



## Historia del razonamiento estadístico

Preparado por Luis M. Molinero (Alce Ingeniería)

Julio 2004

CorreoE: [bioestadistica@alceingenieria.net](mailto:bioestadistica@alceingenieria.net)

 [Artículo en formato PDF](#)

[www.seh-lelha.org/stat1.htm](http://www.seh-lelha.org/stat1.htm)

*Il est impossible que l'improbable n'arrive jamais*  
Emil Gumbel

### Introducción

En la actualidad la estadística ocupa un lugar de gran importancia tanto en la investigación como 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 la utilización de técnicas y conceptos estadísticos en su planteamiento y en el análisis de los datos. Y sin embargo esta situación es bastante reciente, baste señalar que el gran auge de la utilización del método estadístico, tanto para la planificación de experimentos como para el análisis de los datos obtenidos, podemos situarlo en los trabajos de quien sin lugar a dudas se considera como el padre de la estadística moderna, Ronald A. Fisher (1890–1962), aunque su aplicación generaliza en la medicina tardó más en llegar. El comienzo de los ensayos clínicos aleatorizados en medicina, fundamentados en esos trabajos de Fisher, empiezan a despegar con Sir Austin Bradford Hill (1897–1991). Precisamente los editores de Lancet, comprendiendo la necesidad de difundir y explicar las técnicas estadísticas a los médicos, encargaron a Hill escribir una serie de artículos destinados a explicar el uso correcto de la estadística en medicina; artículos que posteriormente darían lugar a un libro, auténtico bestseller del género, *Principles of Medical Statistics*, cuya primera edición corresponde a 1937, y la última a 1991. Esta evolución de la difusión de la estadística, va en paralelo con la de la ciencia en general y con la medicina en particular, cuyo gran desarrollo podemos situarlo en el siglo XIX. Tengo conocidos que todavía recuerdan que cuando eran niños han visto utilizar las sanguijuelas por los médicos para sangrar a los pacientes, o al menos eso dicen.

Precisamente se citan los trabajos de un eminente médico francés, Pierre–Charles–Alexandre Louis (1787–1872), como uno de los primeros en aplicar el razonamiento científico en un trabajo de investigación sobre la fiebre tifoidea, en el que estudió la mortalidad en relación con la edad de los pacientes, el tiempo de estancia en París y la eficacia de las sangrías. También analizó numéricamente la eficacia de las sangrías en el tratamiento de la neumonía. Aunque su intención fue similar a la que actualmente guía los ensayos clínicos, su planteamiento, a la luz de nuestros conocimientos actuales, dista mucho de ser adecuado: los grupos escogidos no eran comparables y sus tablas contenían errores aritméticos de bulto, aunque sus conclusiones en cuanto a la falta de eficacia de las sangrías hoy sabemos que son correctas. La oposición a la utilización de procedimientos numéricos para el análisis de datos clínicos era entonces, y continuó siendo durante mucho tiempo, la "posición oficial", ya que se argumentaba que cada persona, y por lo tanto cada paciente, era un ser "único" y no podían existir conclusiones generalizables; se contemplaba la medicina más como un arte que como una ciencia. Incluso un investigador tan prestigioso como Claude Bernard (1813–1878) rechazaba la utilización de datos estadísticos, considerando la medicina como una disciplina fundamentada únicamente en la fisiología experimental de cada paciente individual. Este debate todavía habría de durar bastante tiempo.

Aunque solemos situar el nacimiento de la ciencia moderna basada en la experimentación en el Renacimiento, es a finales del siglo XVIII y comienzos del siglo XX cuando se produce en Europa la denominada revolución industrial, comenzando en Inglaterra con el empleo de la máquina de vapor y la utilización del carbón como primera fuente de energía, que irá substituyendo poco a poco a los animales y a la mano de obra humana en diferentes tareas. Paralelamente asistimos al declive de los regímenes monárquicos, y a los comienzos de la difusión de la recogida sistemática de datos de tipo social y económico, de los cuales el propio Napoleón era

un gran entusiasta, dada su pasión por las matemáticas, y su tremendo espíritu organizativo. Precisamente contó con gran número de matemáticos franceses en su tarea reformadora, entre los que hay que contar a uno de los padres de la teoría de probabilidades, Pierre Simon de Laplace, que fue profesor suyo, y a quien nombraría incluso ministro del interior en 1799, aunque sólo duró en el cargo unas pocas semanas, pero al que concedería la Legión de Honor en 1805.

