

# LAS CONTRIBUCIONES DE GAUSS A LA ESTADISTICA

David A. Sprott.\*

## I. Introducción

Para el progreso de las ciencias y las matemáticas es importante estudiar en su forma original, o tan cercano del original como sea posible, los trabajos del pasado hechos por los científicos y matemáticos eminentes. Ver, por ejemplo, Fisher (1959). Esta afirmación se aplica igualmente a la estadística moderna. Por esta razón, como se verá más adelante, es muy importante estudiar de primera mano los conceptos y métodos de Gauss. Gauss escribió en latín y alemán. Sus obras sobre estadística fueron traducidas al francés por Bertrand (1855), lo cual fue autorizado y aprobado por Gauss. La versión francesa fue traducida al inglés por Trotter (1957). La presentación de los métodos de Gauss en este artículo se basa en estas traducciones.

Las contribuciones de Gauss a la estadística se conocen bajo el nombre de los mínimos cuadrados. Este nombre, sin embargo, da poca idea de la amplitud de su importancia teórica y práctica. En el Libro 2, Sección 3 de su volumen sobre las órbitas de los planetas, Gauss (1809), discutió la estimación de los seis parámetros que determinan la órbita elíptica de un planeta basado en  $n > 6$  observaciones. Su segunda exposición (Gauss 1821, 1823a, 1826) fue presentada en tres artículos extensos a la Sociedad Real de Göttingen. Es claro de lo dicho por Gauss en la introducción a estos artículos que él consideró este problema como muy importante.



\*Departamento de Estadística y Ciencias Actuariales, Universidad de Waterloo, Ontario, Canadá. Basado en los artículos del mismo autor aparecidos en 1978 y 1983.

No es mi propósito en este artículo trazar las ideas del pasado con la intención de asignar prioridades en sus descubrimientos. Los debates sobre prioridades frecuentemente involucran interpretaciones altamente subjetivas con respecto al periodo exacto en que una idea específica fue formulada. Esto requiere también el definir cuál es la idea exacta. Por ejemplo, hubo (y hay aún) una gran controversia en torno al origen de los mínimos cuadrados. Legendre fue el primero en publicar acerca de los mínimos cuadrados en 1805, y fue responsable del nombre mínimos cuadrados (en francés, *moindre carrés*). Sin embargo, Legendre discutió solamente el principio de los mínimos cuadrados, como se discute en la Sección 3. Pero, como se explicará en las secciones que siguen a la Sección 3, fue Gauss quien desarrolló la teoría de los mínimos cuadrados e hizo una gran parte de la teoría estadística hoy conocida bajo el nombre de regresión lineal con el famoso modelo lineal de Gauss. De hecho, en esta área importante de la estimación, Gauss desarrolló tanto los fundamentos como las aplicaciones de la estadística como se encuentran casi hasta el día de hoy, tal vez aún más avanzado que ahora. No todos los artículos que hoy se escriben son contribuciones positivas.

## 2. El modelo lineal de Gauss

El primer enfoque de Gauss a los mínimos cuadrados se incluye en su libro *Theoria Motus Corporum Caelestium* y ocurre en su discusión del cálculo de las órbitas de los planetas. El problema es estimar las  $k$  cantidades desconocidas  $\theta' = (\theta_1, \theta_2, \dots, \theta_k)$  que determinan la órbita, llamadas *parámetros* en la estadística, basado en  $n > k$  observaciones  $\mathbf{Y}' = (y_1, y_2, \dots, y_n)$ . No es posible medir  $\theta$  directamente; solamente es posible medir funciones de  $\theta$ ,  $\xi_i = \xi_i(\theta)$ ,  $i = 1, 2, \dots, n$ . Si  $\xi$  puede observarse sin error,  $\mathbf{Y} = \xi$ , entonces después de observar  $\mathbf{Y}$ ,  $\xi$  sería conocido numéricamente. Se sigue que, sería necesario solamente resolver las ecuaciones  $\mathbf{Y} = \xi$  para  $\theta$ , lo cual es posible puesto que el sistema de las ecuaciones debe ser consistente. Sin embargo, en la práctica no es posible medir  $\xi$  sin error, así las ecuaciones son

$$\mathbf{Y} = \xi + \mathbf{e}, \quad (1)$$

donde  $\mathbf{e}' = (e_1, e_2, \dots, e_n)$  son los errores en las observaciones  $\mathbf{Y}$  (errores observacionales), y  $\mathbf{Y}$ ,  $\xi$ , y  $\mathbf{e}$  son vectores  $n \times 1$ . La notación  $\mathbf{A}'$  denota la transpuesta de la matriz  $\mathbf{A}$ .

En el caso especial, pero muy común (por lo menos aproximadamente), donde  $\xi$  es lineal en  $\theta$ ,  $\xi_i = \sum x_{ij}\theta_j$ , o en la notación matricial  $\xi = \mathbf{X}\theta$ , las  $x_{ij}$ 's se suponen constantes conocidas, (1) es

$$\mathbf{Y} = \mathbf{X}\theta + \mathbf{e}; \quad (2)$$



Carl Friederich Gauss (1777-1855).

$\mathbf{X} = (x_{ij})$  es una matriz  $n \times k$ , y  $\theta$  es un vector  $k \times 1$ . Este es el modelo lineal de Gauss discutido en los libros de texto de estadística como la regresión lineal.

## 3. El principio de los mínimos cuadrados.

El principio de los mínimos cuadrados consiste en elegir el valor de  $\theta$  que minimiza la suma de los cuadrados  $Q = \sum (y_i - \xi_i)^2$ . Así  $\theta$  es una solución de  $\partial Q / \partial \theta_j = 0$ ,  $j = 1, 2, \dots, k$ . Estas son las ecuaciones de los mínimos cuadrados

$$\sum_{i=1}^n (y_i - \xi_i) \partial \xi_i / \partial \theta_j = 0, \quad j = 1, 2, \dots, k. \quad (3)$$

De (2) donde  $\xi = \mathbf{X}\theta$ , (3) puede escribirse como  $\mathbf{X}'\mathbf{X}\theta = \mathbf{X}'\mathbf{Y}$ , de modo que la solución es el estimador clásico de los mínimos cuadrados  $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ .

Como se dijo anteriormente, han habido discusiones extensas en torno a la prioridad en el desa-

rollo de los mínimos cuadrados, puesto que Legendre consideró también (1) y (2) y produjo (3) y  $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ . Legendre lo publicó en 1805 antes que Gauss, quien lo publicó en 1809. Sin embargo, es importante subrayar el hecho de que Legendre enunció solamente el principio de los mínimos cuadrados como se describió arriba. Gauss, al contrario, desarrolló adicionalmente la teoría de los mínimos cuadrados, como se describe en las secciones siguientes.

