mejor dicho, de esta última parte de la fábula— es que el Cálculo de probabilidades, y la Estadística por medio del Cálculo de probabilidades, aunque se apliquen a conjuntos de casos, no pueden nunca llevar a conclusiones seguras, sino sólo a conclusiones probables. Podemos aclarar dudas más o menos importantes —y ésta es ciertamente una función útil— pero no podemos desentrañarlas de forma definitiva. No podemos proporcionar «test de significación» sino «elementos de sospecha». Por muy alto que sea el grado de probabilidad de sus conclusiones, enfrentadas a un dato seguro obtenido por consciencia o experiencia deberá agachar la cabeza. Reconozcámoslo explícitamente y declarémoslo francamente a los investigadores antes de que se ellos mismos cometan el error.

No sería la primera vez. Nos han cogido en fallo flagrante en la aplicación de los esquemas teóricos.

* * *

Antes de abandonar el argumento que nos ha ocupado en los últimos párrafos, deseo resaltar cómo las críticas así emitidas no se referían más que a aplicaciones particulares, aunque muy importantes, del Cálculo de probabilidades a la Estadística. Es cierto que en algunos casos minan la determinación, de la probabilidad de verificarse una constante estadística, o de la probabilidad de que ésta se vea afectada por cierto error, y sobre todo el uso que generalmente se hace de ello, con el fin de decidir si las divergencias halladas entre los datos estadísticos, o entre estos y los correspondientes valores teóricos, tienen carácter aleatorio o son por el contrario significativos. Otras aplicaciones del Cálculo de probabilidades a la Estadística son legítimas.

Cuando, por conocimientos existentes de la estructura del fenómeno o siguiendo las adecuadas aplicaciones de la teoría de la dispersión, se juzgue que las variaciones del fenómeno dependen exclusivamente de causas aleatorios, la cuestión sobre el carácter significativo o aleatorio de dichas variaciones no se plantea y el uso del error que tiene cierta probabilidad de ocurrir, o de la probabilidad de que ocurra cierto error, resulta lícito para otro fin.

También es posible, entonces, calcular cuál es la probabilidad de cometer cierto error (que *a priori* se sabe puramente aleatorio) o un error no superior a un límite dado, en una investigación (finalizada o por hacer) de una amplitud determinada y que mida el grado de fidelidad que se otorga a las constantes que se han recabado o se recabarán. Y, viceversa, resulta posible decidir cuál es la amplitud que se daría a una encuesta con el fin de tener una probabilidad dada de no cometer un error superior a un límite dado. Así, el jefe de un estudio de estadística que, ante la imposibilidad de recabar todos los datos en una encuesta dada, se contenta con una investigación representativa basada en una fracción de datos elegidos al azar, se puede preguntar cuál es la fracción que deberá considerarse a fin de no incurrir en un error superior a un límite dado. Y el demógrafo, al que se le dice que, en un determinado número de nacimientos habidos en un país dado, se haya verificado una prevalencia de hembras, podrá calcular (sabiendo que el reparto de sexos en el total de nacimientos pueden considerarse dotado de dispersión normal) la probabilidad de que en ese país se verifique el fenómeno excepcional de una mayor probabilidad del nacimiento de hembras o, viceversa, el fenómeno normal de una mayor probabilidad del nacimiento de varones.

Las aplicaciones de tal tipo se llaman *aplicaciones deductivas del Cálculo de probabilidades*, que no están del todo exentas de dificultades; no en balde presuponen la sustitución del valor observado por el valor real (desconocido) de la constante, dificultades de las que ya habíamos hablado; pero se trata, de todas formas, de dificultades de carácter y relieve bien distintos de las que surgen de la sustitución de la variabilidad compleja a su componente aleatorio, por la sustitución de π_a por P_a y de otras circunstancias examinadas en los párrafos anteriores.

Y, frente a un comportamiento anómalo de los datos, parece legítimo, entonces, sospechar (teniendo en cuenta, además del azar, la incertidumbre derivada de la sustitución del valor observado por el valor desconocido de la constante) que es posible aislar errores de la encuesta, en la elaboración o en la transcripción, que por otra parte pasarían desapercibidos o quizá habrían hecho arder el cerebro de los estadísticos en busca de explicaciones artificiosas.