Es en ese entorno cuando podemos datar uno de los primeros y rotundos éxitos de lo que después conoceremos como epidemiología, al descubrir John Snow (1813–1858) en 1854 mediante un cuidadoso y planificado estudio, más de veinte años antes de que Pasteur y Koch sentaran las bases de la moderna microbiología, que el medio de propagación de las epidemias de cólera estaba en el suministro de agua de la ciudad de Londres, en línea diametralmente opuesta a la teoría entonces dominante de la "miasma" y su difusión área. Precisamente en una encuesta realizada a los médicos británicos en 2003 fue elegido por éstos como el médico más grande de la historia. No sólo fue pionero en la utilización de la higiene médica, sino también en el empleo de la anestesia. En cuanto al estudio de los datos de consumo de agua procedente de las distintas estaciones de bombeo de Londres y los números de muertes acaecidos en los diferentes distritos servidos por estas fuentes, resulta revelador la esmerada y cuidadosa recogida de datos, que permiten la obtención de conclusiones claras a partir de cálculos muy sencillos, de aritmética elemental, lo que contrasta poderosamente con la situación actual, en la que quizás se ha invertido la posición, y asignamos a la estadística un papel tan predominante, que incluso se emplea para intentar arreglar las consecuencias de una mala planificación o de un mal diseño del estudio, y otras veces para ornamentar con un halo de tecnicismo trabajos que son a todas luces mediocres o irrelevantes.

### Breve evolución histórica

Uno de los primeros trabajos sobre la probabilidad corresponde al matemático italiano del siglo XVI Girolamo Cardano (1501–1576), aunque fue publicado 86 años después de su fallecimiento.

Ya en el siglo XVII nos encontramos con correspondencia relativa a la probabilidad en los juegos de azar entre los matemáticos franceses Blaise Pascal (1623–1662) y Pierre de Fermat (1601–1665), fundamentos sobre los que Christian Huygens (1629–1695), físico, matemático y astrónomo danés, publicaría un libro en 1656.

Uno de los primeros propulsores de la idea del uso de la teoría de probabilidades en meteorología y medicina fue el matemático suizo Jakob Bernouilli (1654–1705).

Como ya se ha comentado en la introducción, durante el siglo XVIII comienza el auge de la estadística descriptiva en asuntos sociales y económicos y es a finales de este siglo y comienzos del XIX cuando empiezan a sentarse verdaderamente las bases teóricas de la teoría de probabilidades con los trabajos de Joseph Louis Lagrange (1736–1813) y Pierre Simon de Laplace (1749–1827), del brillantísimo y ubicuo matemático y astrónomo alemán Carl Friedrich Gauss (1777–1855), y de Siméon-Denis Poisson (1781–1840). Previamente cabe destacar el descubrimiento de la distribución normal por Abraham de Moivre (1667–1754), distribución que será posteriormente "redescubierta" por Gauss y Poisson.

Una vez sentadas las bases de la teoría de probabilidades, el nacimiento de la estadística moderna y su empleo en el análisis de experimentos, podemos situarlo en los trabajos de Francis Galton (1822–1911) y Karl Pearson (1857–1936). Este último publicó en 1892 el libro *The Grammar of Science*, un clásico en la filosofía de la ciencia y fue él quien ideó el archiconocido test del  $\chi^2$ .

Pero es Ronald Arnold Fisher (1890–1962) sin lugar a dudas la figura más influyente de la estadística, situándola como una poderosa herramienta para la planificación y análisis de experimentos. Contemporáneo de Pearson, desarrolló el análisis de la varianza, 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* publicado en 1925 ha sido probablemente el libro de

estadística más utilizado durante mucho tiempo.