#### 4. La teoría de Gauss: el primer enfoque (1809)

Si se asume que las observaciones son igualmente precisas e independientes, y si la densidad de la probabilidad del error  $e_i$  es  $f(e_i)$ , entonces la densidad de la probabilidad de los errores conjuntos es

$$P = \prod_{i=1}^n f(e_i) = \prod_{i=1}^n f(y_i - \xi_i).$$

Suponga que todos los valores  $\theta_1, \theta_2, \dots, \theta_k$ , son *a priori* igualmente probables. Entonces, la densidad de la probabilidad *a posteriori* de  $\theta$  obtenida mediante el teorema de Bayes (1763) es

$$f(\theta | Y) = C \prod_{i=1}^n f(y_i - \xi_i) = CP, \quad C = 1 / \int_{\Omega} P d\theta, \quad (4)$$

donde  $\Omega$  es el espacio del parámetro  $\theta$ . Ver, por ejemplo, Sprott (1984). El valor  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$  más probable de  $\theta$  corresponde al máximo de  $P$ , y se determina mediante las ecuaciones

$$\begin{aligned} \partial \log P / \partial \theta_j &= \sum_{i=1}^n [d \log f(e_i) / d e_i] \partial e_i / \partial \theta_j \\ &= - \sum_{i=1}^n [f'(y_i - \xi_i) / f(y_i - \xi_i)] \partial \xi_i / \partial \theta_j \\ &= 0, \quad j = 1, 2, \dots, k. \end{aligned} \quad (5)$$

Para obtener una solución explícita  $\hat{\theta}$ , se necesita especificar la forma analítica de  $f(e)$ . Para obtener esto, Gauss asumió que el valor más probable de una sola cantidad  $\xi$  es la media  $\bar{y} = (y_1 + y_2 + \dots + y_n) / n$  de sus valores observados. Para  $k=1$ , y  $\xi_1 = \xi_2 = \dots = \xi_n = \xi$  en (1), las ecuaciones (5) se vuelven  $\sum f'(y_i - \xi) / f(y_i - \xi) = 0$ , y la solución debe ser  $\xi = \bar{y}$ , por lo tanto  $\sum f'(y_i - \bar{y}) / f(y_i - \bar{y}) \equiv 0$  para todos los valores de  $n$ . Esta es una ecuación funcional, cuya solución que maximiza a  $P$  es  $f(z) = k \exp(-h^2 z^2)$ , Aczel (1976, pp. 47-48, 106-109). Esta es la distribución normal o la distribución clásica de Gauss. La cantidad  $h$  fue llamada la "precisión" por Gauss. En la notación moderna,  $h^2 = 1/2\sigma^2$ , y entonces  $k = 1/\sqrt{2\pi}\sigma$ , de modo que  $f(z)$  es la densidad de probabilidad

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{z^2}{2\sigma^2} \right] \quad (6)$$

En esta forma,  $\sigma^2$  es la varianza de  $z$ , y  $\sigma$  se llama error estándar de  $z$ .

Si se usa (6) en (4), entonces

$$P = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[ -\frac{1}{2\sigma^2} \sum (y_i - \xi_i)^2 \right],$$

la cual es un máximo cuando

$$Q = \sum (y_i - \xi_i)^2$$

es un mínimo. Así, el valor más probable de  $\theta$  en (4) es el valor  $\theta$  que hace  $Q$  un mínimo, donde  $Q$  es la suma de los cuadrados de las diferencias de los valores observados  $y_i$  y los valores verdaderos  $\xi_i$ , cuando se asume que los  $y_i$ 's son igualmente precisos. Este es el principio de los mínimos cuadrados, como se discute en la Sección 3.

Cuando los  $y_i$ 's tienen precisiones diferentes, medidas por sus errores estándares  $\sigma_i$ 's, el resultado son los mínimos cuadrados ponderados, también obtenidos por Gauss

$$Q = \sum \frac{1}{\sigma_i^2} (y_i - \xi_i)^2.$$

Si  $\xi = \mathbf{X}\theta$ , como en el modelo lineal de Gauss (2), entonces  $Q = (\mathbf{Y} - \mathbf{X}\theta)'(\mathbf{Y} - \mathbf{X}\theta)$ , las ecuaciones de los mínimos cuadrados son  $\mathbf{X}'\mathbf{X}\theta = \mathbf{X}'\mathbf{Y}$ , con la solución  $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ , como se dice en la Sección 3.

Gauss redujo la forma cuadrática  $Q = (\mathbf{Y} - \mathbf{X}\theta)'(\mathbf{Y} - \mathbf{X}\theta)$  a una suma de cuadrados de la forma  $Q = \sum d_{ii}^2 \alpha_i^2 +$  una constante, la cual es independiente de las  $\alpha$ 's, donde las  $\alpha$ 's son funciones lineales de las  $\theta$ 's. Por lo



tanto, las  $\alpha$ 's son variables aleatorias independientes normales con media cero, y el error estándar de  $\alpha_i$  es  $\sigma/d_{ii}$ . De estos hechos, se sigue que los  $\theta$ 's tienen una distribución normal alrededor del valor más probable  $\hat{\theta}$ , y el error estándar de  $\theta_i$  es  $\sigma\sqrt{a_{ii}}$ , donde  $a_{ii}$  es el elemento  $(i, i)$  de  $A^{-1}=(X'X)^{-1}$ . Los detalles matemáticos de los métodos y los resultados de Gauss pueden obtenerse de Seal (1967).

La interpretación de este resultado dado por Gauss es que la evidencia acerca de  $\theta_i$  contenida en todas las observaciones  $y_1, y_2, \dots, y_n$  es equivalente a la evidencia en la medida sola  $\theta_i$  con una precisión  $1/\sqrt{a_{ii}}$  veces la precisión de las observaciones originales (eso es, con un error estándar de  $\sqrt{a_{ii}}\sigma$ ).

## 5. Discusión

a) Aunque Legendre también consideró  $Y=X\theta+e$ , él se limitó al principio de los mínimos cuadrados como en la Sección 3. Según Seal (1967), el método de Gauss discutido anteriormente fue el primer tratamiento estadístico del problema. Esto significa, un tratamiento que usa explícitamente la distribución de la probabilidad de los errores  $e_i$ .