En caso de que los errores puedan descartarse plausiblemente, el comportamiento anómalo de los datos surgirá la duda de que la serie de datos considerada tenga carácter excepcional desde el punto de vista de la dispersión, que no esté sujeta sólo a variaciones aleatorias, sino que sea también perturbada por factores sistemáticos. No hay anomalía en la que teóricamente no puedan intervenir también factores puramente aleatorios, con probabilidad tanto más baja cuanto más pronunciada sea ésta, de modo que la duda no podrá ya ser resuelta sin conocimientos suplementarios, mediante el Cálculo de probabilidades. Cuanto más pronunciada es la anomalía, tanto más probable será que ésta no dependa únicamente de factores aleatorios; pero no podremos calcular esta probabilidad como se explicó más arriba, hasta que no se tengan, la probabilidad de la intervención de los factores sistemáticos en cuestión y sus efectos eventuales, conocimientos particulares que generalmente no se pueden tener. Tal duda podrá, por otra parte, servir de estímulo para aislar los factores sistemáticos de los que se sospecha han intervenido.

Las aplicaciones de este último tipo difieren de las precedentes en cuanto, en éstas, se compara el comportamiento del fenómeno que se prevé basándose en un esquema probabilista, establecido bajo la hipótesis de que los datos se vean afectados por errores aleatorios, por el comportamiento efectivo del fenómeno y por las eventuales divergencias, hay motivos para persuadirse de la intervención de factores sistemáticos y precisar sus efectos y, eventualmente, su naturaleza. Éstas se llaman, en contraposición a las precedentes, *aplicaciones inductivas del Cálculo de probabilidades* y, no menos que las precedentes, atestiguan la utilidad que el Cálculo de probabilidades tiene para la Estadística; pero no hay que pedirle —ya lo hemos dicho— más de lo que puede dar[9] .

Estas aplicaciones inductivas del Cálculo de probabilidades representan un caso particular de las aplicaciones de los esquemas teóricos sobre los que pasamos a hablar.

* * *

Los sistemas teóricos constituyen una de las zonas más fértiles y atractivas, pero también más peligrosas, de la Estadística, en los que a menudo se aventura, sin las precauciones necesarias, cierta categoría de estadísticos y paraestadísticos. Ya he tenido ocasión de poner en guardia a los especialistas en aplicaciones del Cálculo de probabilidades a la Estadística, contra las insidias que esconde, en un Congreso Iinternacional, recientemente celebrado en la Universidad de Ginebra por iniciativa de ésta y del Instituto Internacional de Cooperación Intelectual[10]. Pasaremos aquí rápidamente a las conclusiones.

El primero y más célebre ejemplo, en la Estadística, de la utilización de un esquema teórico es contemporáneo al surgimiento de la moderna tendencia de nuestra disciplina. De hecho, rebasa a Quetelet, que construyó el esquema de una

[9] Para la contraposición entre aplicaciones inductivas y aplicaciones deductivas del Cálculo de probabilidades y por la utilidad que una y otra revisten para la Estadística, cfr. obra citada *El sexo desde el punto de vista estadístico*, capítulos IV y V, Pp. 63.183

[10] El informe será publicado en las Actas del Congreso a cargo de las dos citadas instituciones. Para evitar repeticiones inútiles, citaré éstos para lo relativo a los autores citados en las páginas siguientes.