El hijo de Karl Pearson, Egon Pearson (1895–1980) y el matemático nacido en Polonia Jerzy Neyman (1894–1981) pueden considerarse los fundadores de las pruebas modernas de contraste de hipótesis.

Mientras tanto en Rusia siempre ha habido una activa y fructífera escuela de matemáticos y la estadística, como no podía ser de otra forma, ha contado con su aportación e influencia. Durante finales del XVIII y comienzos del XIX cabe destacar las figuras de Pafnuty Chebichev (1821–1894) y Andrei Markov (1856–1922), y posteriormente Alexander Khinchin (1894–1959) y Andrey Kolmogorov (1903–1987).

En la siguiente tabla se presenta una relación cronológica de diferentes sucesos que nos permiten tener una idea de la evolución del razonamiento estadístico.

**Tabla 1. Relación de diferentes sucesos de interés en el desarrollo de la estadística, elegidos quizás algo arbitrariamente**

Año	Suceso	Autor
1532	Registro semanal de muertes en Londres	Sir W. Petty
1539	Registro de bautismos, bodas y fallecimientos en Francia	
1654	Correspondencia relativa a los juegos de azar con dados	P. de Fermat B. Pascal
1656	Publicación <i>De ratiociniis in ludo aleae</i>	Ch. Huygens
1662	Primer estudio demográfico publicado basado en actas de defunción	J. Graunt
1663	Publ. póstuma del libro <i>Liber de ludo aleae</i>	G. Cardano
1713	Publ. póstuma de <i>Ars Conjectandi</i>	J. Bernoulli
1714	Publ. de <i>Teoría de la probabilidad</i>	A. De Moivre
1763	Publ. póstuma de <i>Essay Towards Solving a Problem in the Doctrine of Chances</i> . Contiene el famoso teorema de Bayes	Rev. Bayes
1809	Publ. de <i>Theoria Motus Corporum Coelestium</i>	C.F. Gauss
1812	Publ. de <i>Théorie analytique des probabilités</i>	P.S. Laplace
1833	Planteamiento de los conceptos básicos para la creación de una máquina de calcular	Ch. Babbage
1834	Creación de la Statistical Society of London	
1836	Publ. de sus estudios sobre la eficacia de las sangrías terapéuticas	P.C.A. Louis
1837	Distribución de Poisson	D.S. Poisson
1839	Creación de la American Statistical Association (Boston)	
1855	Epidemiología del cólera	J. Snow
1889	Publ. of <i>Natural Inheritance</i>	F. Galton
1895	Introducción de las distribuciones asimétricas	K. Pearson
1900	Desarrollo de la prueba del $\chi^2$	K. Pearson
1900	Redescubrimiento de los trabajos de Mendel	W. Bateson
1903	Desarrollo del Análisis de Componentes Principales	K. Pearson

1908	Publ. de <i>The probable error of a mean</i>	W.S. Gosset ``Student''
1910	Publ. de <i>An introduction to the theory of statistics</i>	G.U. Yule
1911	Publ. <i>The grammar of science</i>	K. Pearson
1915	Distribución del coeficiente de correlación	R.A. Fisher
1921	Publ. <i>Treatise on probability</i>	J.M. Keynes
1925	Publ. <i>Statistical Methods for Research Workers</i>	R.A. Fisher
1925	Métodos de máxima verosimilitud ML	R.A. Fisher
1928	Pruebas para verificación de hipótesis	J. Neyman, E.S. Pearson
1933	Publ. de <i>On the empirical determination of a distribution</i>	A.N. Kolmogorov
1933	Probit análisis	C. Bliss
1935	Publ. de <i>The Design of Experiments</i>	R.A. Fisher
1936	Publ. de <i>Relations between two sets of variables</i>	H. Hotelling
1937	Cálculo de las tasas de desempleo en USA mediante muestreo aleatorio	M. Hansen, F. Stephan
1937	Publ. <i>Principles of Medical Statistics</i>	A.B. Hill
1940	Publ. del libro de texto <i>Statistical Methods</i>	G.W. Snedecor
1945	Pruebas no paramétricas	F. Wilcoxon
1947	Métodos secuenciales	A. Wald
1950	Publ. <i>Experimental Design</i>	W.G. Cochran, G.M. Cox
1950	Construcción del primer ordenador moderno con lámparas de vacío UNIVAC 1	
1974	Comercialización del microprocesador de 8 bit y 2 Mhz Intel 8080	
1976	Publ. <i>Time Series Analysis Forecasting And Control</i>	George Box, Gwilym Jenkins, Holden-Day
1977	Publ. <i>Exploratory Data Analysis</i>	J. Tukey
1977	Comercialización del ordenador personal Apple II. Comienzo de la revolución	
1981	Comercialización del IBM PC 5150	