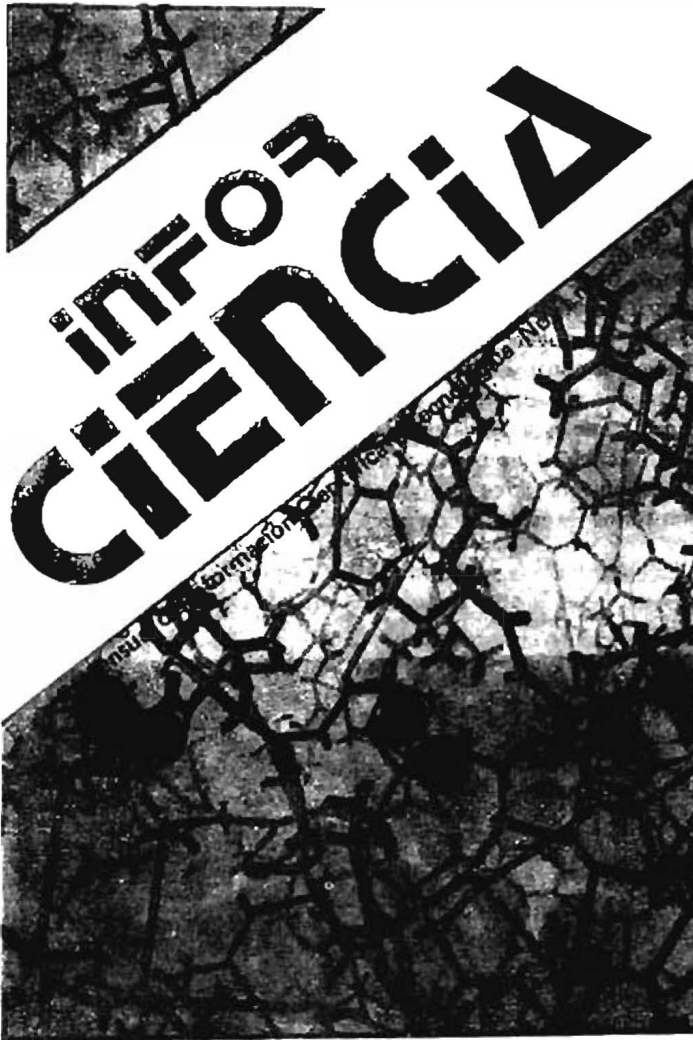
b) Hay una controversia extensa en torno al uso del teorema de Bayes (4). Esta se centra alrededor de la validez de asumir que cualesquiera dos valores de un parámetro  $\theta$  son igualmente probables (o de hecho asumir cualquier otra distribución de probabilidad de  $\theta$ ) cuando nada se conoce acerca de  $\theta$ . Esto es, la ignorancia acerca de  $\theta$  no puede tomarse como base para asumir una función de densidad de probabilidad  $f(\theta)$ , la cual produciría afirmaciones precisas e información exacta acerca de  $\theta$ . Sin embargo, el desarrollo de Gauss arriba discutido es ahora un método estándar de algunos libros de estadística.

Es interesante notar, sin embargo, que Gauss no dio a sus resultados una interpretación bayesiana. No habló de probabilidades ni de precisión del parámetro  $\theta_i$ , las cuales se obtendrían directamente de la distribución normal de  $\theta$  producida arriba. De hecho, dio a sus resultados una interpretación estándar de frecuencia cuando habló de la precisión del estimador  $\hat{\theta}_i$ , como se dijo al final de la Sección 4.

c) Los cálculos de Gauss se han convertido en el álgebra lineal estándar en los libros de texto de regresión. Una gran parte de la regresión lineal y los diseños experimentales, los cuales constituyen la base de una buena parte de la estadística moderna, dependen de la descomposición de  $Q$  en términos de sumas ortogonales de cuadrados.

d) El método de estimación definido por las ecuaciones (5) se llama hoy la estimación de máxima verosimilitud. Es el método de estimación que se usa más frecuentemente. Ver Edwards (1974) para la historia de máxima verosimilitud. Este método fue desarrollado por Fisher, por ejemplo, Fisher (1922, 1925, 1935). Ver Sprott (1984). Gauss derivó los mínimos cuadrados como un caso especial de máxima verosimilitud, apropiado cuando la distribución de los errores es normal, o, igualmente, cuando la media aritmética es el "mejor" estimador del parámetro de posición. Cuando esta distribución no es normal, varios ejemplos muestran que la lógica de arriba no produce los mínimos cuadrados, pero el método de máxima verosimilitud (5) es todavía aplicable. Tal vez el ejemplo más famoso es el obtenido cuando los errores tienen la distribución de Cauchy,

$$f(e) = \frac{1}{\pi} \frac{1}{1+e^2}, \quad e = y - \theta.$$



El consejo de Investigación y Estudios de Posgrado de esta universidad, órgano integrado por investigadores y profesores de posgrado de las diversas escuelas y departamentos de la misma, resolvió por unanimidad publicar un boletín que sirviera de canal de comunicación entre el Consejo y los universitarios de esta institución, de tal manera que éstos conocieran los acuerdos del propio Consejo, las actividades de la Secretaría de Investigación y Estudios de Posgrado y los resultados y estado actual de la investigación en la UAP.

Esta es una distribución simétrica como la distribución normal, pero con mayor concentración en las extremidades. Esta diferencia tiene consecuencias serias, una de las cuales es que el estimador de los mínimos cuadrados, la media aritmética  $\bar{y}$ , es casi el peor estimador posible de  $\theta$ , puesto que es equivalente en precisión a una sola observación.

Con la capacidad de Gauss de generalizar, es tal vez sorprendente el que no haya abandonado la suposición de la supremacía de la media aritmética y haya examinado otras posibilidades. La razón de ello es quizá que las distribuciones continuas, a excepción de la normal, no ocurrían entonces en la práctica. Más probablemente, fue porque más tarde abandonó completamente el método de arriba por un método no-paramétrico, el cual no impone ninguna suposición sobre  $f(e)$  otra que una media de cero y una varianza finita. Como dijo en una carta a Bessel (Plackett, 1972, p. 247), no le pareció tan importante determinar el valor de un parámetro desconocido para el cual la probabilidad es más grande, aunque infinitamente pequeña. (Una distribución continua de una variable aleatoria  $U$  asigna una probabilidad de cero a cada valor específico de  $U$ .) Le pareció más apropiado usar un método que mitiga tanto como sea posible los efectos malos de los errores  $e_i$  de las observaciones. Eso condujo a su segundo método de los mínimos cuadrados.



## 6. La teoría de Gauss: el segundo enfoque (1821, 1823a, 1826)

6.1. El segundo método de Gauss se incluye en tres artículos largos, “*Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*” (Parte I (1821), Parte 2 (1823a), Suplemento (1826)), que fueron presentados a la Sociedad Real de Göttingen. En es-

tos artículos Gauss abandonó el método inferencial de la Sección 4, que involucra “metafísica”, como lo describió en la carta a Bessel citada arriba, por un método basado en una teoría de la decisión.

Comenzó la Parte I comparando el problema de la estimación de un parámetro desconocido con un juego en el cual una pérdida se teme y no hay esperanza de ganancia. Cada error cometido se considera como una pérdida que se sufre. La indeseabilidad relativa de tal juego se mide por la pérdida esperada, esto es, por la suma de los productos de las varias pérdidas posibles con sus probabilidades respectivas. Definió como conveniente, aunque arbitrario, la pérdida como proporcional al cuadrado del error cometido. La pérdida esperada es entonces el error cuadrático medio.