distribución de magnitudes dependientes de una causa constante y de múltiples causas aleatorias, y como tales, independientes una de otra, y concluía que la distribución llamada normal, a la que lleva dicho esquema, se verifica para múltiples características antropométricos humanos. A partir de aquí, concluían que los individuos de un grupo, que se distribuían según la ley normal, pertenecían al mismo tipo y no se diferenciaban nada más que por causas fortuitas. A pesar de algunas voces discordantes[11] dicha conclusión es generalmente aceptada por los estadísticos y estudiosos del Cálculo de probabilidades y es, por ejemplo, reafirmada en un conocido tratado sobre el asunto, realizado por uno de los más eminentes matemáticos vivos: Emilio Borel. Aplicando dicho criterio particularmente a la población francesa, deducía que ésta es, desde el punto de vista racial, relativamente homogénea. Refrendada por nombres también muy autorizados, la conclusión no ha dejado de llamar la atención de los biólogos, que la citan en sus tratados, como prueba del engaño de los resultados a los que puede llevar la Estadística. Se ha demostrado experimentalmente que una población racialmente heterogénea puede muy bien dar lugar a una distribución normal para algunos o para todos los caracteres y que, incluso, las característica de distribución normal que se someten a la investigación estadística se refieren generalmente a poblaciones genéticamente heterogéneas. Y en cuanto a la población francesa, ésta —quizá porque ha sido la primera en ser estudiada seriamente desde el punto de vista racial— se cita por los antropólogos como ejemplo típico de una población marcadamente heterogénea. No hay nada que replicarles.

Otro esquema, que goza de gran renombre, el es esquema sobre el cual Lexis ha construido la teoría de la dispersión. En las aplicaciones que se hacen a la proporción de sexos en los nacimientos, su autor fué llevado a la conclusión de que el sexo se determina en los óvulos de la madres y que la relación de óvulos de sexo masculino y de sexo femenino es igual en todas las mujeres o varía por oscilaciones aleatorias. La Genética al contrario ya ha demostrado que, en la especie humana, como en muchas otras, el sexo se determina en el acto de la fecundación y depende de la presencia o no, en el espermatozoide, del hetero-cromosoma, y la Estadística, por su parte, ha demostrado, como hemos recordado, que las diferencias en la proporción de sexos de las distintas proles son en parte sistemáticas. Aqui se ha evitado el escándalo, porque mientras la Genética y la Estadística llegaban, por una parte, a esos resultados, los mismos estadísticos habían, por otra, rectificado las deducciones de Lexis, llegando a conclusiones menos ambiciosas, pero más fundamentadas.

Luego se han construido decenas esquemas estadísticos para dar cuenta especialmente la curva de desarrollo de las poblaciones y la curva de distribución de la renta; pero, éstos a menudo se han contradicho uno a otro, se han eliminado sucesivamente, o han sido refutados con los datos de la experiencia común o de la introspección directa, así que los estadísticos, a mi parecer más avezados,

[11] Vean, por ejemplo, en las pp. 169-170 de la obra citada: *El sexo desde el punto de vista estadístico*.

nos les han hecho gran caso, considerándolos, todos o casi todos, como ejercicios, a veces ingeniosos, pero sin especial aporte de conocimientos.

El error lógico, que invalidaba tanto las conclusiones de Quetelet y de Borel, como las de Lexis, y que análogamente invalida las deducciones que se desprenden de muchos otros esquemas, es fundamentalmente idéntico y puede aclararse mediante un ejemplo banal. Supogamos que yo haga este razonamiento. Diez apartamentos de 100 m^3 cada uno, son como un edificio de 1000 m^3 (abstracción hecha de paredes, techos, intersticios, etc.). Constatado que un edificio tiene un volumen de 1000 m^3, no podremos conluir que éste tenga 10 apartamentos, cada uno de 100 m^3. Reiréis a carcajadas, por lo ingenua que os parecerá la argumentación; pero las argumentaciones de Quetelet y de Lexis no eran, en el fondo, diferentes. La intervención de una causa constante y de muchas causas aleatorias da lugar —decía Quetelet— a una curva normal. Por tanto, si constatamos que una característica se distribuye según la curva normal, significará –deducían ellos— que ésta depende de una causa constante y de muchas causas aleatorios. Si la probabilidad de un fenómeno es, en los casos aislados, constante e independiente de la verificación del fenómeno en los casos precedentes, la dispersión de su frecuencia de grupo a grupo de casos resulta normal; al contrario —razonaba Lexis— si un fenómeno presenta una dispersión normal, podremos deducir que para los casos aislados su probabilidad es constante o, como mucho, varía con oscilaciones aleatorias entre los grupos considerados. Está claro que, en el entusiasmo de su trabajo de pioneros, estos grandes han cambiado una condición suficiente por una condición necesaria. El equívoco se repite en la utilización de tantos esquemas que hoy se fabrican en serie. Ésto lo favorece el incierto significado que asume la palabra *hipótesis*. Un esquema —decimos— representa una hipótesis; verificada esta hipótesis, el esquema queda verificado. Hay que precisar que todos los esquemas representan sí, una hipótesis, pero una *hipótesis compleja*, o mejor aún, un *sistema de hipótesis*. Para que el esquema se pueda dar por verificado, no basta con verificar el efecto compejo, resultante del sistema de hipótesis, sino que, si las hipótesis aisladas son ν, hay que verificar, además de dicho efecto complejo, también $\nu-1$ hipótesis, y, en tales casos (cuando la enésima hipótesis admite alternativas), deben ser verificadas todas las ν hipótesis singulares. En el caso del edificio, el esquema comportaba la hipótesis de que hubiera 10 apartamentos y de que cada espacio tuviera una capacidad de 100 m^3. Determinada la capacidad total del edificio en 1000 m^3, sería necesario determinar después que éste contenía 10 apartamentos y que 9 de dichos apartamentos tenían cada uno una capacidad de 100 m^3, la capacidad del décimo apartamento se podrá deducir por diferencia. Entonces se podría decir que el esquema está realmente verificado; y así se habría representado la estructura del edificio.