### **Luces y sombras, y algunas anécdotas**

En general en nuestra formación académica estudiamos las asignaturas científicas de un manera lineal y aséptica, como un cuerpo de doctrina fuera de cualquier contexto histórico, desligado de la vida de personas reales que contribuyeron a su desarrollo, así como de las circunstancias históricas y sociales que propiciaron la aparición de nuevas teorías y conocimientos, lo cual contrasta de forma sorprendente con el auge y éxito de ventas de las novelas históricas, de las biografías, ensayos sobre la historia de la ciencia y libros en los que se mezcla la matemática y la ficción. Pero por otro lado mucha gente piensa que las implicaciones éticas y sociales de la ciencia son de gran importancia, existiendo incluso un sector importante de la población que desarrolla una auténtica hostilidad hacia las cuestiones científicas, considerando la ciencia como una especie de gran enemigo de la humanidad. Es por ello que creemos que resulta de gran interés centrar la evolución histórica de la ciencia, de la cual la estadística constituye una pequeña parcela, en su contexto, así como conocer algo más sobre la vida y esfuerzos de las personas que, para bien o para mal, contribuyeron a su

desarrollo. En el caso de la estadística, a quien esté interesado en la historia de sus protagonistas, de una forma amena aunque quizás algo superficial, le recomendamos el libro [\*The lady tasting tea\*](#).

## ■ La estadística y el nazismo

Tal y como queda dicho R.A. Fisher constituye una figura capital en el desarrollo de la estadística moderna, se puede incluso decir que quizás la más importante e influyente, y sin embargo también existen zonas de sombra en su importante trabajo. A raíz de los descubrimientos de Darwin sobre el mecanismo hereditario de evolución de las especies, surgió una nueva teoría científica (?) denominada *eugenesia*, término acuñado por Francis Galton en 1883, quien era por cierto sobrino de Darwin y fue el "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 sobre las que se consideran inadecuadas. Se trata por tanto de dirigir de forma controlada la selección natural. En cuanto escuchamos esta definición en seguida nos viene a la mente el nazismo y sus teorías de superioridad de la raza aria, la limpieza étnica... El desarrollo de la teoría de la evolución de las especies, el auge del método estadístico aplicado a las ciencias y el auge de los fascismos europeos, comparten la secuencia temporal y se influenciaron mutuamente. Desgraciadamente no sólo matemáticos sino también gran número de científicos de otras especialidades fueron defensores de las teorías eugénicas. La lista de científicos que, al menos inicialmente, prestaron apoyo a la teoría eugénica es lamentablemente demasiado grande; entre los estadísticos empezando por Galton, Karl Pearson, y sobre todo Fisher. Otros científicos fueron Haldane, Huxley, Castle, Morgan, Jennings, Davenport, Goddard, Grant, Konrad Lorenz..., inventores como Graham Bell, o escritores como Bernard Shaw (Sin embargo resulta curioso que a éste último se le atribuya la anécdota de que en cierta ocasión una bellísima actriz de la época le comentara en un acto social que harían buena pareja, ya que sus futuros hijos serían tan guapos como ella y tan listos como él, a lo que parece que le contestó que probablemente no, ya que podían resultar tan feos como él y tan tontos como ella).