Estas ideas son formuladas en la teoría de la decisión moderna. En la teoría de la decisión hay una función de pérdida, la pérdida esperada es la función de riesgo, y el objeto es hacer el riesgo mínimo. En el caso de los mínimos cuadrados, la función de pérdida es el cuadrado del error  $(\theta - \hat{\theta})^2$ , donde  $\hat{\theta}$  es cualquier estimador de  $\theta$ , y la función de riesgo es el error cuadrático medio,  $E(\theta - \hat{\theta})^2$ .

6.2 Antes de introducir el problema de los mínimos cuadrados, Gauss comenzó la Parte I con una discusión extensa de varios tópicos estadísticos, incluyendo las propiedades de medias, varianzas, propagación del error, etc., la mayoría de las cuales aparecen en los libros de texto elementales. Notable entre estos resultados es la desigualdad

$$\begin{aligned} \Pr(|x - \theta| \geq \lambda\sigma) &\leq 1 - \lambda/\sqrt{3}, \quad \lambda \leq 2/\sqrt{3}, \\ &\leq 4/9\lambda^2, \quad \lambda \geq 2/\sqrt{3}, \end{aligned}$$

válida para cualquier distribución simétrica con una sola moda (y media)  $\theta$ . Esta desigualdad con  $\lambda = 2$  implica que la probabilidad de que  $x$  caiga en una región de dos errores estándares de  $\theta$  es por lo menos .89. Alternativamente, para cualquier valor de  $x$  observado, un intervalo del 89% de confianza de  $\theta$  sería a lo más  $x \pm 2\sigma$ .

6.3 Regresando a los mínimos cuadrados, el modelo es el modelo lineal de Gauss (2),  $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$ , donde  $\mathbf{X}$ , como antes, se asume conocida sin error, y los errores  $e_i$  son variables aleatorias independientes con media 0 y varianza  $\sigma^2$ . Aunque Gauss no empleó ni la notación ni las propiedades de las matrices, los cálculos se simplifican y acortan mediante su uso.

Para estimar  $\boldsymbol{\theta}$ , Gauss asumió que los errores  $e_i$  eran suficientemente pequeños de modo que sus cuadrados y potencias más altas eran despreciables. Por lo tanto, sin pérdida de generalidad, limitó consideración a los estimadores  $\hat{\boldsymbol{\theta}}$  que son funciones lineales de las observaciones,  $\hat{\theta}_j = \sum_{i=1}^n c_{ji}y_i$ ,  $j = 1, 2, \dots, k$ , o en la notación matricial,  $\hat{\boldsymbol{\theta}} = \mathbf{C}\mathbf{Y}$ .

También, cuando las observaciones son sin errores,  $e \equiv 0$ , las ecuaciones  $Y = X\theta$  deben ser consistentes, así la solución es el valor verdadero de  $\theta$  (Sección 2). Por lo tanto, se requiere que  $\hat{\theta} = CY = CX\theta \equiv \theta$ , lo cual implica la condición  $CX = I$ , donde  $I$  es la matriz de identidad  $k \times k$ .

El teorema clásico de Gauss sobre los mínimos cuadrados es entonces:

*Teorema:* Entre todos los estimadores  $\hat{\theta}$  de  $\theta$  los cuales son funciones lineales de las observaciones,  $\hat{\theta} = CY$ , tal que  $CX = I$ , donde  $I$  es la matriz de identidad  $k \times k$ , el error cuadrático medio es un mínimo cuando  $C = (X'X)^{-1} X'$ , esto es, cuando  $\hat{\theta} = \hat{\theta} = (X'X)^{-1} X'Y$ , el estimador de los mínimos cuadrados.

*Prueba:* Los errores cuadráticos medios requeridos,  $E(\hat{\theta}_j - \theta_j)^2$ , son los elementos diagonales de la matriz  $E(\hat{\theta} - \theta)(\hat{\theta} - \theta)'$ . Ponga  $A = X'X$ ,  $B = A^{-1} X'$ . Puesto que  $CX = I$ , se sigue que  $E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' = CC'\sigma^2$ , y  $(C - B)B' = 0$ , por lo tanto  $CC' = BB' + (C - B)(C - B)'$ . Pero los elementos diagonales de una matriz de la forma  $(C - B)(C - B)'$  son sumas de cuadrados. Así, las sumas de los cuadrados requeridos  $CC'$  serán un mínimo cuando  $C - B = 0$ , por lo tanto  $C = B = A^{-1} X' = (X'X)^{-1} X'$ . Esto da  $\hat{\theta} = \hat{\theta} = (X'X)^{-1} X'Y$ , el estimador de los mínimos cuadrados. Además, los errores cuadráticos medios son los elementos diagonales de  $CC'\sigma^2 = A^{-1} X'X A^{-1} \sigma^2 = A^{-1} \sigma^2$ , como en la Sección 4.



6.4 En la parte 2, Gauss expresó el parámetro  $\theta$  como  $\theta = \hat{\theta} - A^{-1} X'e$ , donde  $\hat{\theta}$  es el estimador de los mínimos cuadrados de  $\theta$ . Esto se sigue de  $\hat{\theta} = A^{-1} X'Y = A^{-1} X'(X\theta + e)$ , y produce  $E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' = A^{-1} X'E(ee')XA^{-1} = A^{-1} \sigma^2$ , de modo que la covarianza entre  $\hat{\theta}_i$  y  $\hat{\theta}_j$  es  $a_{ij}\sigma^2$ , el elemento  $(i, j)$  de  $A^{-1} \sigma^2$ , aunque no le dio un nombre. De esto, mostró que el estimador de los mínimos cuadrados de una función lineal de los parámetros,  $\alpha = \Sigma g_i \theta_i$ , o en la notación

matricial,  $\alpha = G'\theta$ , es la misma función lineal de los estimadores de los mínimos cuadrados,  $\hat{\alpha} = G'\hat{\theta}$ , con la varianza  $\sigma_{\hat{\alpha}}^2 = \Sigma \Sigma g_i g_j a_{ij} \sigma^2 = G'A^{-1} G \sigma^2$ . Esto extiende el dominio de aplicación de los mínimos cuadrados a todas las funciones lineales de los parámetros. Seal (1967) dijo que este resultado se pensó que era nuevo en 1938, (David y Neyman, 1938), donde se describió como una generalización del teorema de Markov (ver Sección 8.3.4).

La ecuación  $\theta = \hat{\theta} - A^{-1} X'e$  expresa el parámetro  $\theta$  explícitamente como una función de la variable aleatoria  $e$ . Aunque Gauss no hizo más uso de ella, es posible considerar a  $\theta$  como una variable aleatoria, sobre la cual pueden ser hechas afirmaciones de probabilidad sin el uso del teorema de Bayes (Secciones 4, 5). Esto conduce a los conceptos de la probabilidad fiducial, (Fisher, 1973), y la probabilidad estructural (Fraser, 1968).