En el caso de la dispersión normal, el esquema comporta dos hipótesis: a) probabilidad constante para los casos aislados; b) independencia de la probabilidad del fenómeno en casos aislados de su verificación en los casos precedentes. No basta observar que un fenómeno tenga dispersión normal para considerar el esquema verificado. La dispersión nomal podría ciertamente corresponder a otros esquemas, por ejemplo al que considera la probabilidad del fenómeno variable de grupo en grupo, pero en cada grupo, sujeto a compensación dependiendo de que se haya o no verificado en los casos precedentes. También hay que verificar, por lo tanto, la hipótesis de que la probabilidad de que se produzca el fenómeno en casos aislados sea independiente haberse producido en los casos precedentes. Así habremos verificado el efecto global del sistema de hipótesis y una de las dos hipótesis aisladas que ello comporta. Pero, en este caso, eso no basta, porque la hipótesis de que la probabilidad del fenómeno sea constante en casos aislados, admite una alternativa que deja igual la probabilidad del fenómeno de grupo en grupo de casos: la alternativa, por tanto, de que la probabilidad del fenómeno sea, en cada grupo, variable por oscilaciones aleatorios en subgrupos con número de casos no fijo, sino variable, por oscilacicones aleatorios. Es necesario por lo tanto observar también la primera hipótesis. En definitiva, no ésta, sino su alternativa aparece conforme a la realidad. Así el análisis del esquema de la dispersión normal resulta completo: y así es posible llegar a conclusiones no carentes de importancia, en cuanto que han llevado a excluir, de una parte las teorías teleológicas que sostenían que existe una compensación entre el número de sexos en grupos y subgrupos sucesivos de nacimientos y, por otra, las numerosas teorías que hacían depender la proporción entre sexos de factores climáticos, económicos, alimentarios, sanitarios, psicológicos, que normalmente varían, en el espacio y en el tiempo, de uno a otro grupo de nacimientos[12].

Los esquemas teóricos, cuando se han analizado a fondo, pueden también permitir penetrar en la estructura íntima de los fenómenos. Pero, fuera del esquema citado de la dispersión normal en su aplicación a la proporción de sexos en los nacimientos, ¿se ha hecho alguna vez un análisis parecido? Aunque se hubiese querido hacer, no siempre se ha podido. Muy pocos esquemas son tan simples como el de la dispersión normal y sólo excepcionalmente se prestan, como en el caso de la proporción de sexos en los nacimientos, a la indispensable verificación de las hipótesis aisladas que comportan.