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

Aunque en 1930 Huxley, Haldane, Hogben, Jennings y otros biólogos comenzaron a reaccionar contra lo descabellado de muchas ideas propugnadas por la eugenesia, ya era demasiado tarde, puesto que dichas ideas había adquirido 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 cumplieron un "registro de rasgos familiares", una especie de pedigríe familiar...

Todo ello revela que, contrariamente a lo que muchas veces creemos o tratamos de creer, no podemos desligar la evolución científica del marco social en el que ésta se desarrolla, existiendo un camino de ida y vuelta entre ambos, y siendo frecuentemente difícil, incluso en ocasiones imposible, desligar las propias creencias y opiniones a priori en la interpretación de los hechos observados, como vamos a ver en el siguiente apartado

## ■ ¿Fumar produce cáncer?

Hacia 1920 se observó un gran incremento de los fallecimientos por cáncer de pulmón. Aunque existen 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 hasta 1950 con los trabajos de Wynder y Graham (1950), y sobre todo Doll y Hill (1950, 1952), que la cuestión cobrará verdadero interés e incluso propiciará agrios debates en la opinión pública. Este último trabajo, publicado en BMJ, es un estudio casos–controles, donde los casos lo constituyen los pacientes que ingresaban 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 les interrogaba sobre sus hábitos tabáquicos, inhalación de otros gases, y otros posibles agentes etiológicos diferentes. Las encuestas fueron efectuados por personal "ciego", en el sentido de que desconocían el propósito del trabajo. El resultado fue que los casos y controles tenían una exposición similar a todos los posibles factores de riesgo, salvo el tabaco, con los siguientes resultados:

	<b>Casos</b>	<b>Controles</b>	<b>Total</b>
<b>Fumador</b>	1350	1296	2646
<b>No fumador</b>	7	61	68
<b>Total</b>	1357	1357	2714

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, el odds ratio 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 probabilidad inferior a 0.001.

Sin embargo estos trabajos recibieron numerosas y fuertes críticas y 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 (a quien en [vemos en muchas fotografías con su pipa en la mano](#)), quien en 1958 publicó un artículo titulado "*Cigarettes, Cancer and Statistics*" en 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 tabáquico, que corroboró no sólo los resultados anteriores sino también una mortalidad más rápida también por 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 con cinco expertos más del National Cancer Institute, de la American Cancer Society, y el Sloan-Kettering Institute, escribieron un artículo en 1959 en el que se revisaba los diferentes trabajos publicados al respecto, así como las objeciones que habían sido planteadas tanto por Fisher, como Berkson, y Neyman, y el propio Tobacco Institute, demostrando la abrumadora evidencia a favor de que la tesis de que el hábito de fumar es una causa importante del aumento de la incidencia de cáncer de pulmón.

Esta historia ilustra tanto la dificultad que entraña probar mediante argumentos estadísticos una relación causa efecto, como la importancia que tienen las propias creencias en el análisis de los datos, incluso entre investigadores que supuestamente deberían tener una opinión ecuaníme.

## ■ Primer ensayo clínico aleatorizado

Bajo la influencia de los trabajos de Fisher y de sus publicaciones sobre el correcto diseño de experimentos, así como la utilización de la aleatorización como metodología para eliminar el sesgo y controlar la variabilidad experimental, la estadística comenzó a ser empleada también en la experimentación médica, de la mano de uno de sus principales impulsores Hill, como ya se ha comentado. Se suele citar como primer ensayo clínico aleatorizado el efectuado por el British Medical Research Council en 1946, sobre el uso de la estreptomycin en el tratamiento de la tuberculosis pulmonar, en el que intervino Hill y que fue planificado, ejecutado y publicado con un cuidado esmerado.

También en América la metodología estadística comenzaba su exitosa carrera en el mundo de la investigación médica. En 1954 se llevó a cabo el mayor y más costoso estudio médico jamás realizado, para evaluar la eficacia de la vacuna de Salk como protección contra la poliomielitis. Puesto que la incidencia anual de polio era solo del orden de 1/2000, fue necesaria la participación de alrededor de 2 millones de niños, con un coste superior a 5 millones de dólares, para comprobar definitivamente la eficacia de la vacuna.