6.5 Gauss entonces derivó algunos resultados sobre la suma de las desviaciones cuadráticas,  $Q = (Y - X\theta)'(Y - X\theta)$  de la Sección 4. El estimador de los mínimos cuadrados  $\hat{\theta}$  hace  $Q$  un mínimo. El mínimo resultante es  $Q_m = (Y - X\hat{\theta})'(Y - X\hat{\theta}) = Y'(Y - X\hat{\theta})$ ;  $Q_m$  se llama la suma residual o suma de error de los cuadrados.

Si la componente  $\theta_1$  del vector  $\theta$  se mantiene constante mientras las componentes  $\theta_2, \theta_3, \dots, \theta_k$  varían, entonces  $Q$  puede tomar un valor mínimo relativo  $Q_r$ . Gauss mostró que si  $(Q_r - Q_m) \leq c^2$ , entonces  $(\hat{\theta}_1 - \theta_1)^2 \leq a^{11} c^2 = c^2 \sigma_{\hat{\theta}_1}^2 / \sigma^2$ , donde  $\sigma_{\hat{\theta}_1}^2$  es la varianza de  $\hat{\theta}_1$ , el estimador de los mínimos cuadrados de  $\theta_1$ . También, si  $\alpha = \Sigma g_i \theta_i = G'\theta$  se mantiene constante, y en otros aspectos los  $\theta$ 's se dejan variar, el mínimo relativo correspondiente  $Q_r$  de  $Q$  es tal que  $(Q_r - Q_m) \leq c^2$  implica  $(\alpha - \hat{\alpha})^2 \leq c^2 (G'A^{-1}G) = c^2 \sigma_{\hat{\alpha}}^2 / \sigma^2$  de la Sección 6.4. Gauss no pareció usar estos resultados, pero forman la base de pruebas de la significación moderna en regresión. De hecho, para probar la significación de los valores específicos de un subconjunto de los  $\theta$ 's, el método estándar consiste en calcular la razón  $(Q_r - Q_m)/Q_m$  correspondiente. Bajo el modelo lineal con errores  $e_i$  normales, ésta es proporcional a una razón  $F$  cuyo valor numérico determina el nivel de la significación.

6.6 Supongamos que un valor adicional  $y_{n+1}$  se observa, para el cual el vector de los valores de  $x$  es  $x = (x_{(n+1)1}, x_{(n+1)2}, \dots, x_{(n+1)k})$ . Gauss mostró que el estimador de los mínimos cuadrados basado en todas las  $(n + 1)$  observaciones es

$$\theta^* = \hat{\theta} - [M(x\hat{\theta} - y_{n+1}) / (1 + w)],$$

donde  $M = A^{-1} x'$ ,  $w = x A^{-1} x' = x M$  (que es una cantidad escalar), y  $\hat{\theta}$  es el estimador de los mínimos cuadrados basado en las  $n$  observaciones originales. Mostró también que la matriz de covarianza de  $\theta^*$  es

$$A^* \sigma^2 = [A^{-1} - M M' / (1 + w)] \sigma^2,$$

y la nueva suma mínima de los cuadrados es

$$Q_m' = Q_m + (\mathbf{x}\hat{\boldsymbol{\theta}} - y_{n+1})^2 / (1 + w).$$

Estas fórmulas permiten la incorporación de una nueva observación sin tener que calcular de nuevo el estimador, su varianza, y la suma mínima resultante de los cuadrados del conjunto complejo de las ecuaciones de los mínimos cuadrados. Esto es, los resultados  $\hat{\boldsymbol{\theta}}$ ,  $\mathbf{A}^{-1}$ , y  $Q_m$  de los cálculos ya hechos con las  $n$  observaciones originales pueden usarse en las fórmulas simples de arriba para obtener las cantidades revisadas  $\boldsymbol{\theta}'$ ,  $\mathbf{A}^{-1}$ , y  $Q_m'$  basadas en todas las observaciones, sin tener que invertir otra vez una matriz.

Estos resultados forman la base de los mínimos cuadrados recursivos. Fueron obtenidos por Young (1974, ecuación 11), quien los atribuyó a Plackett (1950). Plackett consideró el caso general donde las observaciones adicionales ocurren en conjuntos de  $s > 1$ , y citó a Gauss por haber presentado las fórmulas necesarias mencionadas arriba. Young (1974) también describe las aplicaciones de los métodos recursivos al problema de la estimación de la órbita de la Misión Apolo. Según Young, estos métodos recursivos, no sólo iniciados sino desarrollados considerablemente por Gauss, juegan un papel importante en las misiones espaciales de hoy.

6.7 Las últimas secciones de la Parte 2 tratan de la estimación de la cantidad  $\sigma$  que ocurre en todas las fórmulas anteriores. El método usualmente empleado hasta entonces conduciría a  $\sqrt{Q_m/n}$  como el estimador de  $\sigma$ . Gauss notó que  $\sqrt{Q_m/n}$  es demasiado pequeño como un estimador de  $\sigma$  puesto que, si  $\boldsymbol{\theta}$  fuera conocido, el estimador de  $\sigma$  sería  $\sqrt{Q/n}$ , y  $Q_m < Q$ , a menos que  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ . Mostró que  $E(Q_m) = (n - k)\sigma^2$ , y así recomendó usar  $\hat{\sigma} = \sqrt{Q_m/(n - k)}$  como el estimador de  $\sigma$ . También calculó su error estándar, y notó que cuando los  $e_i$ 's tienen la distribución normal, este error estándar es el error estándar de la suma de  $(n-k)$  errores independientes de  $e_i$ . Esto se relaciona con los resultados distribucionales en la teoría de la regresión donde los errores  $e_i$  se asumen tener una distribución normal.

6.8 El Suplemento trata de la estimación de los mínimos cuadrados cuando hay restricciones sobre los parámetros. Específicamente, Gauss asumió que  $\mathbf{X} = \mathbf{I}_n$ , de modo que hay  $n$  parámetros  $\xi_i = \theta_i$ ,  $n$  observaciones  $y_i = \xi_i + e_i$ , y  $r$  restricciones lineales  $\sum_{i=1}^n f_{ki}\theta_i = 0$ ,  $k = 1, 2, \dots, r$ , o en la notación matricial,  $\mathbf{F}\boldsymbol{\theta} = 0$ . Entonces, entre las funciones lineales, el estimador  $\hat{\boldsymbol{\alpha}} = \sum a_i \theta_i^*$  es el estimador con el error cuadrático medio mínimo de  $\boldsymbol{\alpha} = \sum g_i \theta_i$  si los  $\theta_i^*$ 's son los estimadores de los mínimos cuadrados de los  $\theta_i$ 's bajo las restricciones  $\mathbf{F}\boldsymbol{\theta} = 0$ . Además, el estimador correspondiente de  $\sigma$  es  $\hat{\sigma} = \sqrt{Q_m/r}$ .