Entre tanto, hasta que no se haga tal análisis, los esquemas teóricos pueden satisfacer solo a los que aspiran a tener una representación cualquiera fenómenos, sin preocuparse de si responde o no a la realidad; también pueden servir cuando conducen a fórmulas matemáticas, a fines de interpolación, pero no pueden aumentar nuestro conocimiento de la estructura real de los fenómenos, y se exponen a graves errores cuando se quiere deducir consecuencias sobre la naturaleza y sobre el desarrollo de éstos.

No se diga que las construcciones de Quetelet y Lexis representaban primeras apoximaciones, de acuerdo con el estado de desarrollo de las ciencias especiales, y destinadas a dejar lugar a las mayores aproximaciones, cuando los conocimientos científicos se mejorasen. ¡Sería vivir de ilusiones!

[12] Cfr. El sexo desde el punto de vista estadístico, op. Cit. Pag. 167-168

En realidad, la conclusión de Queteleet, sobre la pureza racial de las poblaciones con caracteres de distribución gausiana, al igual que la de Lexis, sobre la predeterminación del sexo desde los ovarios de la mujer, no tienen carácter relativo o aproximado, sino absoluto y definitivo; ni había nada, desde el punto de vista dominante entonces en antropología y biología, que lo impusiera o aconsejara. Éstas se apartan de la opinión dominante del tiempo, a la cual los estadísticos y investigadores deben, con sentido convergente, aproximarse sucesivamente, los primeros según un análisis profundo y lógicamente riguroso de las aplicaciones del Cálculo de probabilidades, los segundos en base a investigaciones ingeniosamente concebidas y actuando con los avanzados recursos de la técnica. Reconocer los propios errores es, también para la Estadística, la primera condición para enmendarse y progresar.

* * *

Merece mención aparte cierta categoría de aplicaciones de esquemas teóricos, en la cual, o por el modo con el que el esquema se ha construido, o por el modo en que se procede a su verificación, es el de esperar *a priori* que los datos empíricos concuerden con ella.

Un ejemplo apropiado de parecidas aplicaciones puede citarse de una verificación que Boas ha realizado de su teoría sobre la modificabilidad del índice cefálico.

Se sabe que Boas, estudiando los índices cefálicos de los inmigrantes a Nueva York, pertenecientes a distintas poblaciones europeas y comparándolos con los de sus hijos nacidos en Europa y en América, vió que los hijos nacidos en América, pertenecientes a poblaciones dolicocéfalas, como los sicilianos, resultaban menos dolicocéfalos y los que pertenecían a poblaciones braquicéfalas, como los judíos o los eslavos, resultaban menos braquicéfalos que los padres y hermanos nacidos en Europa; la diferencia aparece tanto más acentuada cuanto más largo es el intervalo desde la inmigración hasta el nacimiento. Una bomba que explosiona en una asamblea de ciudadanos no produce mayor conmoción que la que tal resultado produjo en el mundo de los antropólogos. Cuando se piensa que los antropólogos pretendían y pretenden reconocer, basándose en el índice cefálico, la permamencia y el resurgimento, entre las poblaciones actuales, de las razas que poblaron la tierra en el Paleolítico, hace centenares de miles de años, se comprende cómo han sentido segar la hierba bajo sus pies frente a tales resultados, presentados por una de las luminarias de su ciencia, que habrían demostrado que pocos años de entorno moderadamente distinto, como es el entorno americano frente al europeo, hacen modificar sustancialemtne el índice cefálico a lo largo de la misma línea familiar.