## ■ Breve historia de la regresión logística

*It is an universal rule that the unknown kinsman in any degree of any specified man, is probably more mediocre than he.*  
Francis Galton 1886

El concepto de regresión es uno de los pilares de la estadística, y data al menos desde principios de 1800 con los trabajos de Legendre, Gauss y Laplace. Es posible que el término *regresión* sea debido a Francis Galton, quien acuñó el término "regresión hacia la media" para describir la observación de que los hijos de padres muy altos tienden a ser algo más bajos que sus progenitores, y por el contrario los hijos de padres muy bajos suelen ser algo más altos, y por lo tanto acercarse en ambos casos más a la media de la población. Este fenómeno que se produce en muchas más circunstancias en la naturaleza, queda muy bien explicado por Stephen M. Stigler con el siguiente ejemplo. Supongamos que efectuamos en dos momentos diferentes de tiempo un examen sobre una materia concreta a un alumno, y evaluamos primero uno de ellos, observando que obtiene una nota mucho más alta que la media de sus compañeros de clase. ¿Cómo de buena esperamos que sea la puntuación en el segundo examen? Probablemente alta, pero también probablemente no tan alta como en la primera ocasión, ya que probablemente el gran éxito en la primera ocasión se deba a dos componentes: por un lado la capacidad del alumno (componente estable o permanente) y por otro un cierto grado de suerte (componente transitorio y en cierta medida aleatorio). El coeficiente que medía esa *regresión* hacia la media pasó desde entonces a indicarse con la letra *r*.

La posible presencia de regresión hacia la media entraña una "trampa" en la que es fácil caer ignorando su presencia. En el estudio de datos de tratamientos que reducen los niveles anormalmente altos de una variable biológica, como puede ser el caso de la tensión arterial, podemos analizar el subgrupo de pacientes con valores más elevados, observando un mayor descenso que puede no ser debido en su totalidad al tratamiento, sino precisamente a esa regresión hacia la media, que aparecería incluso sin tratamiento alguno. Esa es una de las ventajas de la aleatorización, entre otras, ya que permite compensar estos sesgos.

Aunque este primer concepto de regresión no tenga nada que ver con el sentido que actualmente se utiliza para esa palabra, que designa las técnicas empleadas para construir funciones matemáticas que permiten calcular o predecir el resultado de una o más variables a partir de otras variables.

Una de las técnicas de regresión más utilizada actualmente en medicina es la *regresión logística*. Ya en 1937 Bartlett utiliza la transformación  $\log[y/(1-y)]$  para analizar proporciones. También Fisher y Yates sugieren en 1938 el uso de esa transformación para analizar datos binarios. El término logit fue introducido por Joseph Berkson en 1944 para designar esta transformación y sus trabajos popularizaron la utilización de la regresión

logística. Jerome Cornfield utilizó la regresión logística para el cálculo de odds ratio como valores aproximados del riesgo relativo en estudios de casos y controles.

Pero quizás el principal difusor de la regresión logística fue David R. Cox en 1970 con su libro *The Analysis of Binary Data*.

Uno de los últimos avances más importantes en el campo de la regresión de los últimos años corresponde a la introducción de los *modelos lineales generalizados GLM*, por los estadísticos británicos John Nelder y R.W.M. Wedderburn, en 1972, que unifican toda la teoría existente en cuanto a modelos probit y modelos logísticos, con los modelos lineales basados en la distribución normal, así como el análisis de la varianza. Curiosamente el principal algoritmo para ajustar estos modelos se denomina "*Fisher scoring*", debido a que fue introducido por Fisher en 1935 para ajustar modelos probit de máxima verosimilitud.