Este método permite fácilmente tratar con el modelo general  $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$  de las secciones anteriores

bajo las restricciones  $\mathbf{F}\boldsymbol{\theta} = 0$ . Eso forma la base del análisis de varianza discutido hoy en los diseños experimentales.

## 7. Otras contribuciones de Gauss (1803-1809, 1816, 1823b, 1824)

En la segunda de estas contribuciones (1816), él determinó el estimador de máxima verosimilitud  $\hat{h}$  de la precisión  $h = 1/\sqrt{2}\sigma$  de la distribución normal. Obtuvo los intervalos de probabilidad de 50% del error probable,  $.6744897\sigma$ , mediante el uso de una aproximación normal a la distribución de  $S_k = \sum_{j=1}^k e_j^2$ ,  $k = 1, 2, \dots, 6$ . El artículo es interesante particularmente porque mostró que el estimador obtenido en  $S_2$  es el más preciso en el sentido de que produce el intervalo de 50% más corto. Con  $S_2$ , 100 errores observados  $e_i$  producen la misma precisión (esto es, el intervalo de probabilidad de 50% de la misma anchura) que 114 errores observados con  $S_1$ , 109 con  $S_3$ , 103 con  $S_4$ , 178 con  $S_5$ , y 251 con  $S_6$ .

Fisher (1920) redescubrió estos hechos, y aisló la propiedad de la suficiencia, mediante la cual  $S_2$  puede decirse que contiene toda la información de la muestra sobre  $\sigma$  en la distribución normal, es decir,  $S_2$  es suficiente para la estimación de  $\sigma$  en la distribución normal. En el mismo caso (Sección 4), el estimador de los mínimos cuadrados  $\hat{\boldsymbol{\theta}}$ , que es el estimador de máxima verosimilitud, es suficiente para  $\boldsymbol{\theta}$ , y por lo tanto contiene toda la información de la muestra sobre  $\boldsymbol{\theta}$ .

El último (1824) de los artículos de arriba es interesante como un ejemplo de los mínimos cuadrados ponderados. Los errores estándares  $\sigma_i$  de los errores  $e_i$  se asumieron proporcionales a la raíz cuadrada del tiempo que transcurre entre dos observaciones en el problema de determinar la longitud mediante un cronómetro.



## 8. Discusión

8.1 La única manera en que los libros de texto de estadística agregan algo a la teoría anterior de Gauss es en las pruebas específicas de la hipótesis lineal  $C\theta = 0$ , y en la teoría exacta correspondiente de las distribuciones asociadas con errores que tienen la distribución normal, esto es, las distribuciones  $t$  de Student,  $F$ , y  $\chi^2$ .

8.2 Vale la pena hacer notar las diferencias en la generalidad entre los dos métodos de las Secciones 4 y 6 de los mínimos cuadrados. El primero es más restrictivo por el hecho de que los errores  $\epsilon_i$  deben tener la distribución normal, pero a su vez tiene un marco más amplio puesto que permite la consideración de cualesquiera funciones  $\xi_i(\theta)$ . El último método es más general por el hecho de que los errores pueden tener cualquier distribución  $f(\epsilon)$  con media cero y varianza finita, pero es más limitado puesto que restringe la atención a las funciones  $\xi = X\theta$  que son lineales (aproximadamente) en  $\theta$ .

8.3 La presentación moderna de la estimación difiere de la de Gauss discutida anteriormente en varios aspectos desafortunados.

8.3.1 Gauss justificó la restricción a los estimadores lineales mediante la suposición de que los errores eran suficientemente pequeños de modo que sus cuadrados y potencias más altas pudieran ignorarse. Las aplicaciones de Gauss fueron en los tópicos de la astronomía y en la geodesia donde las medidas tienen un alto grado de precisión. Al contrario, los libros de texto de estadística modernos utilizan frecuentemente los estimadores lineales solo por conveniencia matemática y no por causa de las propiedades científicas de las observaciones.

8.3.2 Gauss subrayó que el criterio del error cuadrático medio es arbitrario. Pero, note que bajo la condición de que los errores  $\epsilon_i$  sean suficientemente pequeños, de modo que los estimadores pueden ser lineales, entonces cualquier función de pérdida puede aproximarse localmente por el error cuadrático medio. Sin embargo, muchas presentaciones modernas ponen demasiado énfasis en el criterio del error cuadrático medio, elevándolo a la categoría de un concepto básico en la teoría de la estimación estadística. Ver, por ejemplo, Berkson (1980).

8.3.3 La condición  $CX = I$  de la Sección 6.3 es una condición de *consistencia*, y puede llamarse la condición de *consistencia del error*. Esto la distingue de otros criterios de consistencia, como el de consistencia de Fisher. El objeto de los criterios de consistencia es identificar el parámetro que se estima. Esto es,  $\hat{\theta}$  debe estimar  $\theta$ , y no alguna otra función  $g(\theta)$ . Por ejemplo,  $s = \sqrt{\sum(x_i - \bar{x})^2/n}$  estima  $\sigma$ , y no  $\sigma^2$  ó  $1/\sigma$ ;  $1/\sigma$  se estima por  $1/s$ , y no por  $s$  ó  $s^2$ .

Sin embargo, la condición  $CX = I$  se interpreta usualmente para implicar que Gauss buscaba los estimadores *insesgados*,  $E(\hat{\theta}) = \theta$ . Puesto que  $CX = I$  im-

plica que el estimador  $\hat{\theta}$  es insesgado ( $E(\hat{\theta}) = E(CY) = E(CX\theta + Ce) = CX\theta = \theta$ ), el estimador de los mínimos cuadrados es por cierto insesgado. Pero, la necesidad de insesgamiento no es en general razonable, porque hay ejemplos donde todos los valores que toma el estimador insesgado *único* son imposibles. Por ejemplo, suponga que artículos tales como bulbos electrónicos fallan de acuerdo a un proceso de Poisson a razón  $\lambda$ , y que 3 fallas se han observado en una unidad de tiempo. Entonces, el estimador insesgado *único* de  $\theta = \exp(-\lambda)$ , el cual es la probabilidad de que un artículo especificado tenga una duración mayor de 4 unidades de tiempo, es  $-27$ . Puesto que cualquier probabilidad debe estar entre 0 y 1, este estimador toma un valor imposible y por lo tanto es absurdo. Además, puesto que no hay otro estimador insesgado, esta dificultad es solamente el resultado de requerir que el estimador sea insesgado. Tales ejemplos no son raros.

El que Gauss no buscara los estimadores insesgados parece evidente por el hecho de que el estimador  $\hat{\sigma}$  de  $\sigma$  de la Sección 6.7 sugerido por Gauss no es insesgado.

