

Historia de la estadística

Sergio Hernández González

La palabra “estadística” a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y demás. Al instante de escuchar esa palabra, son estas las imágenes que llegan a nuestra imaginación.

La estadística es mucho más que sólo números apilados y gráficas bonitas. Es una ciencia con tanta antigüedad como la escritura, y es por sí misma auxiliar de todas las ciencias –medicina, ingeniería, sociología, psicología, economía, etcétera–, así como de los gobiernos, mercados y otras actividades humanas.

En la actualidad, la estadística ocupa un lugar de gran importancia en la investigación y en la práctica médica. En los estudios de medicina de cualquier país se incluyen varias asignaturas dedicadas a la estadística; es difícil, por no decir imposible, que un trabajo de investigación sea aceptado por una revista médica sin que sus autores hayan utilizado técnicas y conceptos estadísticos en su planteamiento y en el análisis de los datos.

La estadística que conocemos hoy día debe gran parte de sus logros a los trabajos matemáticos de aquellos hombres que desarrollaron la teoría de las probabilidades, con la cual se adhirió la estadística a las ciencias formales.

Desde los comienzos de la civilización han existido formas sencillas de estadísticas, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales y otras cosas.

Hacia el año 3000 a. de C. los babilonios utilizaban ya pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque. En el antiguo Egipto, los faraones

lograron recopilar, alrededor del año 3050 a. de C., prolijos datos relativos a la población y la riqueza del país; de acuerdo con el historiador

griego Heródoto, dicho registro de la riqueza y la población se hizo con el propósito de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

En el antiguo Israel, la Biblia da referencia, en el libro de los N ú m e r o s, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David, por otra parte, ordenó a Joab, general del ejército, hacer un censo de Israel con la finalidad de conocer el número de habitantes, y el libro C r ó n i c a s describe el bienestar material de las diversas tribus judías.

En China ya había registros numéricos similares con anterioridad al año 2000 a. de C. Los griegos, hacia el año 594 a. de C., efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años llevaban a cabo un censo de la población, y los funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. En la época del nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del Imperio.

Durante los mil años posteriores a la caída del Imperio Romano se hicieron muy pocas operaciones estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve y por Carlomagno en los años 758 y 762, respectivamente. En Francia se realizaron algunos censos parciales de siervos durante el siglo IX.

Después de la conquista normanda de Inglaterra en 1066, el rey Guillermo I encargó un censo en el año 1086. La información en él obtenida se recoge en el Domesday Book, o Libro del Gran Catastro, que es un documento acerca de la propiedad, la extensión y el valor de las tierras en Inglaterra. Esta obra fue el primer compendio estadístico de ese país.

Aunque Carlomagno en Francia y Guillermo el Conquistador en Inglaterra trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media. Durante los siglos XV, XVI y XVII, hombres como Leonardo de Vinci, Nicolás Copérnico, Galileo Galilei, William Harvey, Francis Bacon y René Descartes hicieron grandes operaciones con base en el método científico, de tal forma que cuando se crearon los Estados nacionales y surgió como fuerza el comercio

internacional, había ya un método capaz de aplicarse a los datos económicos.

Debido al temor que Enrique VII tenía de la peste, en el año 1532 empezaron a registrarse en Inglaterra las defunciones causadas por esta enfermedad. En Francia, más o menos por la misma época, la ley exigía a los clérigos registrar los bautismos, fallecimientos y matrimonios.

Durante un brote de peste que apareció a fines del siglo XVI, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos. Esa costumbre continuó muchos años, y en 1632 los llamados Bills of Mortality (Cuentas de Mortalidad) ya contenían datos sobre los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt compiló documentos que abarcaban treinta años, mediante los cuales efectuó predicciones sobre el número de personas que morirían de diversas enfermedades, así como de las proporciones de nacimientos de hombres y mujeres que cabía esperar. El trabajo de Graunt, condensado en su obra *Natural and political observations... made upon the Bills of Mortality* (Observaciones políticas y naturales... hechas a partir de las Cuentas de Mortalidad), fue un esfuerzo de inferencia y teoría estadística.

Alrededor del año 1540, el alemán Sebastián Muster realizó una compilación estadística de los recursos nacionales, que comprendía datos acerca de la organización política, instrucciones sociales, comercio y poderío militar. Durante el siglo XVII se aportaron indicaciones más concretas sobre los métodos de observación y análisis cuantitativo y se ampliaron los campos de la inferencia y la teoría estadística.

Los eruditos del siglo XVII demostraron especial interés por la estadística demográfica como resultado de la especulación sobre si la población aumentaba, disminuía o permanecía estática.

En los tiempos modernos, tales métodos fueron resucitados por algunos reyes que necesitaban conocer las riquezas monetarias y el potencial humano de sus respectivos países. El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en 7 moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después

de revisar miles de partidas de defunción, pudo demostrar que en tales años no fallecían más personas que en los demás. Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros.

Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística, que extrajo del término italiano *statista* (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia

serían el aliado más eficaz del gobernante consciente. La raíz remota de la palabra se halla en el término latino *status*, que significa “estado” o “situación”. Esta etimología aumenta el valor intrínseco de la palabra por cuanto que la estadística revela el sentido cuantitativo de las más variadas situaciones.

Uno de los primeros trabajos sobre las probabilidades corresponde al matemático italiano del siglo XVI Girolamo Cardano, aunque fue publicado 86 años después de su fallecimiento. En el siglo XVII encontramos correspondencia relativa a la probabilidad en los juegos de azar entre los matemáticos franceses Blaise Pascal y Pierre de Fermat, fundamentos sobre los que Christian Huygens, físico, matemático y astrónomo danés, publicaría un libro en 1656. Durante ese mismo siglo y principios del XVIII, matemáticos como Bernoulli, Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante, durante cierto tiempo la teoría de las probabilidades limitó su aplicación a los juegos de azar, y no fue sino hasta el siglo siguiente que comenzó a aplicarse a los grandes problemas científicos.

Durante el siglo XVIII empieza el auge de la estadística descriptiva en asuntos sociales y económicos, y es a finales de ese siglo y comienzos del XIX cuando se comienzan a asentar verdaderamente las bases teóricas de la teoría de probabilidades con los trabajos de Joseph Louis Lagrange y Pierre Simon de Laplace, del brillantísimo y ubicuo matemático y astrónomo alemán Carl Friedrich Gauss, y de Simeón-Denis Poisson. Previamente, cabe destacar el descubrimiento de la distribución normal por Abraham de Moivre, distribución que será posteriormente “redescubierta” por Gauss y Poisson.

Jacques Quételet es quien aplica la estadística a las ciencias sociales. Interpretó la teoría de la probabilidad para su uso en esas ciencias y aplicó el principio de promedios y de la variabilidad a los fenómenos sociales. Quételet fue el primero en efectuar la aplicación práctica de todo el método estadístico entonces conocido a las diversas ramas de la ciencia.

En el periodo de 1800 a 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría estadística: la teoría de los errores de observación, aportada por Laplace y Gauss, y la teoría de los mínimos cuadrados, realizada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Galton ideó el método conocido como *correlación*, que tenía por objeto medir la influencia relativa de los factores sobre las variables. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica, tales como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones.

Una vez sentadas las bases de la teoría de probabilidades, podemos situar el nacimiento de la estadística moderna y su empleo en el análisis de experimentos en los trabajos de Francis Galton y Kurt Pearson. Este último publicó en 1892 el libro *The Grammar of Science* (La gramática de la ciencia), un clásico en la filosofía de la ciencia, y fue él quien ideó el conocido test de Chi-cuadrado.

El hijo de Pearson, Egon, y el matemático nacido en Polonia Jerzy Neyman pueden considerarse los fundadores de las pruebas modernas de contraste de hipótesis.

Pero es sin lugar a dudas Ronald Arnold Fisher la figura más influyente de la estadística, pues la situó como una poderosa herramienta para la planeación y análisis de experimentos. Contemporáneo de Pearson, desarrolló el análisis de varianza y fue pionero en el desarrollo de numerosas técnicas de análisis multivariante y en la introducción del método de máxima verosimilitud para la estimación de parámetros. Su libro *Statistical Methods for Research Workers* (Métodos estadísticos para los investigadores), publicado en 1925, ha sido probablemente el libro de estadística más utilizado a lo largo de muchos años.

Mientras tanto, en Rusia, una activa y fructífera escuela de matemáticas y estadística aportó asimismo –como no podía ser de otro modo– su considerable influencia. Desde finales del siglo XVIII y comienzos del XIX cabe destacar las figuras de Pafnuty Chebichev y Andrei Harkov, y posteriormente las de Alexander Khinchin y Andrey Kolmogorov.

En el siglo XIX, con la generalización del método científico para estudiar todos los fenómenos de las ciencias naturales y sociales, los investigadores vieron la necesidad de reducir la información a valores numéricos para evitar la ambigüedad de las descripciones verbales.

En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de los datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El trabajo del experto estadístico no consiste ya sólo en reunir y tabular los datos, sino sobre todo en interpretar esa información.

El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchos conjuntos de datos se pueden estudiar con gran exactitud utilizando determinadas distribuciones probabilísticas. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

Sucesos de interés en el desarrollo de la estadística

A continuación se presenta una relación cronológica de diferentes sucesos que nos permiten tener una idea general de la evolución de la estadística.

Dos hechos contradictorios en la historia de la estadística

La estadística y el nazismo

Tal y como quedó dicho, R. A. Fisher constituye una figura capital en el desarrollo de la estadística moderna, y se puede incluso decir que es quizás

la más importante e influyente; sin embargo, también existen zonas de sombra en su importante trabajo. A raíz de los descubrimientos de Charles Darwin sobre el mecanismo hereditario de evolución de las especies, surgió una nueva teoría científica (?) denominada e u g e n e s i a, término acuñado por Francis Galton en 1883, quien era por cierto sobrino de Darwin y “descubridor” de las huellas digitales.