Una avalancha de críticas, más o menos serias, cayó sobre Boas, el cual, frente a ellas, y manteniendo los resultados obtenidos, dió una explicación, que puede parecer menos desastrosa para las concepciones tradicionales, en cuanto que no niega que las diferencias de carácter genético persistan en el mismo entorno, pero admite que otras diferencias, que tendrían carácter fisiológico, como serían las diferencias entre los índices cefálicos de las distintas poblaciones europeas, se puede comparar a través de generaciones y se atenúan, incluso de una generación a otra, en el entorno común ciudadano. Como confirmación de esta teoría, Boas recurre a los datos de la estadística italiana, los cuales por una parte, la *Antropometría Militar*, proporciona separadamente las distribuciones de índice cefálico de los militares según el registro de nacimientos y por otra, el censo, indica la composición de la población de las ciudades según el lugar de nacimiento. Si los índices cefálicos —razonaba Boas— presentan, al pasar al entorno ciudadano, la convergencia que yo admito, entonces la desviación cuadrática media del índice cefálico de los militares nacidos en una ciudad debería resultar inferior a la desviación cuadrática media del índice cefálico de sus padres de distintos orígenes, que se puede calcular admitiendo que los inmigrantes no se diferencian sistemáticamente por su índice cefálico de la población de la que provienen. Y está bien, pero la *Antropología Militar* facilita sólo los datos de una generación, y entonces Boas supone, en la verificación de su esquema, que la desviación cuadrática media del índice cefeálico fuese, en cada ciudad, la misma en las dos generaciones y compara por tanto la desviación cuadrática media del índice cefálico de los militares nacidos en la ciudad con la desviación cuadrática media del índice cefálico calculado en una población ficticia, resultante de militares nacidos en la ciudad y de militares inmigrados de las distintas localidades según las frecuencias recabadas del censo[13].

La última hipótesis, así presentada, llevaba prácticamente a un resultado obligado, dado que, si las medias de un carácter son distintas en varios grupos A, B, C, etc., y la desviación cuadrática media es, en varios grupos, del mismo orden de tamaño, es inevitable, por una conocida propiedad de la media aritmética, que la desviación cuadrática media en uno de los grupos sea menor que en la población global resultante de los distintos grupos ponderados[14].

[13] Cfr. Franz Boas y Helene M. Boas. «The head-forms of the Italians as influenced by heredity and environment» *American Anthropologist*, Vol. XV, n. 2, Abril-Junio 1913, particularmente en las pags. 185-188.

[14] Tal objeción ya se ha formulado de forma distinta en las pags. 224-227 de nuestra reseña *Las recientes publicaciones de estadística biológica de los italianos y sobre los italianos* («Archivo de Antropología y Etnografía», Vol.XLIV, Fasc. 2.º - 3.º, 1914) traducida después al alemán, bajo el título *Die neuere biologischstadistiche Literatur del Italiener und über die Italiener* («Archiv für Soziale Hygiene und Demographie·, IX Band 3 u 4 Heft. Cfr., en particular sobre el argumento, pag. 397-401). Mientras, en un escrito posterior (New evidence in regard to the instability of Human Types, «Proceedings of the National Academy os Sciences», Vol. 2, Diciembre 1916, pag 713 y sgtes.) Boas citaba una crítica mía y reconocía su fundamento, no parecía en cambio darse cuenta del contenido de la expuesta en el texto, a la que indicaba. Debo decir, sin embargo, a este propósito que, por lo que respecta a la teoría de Boas mis críticas tienen un carácter puramente metodológico, porque desde un punto de vista sustancial, no creo infundada su tesis de la plasticidad de los tipos humanos bajo la influencia, directa o indirecta, del entorno, tesis que resulta confirmada también por las investigaciones que, bajo mi dirección, se hicieron sobre los caracteres de las colonias alógenas de Italia, de campañas concretas, del Comité Italiano para el estudio de los problemas de la población.

Por lo tanto, no es para maravillarse que Boas haya verificado su esquema de esa forma, sería muy difícil que ocurriese de otra. Su verificación era puramente ilusoria.

Hipótesis análoga, u omisiones en el procedimiento de verificación, muestran como ilusoria la correspondencia de los hechos que se cree haber comprobado con otros esquemas, que circulan como moneda válida, o al menos aceptada por muchos, en el mercado estadístico.

Tal es el caso del esquema de la logística en sus múltiples variantes y complicaciones. Aquí se omite verificar (y a menudo no se podría hacer) el origen de la curva y la asíntota superior y, por otra parte, se consideran los datos de la población a intervalos bastante amplios, para eliminar variaciones demasiado bruscas que puedan ser intervenidas por causas excepcionales; y entonces basta echar un vistazo a la forma de la curva para entender fácilmente como, en una u otra de sus fases, se pueda adaptar el curso de cualquier población según ésta se desarrolle.