8.3.4 La teoría presentada anteriormente se conoce usualmente como la teoría de Gauss-Markov. Pero parece que Markov no produjo nada nuevo sobre los mínimos cuadrados. Su nombre se asocia con la teoría aparentemente porque su prueba de 1912 fue pensada como original por Neyman (1934), ver Plackett (1949), Seal (1967), aunque en la discusión del artículo de Neyman, Fisher indicó la prioridad de Gauss.



28.4.66 47

8.4 Hay varios desarrollos modernos en la teoría de Gauss que tienen utilidad dudable. Uno es la estimación mediante el error cuadrático medio mínimo. Otro es la teoría de los mejores estimadores lineales insesgados o BLUES (en inglés, *best linear unbiased estimates*). Y otro es la teoría de la estimación insesgada con varianza uniformemente mínima o UMVU (en inglés, *uniformly minimum variance unbiased*). Las dificultades con éstos son las siguientes.

a) No es posible en general minimizar el error cuadrático medio, u otra función de pérdida, cuando se dan solamente la función de pérdida, las observaciones, y su distribución, Barnard (1977). Por lo tanto, una teoría basada solamente en el error cuadrático medio mínimo no puede obtenerse. Esta es tal vez la razón de que las condiciones de linealidad o insesgamiento se imponen, las cuales conducen a los estimadores BLUES y UMVU.

b) La necesidad de la linealidad o el insesgamiento es en general demasiado restrictiva, puesto que puede dar lugar a estimaciones que toman valores imposibles, como se ejemplificó anteriormente.

c) Si  $\hat{\theta}$  es un estimador insesgado de  $\theta$ ,  $g(\hat{\theta})$  no es en general un estimador insesgado de  $g(\theta)$ , a menos que  $g(\theta)$  sea lineal en  $\theta$ .

d) La varianza no es una medida adecuada de la precisión de un estimador  $\hat{\theta}$ , a menos que  $\hat{\theta}$  tenga la distribución normal, Fischer (1935, pp. 46, 82; 1973, p. 158). Ver Sprott y Viveros (1984, 1985).

8.5 Los desarrollos más importantes de la teoría de Gauss parecerían ser la teoría de la decisión y la teoría de la estimación de R.A. Fisher. Estas son discutidas por Sprott y Viveros (1984). En el método de máxima verosimilitud (Sección 5.4) los dos objetivos de Gauss son, en cierto sentido, reconciliados, Fisher (1936, p. 249). En lugar de maximizar las probabilidades de  $\theta$ , todas las cuales son infinitamente pequeñas, como Gauss objetó (Sección 5.4), máxima verosimilitud maximiza una nueva medida de incertidumbre llamada verosimilitud, la cual no es infinitamente pequeña. Además minimiza, por lo menos asintóticamente, el error cuadrático medio (o varianza). Ver Sprott (1984). Como se discutió anteriormente, cuando los errores tienen la distribución normal, la estimación de máxima verosimilitud es también la estimación de los mínimos cuadrados (Secciones 5.4, 7); así cuando las funciones son lineales, maximiza exactamente los errores cuadráticos medios. Tal método conduce a las consideraciones de la consistencia (Sección 8.3.3), la suficiencia (Sección 7), la verosimilitud (ver Sprott 1984), y los conceptos relacionados, en lugar del sesgo, la varianza, y los errores cuadráticos medios.

Un desarrollo reciente es el análisis del modelo lineal de Gauss donde los errores  $e_i$  se supone que tienen cualquier distribución *específica*. Esto se relaciona con la inferencia condicional y los estadísticos auxiliares, métodos y conceptos desarrollados



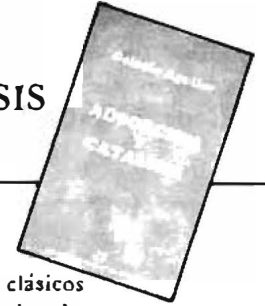
Angelo Altieri Megale

## LOS PRESOCRATICOS

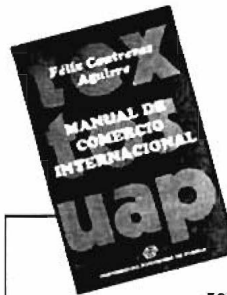
Cualquiera que haya intentado enseñar filosofía griega presocrática en los propios fragmentos, sabe que la comparación entre distintas versiones constituye una inestimable ayuda para desarrollar en los estudiantes la clara conciencia de que el mundo griego es vertido al español de muy diversas maneras. Esta es una razón adicional para ver en esta nueva versión un aporte más a la comprensión del pensamiento de los fundadores de la filosofía.

Gelasio Aguilar

## ADSORCION Y CATALISIS



El objetivo fundamental del libro consiste en describir en forma accesible las teorías básicas del fenómeno de la adsorción y los métodos experimentales clásicos para el estudio de las propiedades de los absorbentes y catalizadores. Asimismo, examinar el fenómeno catalítico sobre los cuerpos sólidos (catálisis heterogénea).



Félix Contreras Aguirre

## MANUAL DE COMERCIO INTERNACIONAL

Por su expresión didáctica, éste es un libro destinado a cubrir no sólo los requerimientos de la academia sino también las inquietudes del lector no especializado. Están aquí desde los conceptos más elementales hasta los pasos específicos del comercio exterior: distintos códigos tanto para el comprador como para el vendedor, cuadros y ejemplos ilustrativos, etc.

Jorge Bouton

## EL SINTOMA Y EL SIGNO

2 Tomos



Esta obra, destinada a estudiantes de medicina del nivel clínico, es esencialmente didáctica para la enseñanza- aprendizaje de la clínica cotidiana. Ha sido escrita prácticamente sin bibliografía y sobre la base de una larga experiencia de 30 años en la docencia médica.

Toma como base la clínica —el signo y el síntoma— como "lenguaje fenoménico" de la enfermedad y su cronología, ordenación y jerarquización como orientación diagnóstica.

por Fisher (1934; 1973, pp. 165-169). Los detalles pueden encontrarse en Barnard y Sprott (1977), y también desde un punto de vista diferente, en Fraser (1968).

## 9. Conclusión

Una medida de la amplitud y la influencia de las contribuciones de Gauss a la estadística es el hecho de que sus contribuciones han conducido a la teoría de la decisión estadística, Wald (1971); al diseño y análisis experimentales, Fisher (1966); y a una teoría de la estimación estadística, Fisher (1973). Aunque la aplicación práctica de la teoría de la decisión estadística tal vez no es muy dramática, las aplicaciones prácticas de los diseños experimentales y de la teoría de la estimación estadística desarrollados por

Fisher están fuera de duda. Los diseños experimentales de Fisher explotan totalmente las propiedades del modelo lineal de Gauss y los mínimos cuadrados. Ambos, con la teoría general de la inferencia estadística de Fisher, forman una estructura lógica muy interesante de la experimentación y la inferencia estadística que tiene un uso práctico inmenso. Es muy importante subrayar que las investigaciones teóricas de Gauss fueron respuestas a problemas prácticos. Por lo tanto, todas sus contribuciones a la estadística muestran una mezcla entre la teoría y la aplicación la cual es poco vista en nuestros días.