Podríamos definir la eugenesia como la ciencia que estudia cómo mejorar la raza humana, proporcionando los mecanismos para que las características que se consideran como mejores se desarrollen más rápidamente que las inadecuadas. Se trata por tanto de dirigir de forma controlada la selección natural. En cuanto escuchamos esta definición, enseguida nos viene a la mente el nazismo y sus teorías de superioridad de la raza aria, limpieza étnica y demás.

Desgraciadamente, no sólo muchos matemáticos sino también un gran número de científicos de otras especialidades fueron defensores de las teorías eugenésicas. La lista de los científicos que, al menos inicialmente, prestaron su apoyo a dicha teoría es lamentablemente muy grande. Entre los estadísticos hallamos a Galton, a Pearson y sobre todo a Fisher.

En 1933, el gobierno alemán, presidido por Hitler, promulgó la ley de esterilización eugenésica, que puede considerarse ya como el antecedente de los experimentos perpetrados en los campos de concentración y de las atrocidades cometidas en nombre de una supuesta experimentación médica en dichos campos.

Aunque en 1930 Huxley, Haldane, Hogben, Jennings y otros biólogos renombrados comenzaron a reaccionar en contra de lo descabellado de muchas ideas propugnadas por la eugenesia, ya era demasiado tarde puesto que dichas ideas habían logrado difusión e importancia, y no sólo en los regímenes fascistas europeos: un importante biólogo americano, Charles Davenport, financiado por la Carnegie Foundation, creó el Eugenics Record Office en 1910, y miles de americanos llenaron un “registro de rasgos familiares”, que era una especie de pedigree familiar.

Y fumar, ¿produce cáncer?

Hacia 1920 se observó un gran incremento de los fallecimientos debidos al cáncer pulmonar. Aunque había trabajos previos sobre la posible relación entre el hábito de fumar y el cáncer de pulmón, como los de Lombard y Doering (1928) y Müller (1939), no será sino hasta la década de los cincuenta –con los trabajos de Wynder y Graham (1950) y sobre todo de Doll y Hill (1952 y 1959)– que la cuestión cobrará verdadero interés e incluso propiciará agrios debates en la opinión pública. Este último trabajo, publicado en el British Medical Journal, es un estudio de casoscontroles, donde los casos eran los pacientes que habían ingresado en ciertos hospitales con diagnóstico de cáncer de pulmón, mientras que los controles eran pacientes cuyo ingreso se debía a otras causas. A ambos tipos de pacientes se le interrogaba sobre sus hábitos de fumar tabaco, de inhalar otros gases y otros posibles agentes etiológicos. Las encuestas fueron efectuadas por personal “ciego”, en el sentido de que desconocía el propósito del trabajo. El resultado fue que los casos y los controles tenían una exposición similar a todos los posibles factores de riesgo, salvo el tabaco, con los siguientes resultados:

Si efectuamos los cálculos, el odds ratio es de 9.1, y dado que las tasas de cáncer de pulmón en la población son bajas, puede interpretarse como un riesgo relativo de padecer cáncer de pulmón de los fumadores frente a los no fumadores. El resultado es estadísticamente significativo, con un nivel de confianza inferior a 0.001.

Sin embargo, estos trabajos recibieron numerosas y fuertes críticas de personalidades tan respetadas como Joseph Berkson, estadístico principal de la Clínica Mayo. Incluso Jerzy Neyman puso objeciones. Pero quizás el principal paladín de esas críticas fue nada menos que el gran R. A. Fisher, quien en 1958 publicó un artículo titulado “Cigarettes, cancer and statistics” en el Centennial Review, y dos artículos en la prestigiosa revista Nature titulados “Lung cancer and cigarettes” y “Cancer and smoking”.

En 1954, Doll y Hill comenzaron un estudio prospectivo, de cohortes, en el que se efectuaba un seguimiento de médicos británicos y se estudiaba la posible asociación entre las tasas de mortalidad y el hábito de fumar tabaco, que corroboró no sólo los resultados anteriores sino también una mortalidad más rápida debida también a otras causas –fundamentalmente enfermedades coronarias– entre los fumadores.

A medida que la evidencia se fue acumulando, tanto Berkson como Neyman fueron cambiando de opinión, aunque Fisher permaneció irreductible en su posición. Otro gran estadístico, Jerome

Cornfield, y cinco expertos más del Nacional Cancer Institute, de la American Cancer Society y del Sloan-Kettering Institute, escribieron un artículo en 1959 en el que se revisaban los diferentes trabajos publicados al respecto, así como las objeciones que habían sido planteadas tanto por Fisher como por Berkson y Neyman y el propio Tobacco Institute, demostrando la abrumadora evidencia a favor de la tesis de que el hábito de fumar es una causa importante del aumento en la incidencia de cáncer de pulmón.