De parecida forma, curvas con un número suficiente de parámetros necesariamente se adaptan bien a series de datos cualquiera, independientemente de cómo se dispongan, sin que dicha adaptación pueda ser interpretada como una verificación de los esquemas correspondientes a las curvas.

Estas pretendidas verificaciones me traen a la mente la enseñanza que las malas lenguas de los estudiantes atribuían, creo que en broma, a un colega nuestro de zoología, según la cual el mordisco de la víbora se reconoce infaliblemente en el hecho de que deja dos agujeritos exactamente equidistantes entre ellos invariablemente dispuestos en una misma recta. Que los dos agujeritos tengan dichas características es incontestablemente cierto, pero que pudieran tener otros es obviamente imposible, ya que no se podría reconocer en ellos el bocado de la víbora. Análogamente, habiendo reconocido que los datos empíricos presentan ciertos caracteres que la teoría de Boas, o la logística, o una curva con muchos parámetros harían esperar, debemos preguntarnos si éstos podrían en la práctica presentar caracteres distintos, y, en caso negativo, concluir que se trata de una verificación ilusoria que no aumenta nada nuestro conocimiento de la estructura del fenómeno.

* * *

Los errores descritos más arriba, que se cometen en la verificación y en la utilización de los esquemas teóricos, no son más que casos particulares de un error muy conocido, en el cual se incurre más o menos fácilmente en todas las ciencias: el de olvidar al llegar a la meta las hipótesis de las que se ha partido o que se han adaptado por el camino. Los matemáticos son acusados de incurrir en él con cierta frecuencia. No sé si la acusación es fundada por otros campos de su actividad: sí lo ha sido por las aplicaciones que inicialmente se han hecho del Cálculo de probabilidades a la Estadística. Los esquemas teóricos, aunque erróneos y mal interpretados a veces, representan de hecho y en cualquier caso, un gran paso adelante en comparación con la situación en la que el Cálculo de probabilidades se aplicaba directamente a los distintos fenómenos estadísticos sin averiguar de ninguna forma si en éstos se verificaban las hipótesis que servían de base a dicho cálculo. Así, Bernoulli, Laplace y Poisson y lo mismo Quetelet y Herschel lo habían aplicado a las decisiones de cuerpos deliberativos y judiciales, y tras sus huellas otros, aunque no matemáticos, a la Estadística sanitaria. Corresponde a Quetelet, por la magnitud extensiva, y a Cournot, por la magnitud intensiva, el mérito de haber demostrado que no se podía proceder a la aplicación del Cálculo de probabilidades sin una previa comprobación de la existencia de las condiciones que esto presupone; pero sólo más tarde, con Dormoy y Lexis, la idea comenzó a tener aplicaciones sistemáticas. Así se desarrollaron aquéllas que después se llamarían aplicaciones inductivas del Cálculo de probabilidades[15].

A menudo me vuelve a la memoria una anécdota que leí de pequeño en una revista alemana. Un profesor, ensimismado en sus elucubraciones, es molestado por chavales que juegan bajo su ventana, y, para librarse de ellos, tiene una feliz idea: —Pero ¿qué hacéis aquí? —les grita—; ¿por qué no vais a ver la gallina de tres patas que enseñan en la plaza del Pueblo?. Los chavales echan a correr como saetas (se trata de ir al otro extremo de la ciudad) y difunden la noticia entre la gente que se cruzan. Y la noticia, al difundirse, se agranda. Ya no es una gallina, sino un pavo, un cordero, un buey, un elefante. No se dice ya que está en la plaza, sino en la casa de fieras. —Si, la casa de fieras que está más allá de la plaza del Pueblo. —Y es un elefante albino con dos trompas. — Sí, y el famoso explorador X, que acaba de volver de Sumatra, hablará sobre el fenómeno. — ¿A qué hora? — A las seis. ¿Quién puede abrigar dudas ante la indicación precisa de la hora y del nombre del explorador? Toda una multitud de personas serias iba hacia la casa de fieras, deseando ver el espectáculo. El profesor, dándose cuenta del insólito movimiento, pide información y, a su vez, coge precipitadamente su abrigo y su sombrero para llegar a tiempo, el también, a la exposición de un elefante albino con dos trompas. Así, en la ciencia, muchas veces se ponen en circulación hipótesis científicamente irreales y luego, olvidándose o sin advertir las consecuencias, se va tras ellas, dándolas por descubrimientos, conclusiones que no representan más que el eco multiplicado de la propia imaginación.