## Reconocimientos

Quiero agradecer al Dr. R. Viveros por leer las versiones preliminares de este artículo, y por sugerencias provechosas.

## Referencias

- 1 Aczel, J. (1976), *Lectures on functional equations and their applications*, Academic Press, New York.
- 2 Barnard, G.A. (1977), "On ridge regression and the general principles of estimation", *Utilitas Math*, núm. 11, pp. 299-311.
- 3 Barnard, G.A. y Sprott, D.A. (1977), "Theory of estimation", *University of Waterloo Technical Reports*.
- 4 Bayes, T. (1763), "Essay towards solving a problem in the doctrine of chances", *Phil. Trans. Roy. Soc.* núm. 53, pp. 370-418. Reproducido con una nota bibliográfica por G.A. Barnard en *Biometrika* núm. 59 (1958), pp. 293-315.
- 5 Berkson, J. (1980), "Minimum chi-square, not maximum likelihood" (con discusión), *Ann. Statist.* núm. 8, pp. 457-487.
- 6 Bertrand, J. (1855), *Méthode des moindres carrés. Mémoires sur la combinaison des observations, par Ch. Fr. Gauss*. Traduits en français et publiés avec l'autorisation de l'auteur. Mallet-Bachelier, Imprimeur-Libraire de l'École Polytechnique, du Bureau des Longitudes, Quai des Augustins, núm. 55, Paris.
- 7 David, F. N. y Neyman, J. (1938), "Extension of the Markov theorem on least squares", *Statist. Res. Mem.* núm. 2, pp. 105-116.
- 8 Edwards, A.W.F. (1974), "The history of likelihood", *Int. Statist. Rev.* núm. 42, pp. 9-15.
- 9 Fisher, R.A. (1920), "A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error", *Monthly Notices Roy. Astron. Soc.* núm. 80, pp. 758-770.
- 10 Fisher, R.A. (1922), "On the mathematical foundations of theoretical statistics", *Phil. Trans. Roy. Soc. A* núm. 222, pp. 309-368.
- 11 Fisher, R.A. (1925), "Theory of statistical estimation", *Proc. Cambridge Phil. Soc.* núm. 22, pp. 700-725.
- 12 Fisher, R.A. (1934), "Two new properties of mathematical likelihood", *Proc. Roy. Soc. London Ser. A* núm. 144, pp. 285-307.
- 13 Fisher, R.A. (1935), "The logic of inductive inference" (con discusión), *Jour. Roy. Statist. Soc.* núm. 98, pp. 39-82.
- 14 Fisher, R.A. (1936), "Uncertain inference", *Proc. Amer. Acad. Arts and Sciences*, núm. 71, pp. 245-258.
- 15 Fisher, R. A. (1959), "Natural selection from the genetical standpoint", *Aust. J. Sci.* núm. 22, pp. 16-17.
- 16 Fisher, R.A. (1966), *Design of experiments*. Oliver and Boyd, 8th ed., Edinburgh.
- 17 Fisher, R.A. (1973), *Statistical methods and scientific inference*, Hafner Press, New York.
- 18 Fraser, D.A.S. (1968), *The structure of inference*, Wiley, New York.
- 19 Gauss, C.F. (1809), *Theoria Motus Corporum Coelestium*, *Werke* 7, traducido al inglés por C.H. Davis (1963) Dover, New York.
- 20 Gauss, C.F. (1821, 1823a, 1826), "Theoria combinationis observationum erroribus minimis obnoxiae", Partes 1, 2, y Suplemento, *Werke* 4, pp. 1-108.
- 21 Gauss, C.F. (1803-1809), "Disquisitiones de elementis ellipticis Palliis", *Werke* 6, pp. 1-24.
- 22 Gauss, C.F. (1816), "Bestimmung der Genauigkeit der Beobachtungen", *Werke* 4, pp. 109-117.
- 23 Gauss, C.F. (1823b), "Anwendungen der Wahrscheinlichkeitsrechnung auf eine Aufgabe der praktischen Geometrie", *Werke* 9, pp. 231-237.
- 24 Gauss, C.F. (1824), "Chronometrische Längenbestimmungen", *Astronomische Nachrichten* núm. 4, 227.
- 25 Neyman, J. (1934), "On two different aspects of the representative method" (con discusión), *J. Roy. Statist. Soc.* núm. 97, pp. 558-625.
- 26 Plackett, R.L. (1949), "A historical note on the method of least squares", *Biometrika*, núm. 36, pp. 458-460.
- 27 Plackett, R.L. (1950), "Some theorems on least squares", *Biometrika* núm. 37, pp. 149-157.
- 28 Plackett, R.L. (1972), "Studies in the history of probability and statistics XXIX. The discovery of the method of least squares", *Biometrika*, núm. 59, pp. 239-251.
- 29 Seal, H.L. (1967), "Studies in the history of probability and statistics XV. The historical development of the Gauss linear model", *Biometrika* núm. 62, pp. 1-24.
- 30 Sprott, D.A. (1978), "Gauss's contributions to statistics", *Historia mathematica* núm. 5, pp. 183-203.
- 31 Sprott, D.A. (1983), "Gauss, Carl Friedrich", *Encyclopedia of statistical sciences*, vol. 3, pp. 305-309. Kotz y Johnson, eds. Wiley & Sons, New York.
- 32 Sprott, D.A. (1984), "Likelihood and maximum likelihood estimation", *Mathematical Reports of the Academy of Science, Royal Society of Canada* núm. 6, pp. 225-242.
- 33 Sprott, D.A. y Viveros, R. (1984), "The interpretation of maximum likelihood estimation", *Canad. Jour. Statist.* núm. 12, pp. 27-38.
- 34 Sprott, D.A. y Viveros, R. (1985), "The estimation of ratios and related quantities", *Contributions to Economics and Statistics Today*, pp. 242-251. Schneeweiss y Strecker, eds. Springer-Verlag, Berlin.
- 35 Trotter, H.F. (1957), "Gauss's work on least squares", based on the french translation by Bertrand (1855) authorized and approved by Gauss, (Fotocopiado), *Dept. Sci. & Tech., Firestone Library, Princeton Univ.*
- 36 Wald, A. (1971), *Statistical decision functions*, Chelsea Pub. Co., New York.
- 37 Young, P. (1974), "Recursive approaches to time series analysis", *Bull. Inst. Math. and its Application* núm. 10, pp. 209-224.