## ■ Los ordenadores

Muchos de los problemas de estimación de parámetros de modelos estadísticos no tienen una solución algebraica fácil o ni siquiera es posible obtener una fórmula para calcularla. Es el caso de la mayoría de las estimaciones por el método de máxima verosimilitud. Sin embargo es posible obtener una solución suficientemente aproximada mediante un método que se denomina *iterativo*. En esencia la idea es la siguiente: se trata de determinar el máximo de una función matemática (la función de verosimilitud). Si imaginamos la función como una colina que subimos con los ojos cerrados podemos suponer que mientras al avanzar sigamos ascendiendo todavía no hemos llegado a la cima, pero si en un momento determinado, al avanzar pasamos de ascender a descender entonces quiere decir que hemos llegado a la cumbre, la posición del máximo. El mecanismo por tanto es proponer un valor para el parámetro que se va a estimar y calcular el valor de la función. Aumentamos por ejemplo ligeramente el valor del parámetro y volvemos a calcular el valor de la función, si éste aumenta seguimos aumentando el valor de la solución y en caso contrario lo disminuimos. El tamaño del nuevo paso va a depender en cada momento de la pendiente observada: si la pendiente es grande daremos un paso grande, si es pequeña el siguiente paso será más pequeño. Aunque parece ser que este tipo de métodos numéricos ya eran utilizados en la antigua China y en el antiguo Egipto, sí que se sabe que era empleado por los bancos en la época del Renacimiento para cálculos de interés, guardando en secreto cada banco la fórmula concreta que utilizaba.

Está claro que incluso sin estos cálculos iterativos, los cálculos necesarios en cualquier procedimiento estadístico, aunque sólo sea para determinar la desviación estándar de un conjunto de datos, son tremendamente tediosos para ser efectuados con una calculadora básica no programable y no digamos a mano. De ahí que no resulte extraño que la verdadera "edad de oro" de la estadística haya ido pareja con el auge del ordenador personal. Aunque el enfoque estadístico de Fisher sigue siendo hoy igual de correcto y podemos decir que la estimación por el método de máxima verosimilitud es la base de la mayor parte de los nuevos métodos desarrollados, la diferencia en cuanto a su empleo universal radica ahora en la potencia de cálculo de la que disponemos, que ha permitido resolver problemas antes intratables y sobre todo divulgar su uso.

Un gran parte de las técnicas de análisis de la estadística moderna, como por ejemplo los métodos de montecarlo, la estimación autosuficiente, la utilización de modelos bayesianos, etc. no son posibles sin la utilización del ordenador para efectuar los cálculos.

## ■ Los avances más recientes

Después de la segunda guerra mundial los avances recientes en física, economía y telecomunicaciones, abrieron nuevos campos para la utilización de técnicas estadísticas, entre las que cabe destacar el análisis de series temporales, la teoría de juegos, teoría del caos, redes neuronales, aunque la influencia de todas ellas en el campo de la estadística médica ha sido escasa.

Sin embargo uno de los campos en los que la bioestadística va a experimentar sin lugar a dudas un mayor crecimiento en un futuro reciente, como en muchas otras áreas de la medicina, es el campo de la genética y concretamente de la epidemiología genética.

## Enlaces

- [History of statistical thinking in medicine](#)  
Tar Timothy Chen  
ADVANCED MEDICAL STATISTICS  
edited by Ying Lu (University of California, San Francisco, USA) & Ji-Qian Fang (Sun Yat-Sen University, Guangzhou, China)
- [From Association to Causation: Some Remarks on the History of Statistics](#)  
David Freedman  
University of California, Berkeley.  
30-Mayo-1999
- [Materials for the History of Statistics](#)  
University of York
- [The MacTutor History of Mathematics archive](#)
- [Historia de la estadística. Biografías](#)

## Bibliografía

- **The Lady Tasting Tea. How statistics revolutionized science in the twentieth century.**  
David Salsburg  
Owl Books (NY); 2nd Rep edition (May, 2002).
- **The Taming of Chance** (Ideas in Context)  
  
Ian Hacking  
Cambridge University Press; (August , 1990)
- **Statistics on the Table: The History of Statistical Concepts and Methods**  
Stephen M. Stigler  
Harvard University Press; Reprint edition (August, 2002)



[Indice de artículos](#)

[Principio de la página](#)