* * *

En conclusión, debemos reconocer que, no sólo el terreno de la Estadística está sembrado de peligros, sino también que quienes se mueven en ellos, caen a menudo. Efectivamente, al ponerle título a este discurso, dudaba entre *Los peligros de la Estadística* o *Los pecados de la Estadística*.

Una bella dama, que a menudo se aventura en países inexplorados y no pone objeciones a hacer componenda con cualquier tipo de gente y a veces se

---

[15] Para más detalles sobre el asunto, se puede ver en *El sexo desde el punto de vista estadístico*, op. Cit. Cap. IV pag. 73-89

acompaña de jovenzuelos aventureros, está inevitablemente expuesta a peligros y difícilmente se verá inmune. La Estadística es indudablemente un ciencia muy seductora, como lo prueba el hecho de que a ella se acercan no sólo los muchos que la conocen a fondo, sino aquéllos —y son muchos más— que la conocen apenas de vista. Es una ciencia que, para algunos campos, tiene una función de vanguardia y, en otros, es la única en aventurarse; es una ciencia que se asocia sin dificultad a cualquier otra rama del saber y luego a menudo agradece o busca la compañía de un atractivo, pero muy peligroso, sujeto, denominado el Cálculo de probabilidades. Parecía un «*niño prodigio*» este Cálculo de probabilidades, ¡cuántas esperanzas nos había hecho tener, en la puritana familia de las matemáticas a la que pertenece! Y no es que las esperanzas hayan decaído; pero ¡cuántas preocupaciones nos ha dado mientras tanto a los padres, siempre inclinados, como todos los padres, a echar la culpa de sus errores a las malas compañías! Las crónicas cuentan un escándalo provocado cuando éste ha querido meter la nariz en las decisiones de asambleas y en las sentencias de los jueces, pretendiendo dictar normas generales para la composición de los cuerpos deliberativos, y se murmura hasta que se había pensado seriamente —audacia casi increíble— calcular la probabilidad de que el padre Sol una mañana no se levantase del tálamo nupcial. Entretanto, los hermanos geómetras y primos estadísticos de vez en cuando se exprimen el cerebro para explicar ciertas contradicciones paradójicas de su comportamiento.

En tal compañía, con tantos atractivos y tantas tentaciones, no es para sorprenderse si la conducta de la Estadística no es siempre rigurosa a toda prueba. Ciertamente, con el pasar de los años, su individualidad se forma y su poder de autocrítica se desarrolla, de forma que se va haciendo más cauta en las iniciativas, más mirada con las amistades, más práctica en los fines, más coherente en las directivas, más mesurada en su comportamiento. La naturaleza, variedad y complejidad de sus manifestaciones son tales y tantas, por otra parte, las envidias de las que está rodeada en este bajo mundo, que difícilmente podremos esperar que su conducta esté exenta de crítica. Pero, por otra parte, aunque éstas pudieran ser parcialmente justificadas, no deberá olvidarse nunca, y nosotros no lo olvidaremos, que nadie podrá llegar donde ésta llega, ver o entrever lo que ésta ve o entrevé, recoger los frutos, aunque a veces no estén maduros, que ésta recoge, que nadie podrá, en una palabra, sustituirla, y por otra parte, que es una disciplina a la que el lema nietzschiano *gefährlich leben*[16] se aplica como una necesidad de vida, esta es precisamente la Estadística.

[16] «Der glaub er mir! – das Geheimniss, um die grösste Fruchtbarkeit und den grösten Genuss von Dasein einzuernten, heisst: gefährlich leben!» Nietzsche, Die fröliche Wissenschaft (la gaya scienza), Leipzig, C.G. Naumann, 1900, pag. 215.

Gefährlich leben : vive peligrosamente. N. T.