Int. Stat. Rev., Vol. 42, No. 2, 1974, pp. 147–174/*Longman Group Ltd/Printed in Great Britain*

The Method of Least Squares and Some Alternatives—Part I

H. Leon Harter

Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio, USA

Summary

A very important problem in mathematical statistics is that of finding the best linear or non-linear regression equation to express the relation between a dependent variable and one or more independent variables. Given are observations, each subject to random error, greater in number than the parameters in the regression equation, on the dependent variable and the related values of the independent variable(s), which may be known exactly or may also be subject to random error. Related problems are those of choosing the best measures of central tendency and dispersion of the observations. The best solutions of all three problems depend upon the distribution of the random errors. If one assumes that the values of the independent variable(s) are known exactly and that the errors in the observations on the dependent variable are normally distributed, then it is well known that the mean is the best measure of central tendency, the standard deviation is the best measure of dispersion and the method of least squares is the best method of fitting a regression equation. Other assumptions lead to different choices. Most practitioners have tended to make the assumption of normality and not to worry about the consequences when it is not justified. Another problem arises when the data are contaminated by spurious observations (outliers) which come from distributions with different means and/or larger standard deviations. Many methods have been proposed for rejecting outliers or modifying them (or their weights). After summarizing (chronologically) the voluminous literature on measures of central tendency and dispersion, the method of least squares and numerous alternatives, the treatment of outliers and robust estimation, the author recommends a simple and reasonably robust set of procedures.

1. Introduction

Since very early times, people have been interested in the problem of choosing the best single value (average or mean) to summarize the information given by a number of independent observations or measurements, each subject to error, of the same quantity. Eisenhart (1971) presents evidence of the use of such averages as the mode (the value occurring most frequently) and the midrange (the value midway between the largest and smallest observations) by the ancient Greeks and Egyptians and by the Arabs during the Middle Ages. The median (the value such that there are the same number of observations above as below it) and the arithmetic mean (the sum of all the observations divided by their number) seem not to have come into use until early in the modern era.

The problem of determining the constants in the equation of the straight line which best fits (in some specified sense) three or more non-collinear points in the (x, y) plane whose coordinates are pairs of associated values of two related variables, x and y, dates back at least as far as Galileo Galilei (1632). This problem can be generalized in two ways: (1) instead of finding the best linear equation in two variables (best line in a plane), one may wish to find the best linear equation in three variables (best plane in three-dimensional space) or in more than three variables (best hyperplane in a hyperspace); (2) one may drop the requirement that the equation be linear, and find the best curve in a plane, the best surface in three-dimensional space, or the best hypersurface in a hyperspace. Statisticians speak of these problems as those of linear and non-linear regression.

The problems of determining the best average or measure of central tendency and the best linear or non-linear regression equation are related to each other and to the problem of choosing the best measure of variability or dispersion. The solutions of all three problems depend upon the distribution of the errors or residuals (deviations of the observed values from those predicted by the regression equation). The body of statistical theory which treats all these related problems is called the theory of errors. In the following sections we shall trace the development of the theory of errors from the time of Galileo to the present day. We shall see that, almost from the time early in the nineteenth century when it was first proposed, the method of least squares has enjoyed a pre-eminence over other methods in the theory of errors. We shall examine the question as to the conditions under which this pre-eminence is deserved and when other methods are theoretically superior to the method of least squares.

2. Pre-Least-Squares Era (1632–1804)

Galileo Galilei (1632) considers the question of determining the distance from the earth of a new star, given observations on its maximum and minimum elevation (in degrees) and the elevation of the pole star by 13 observers at different points on the earth's surface. If the observations were exact, the distance could be determined from the observations of any two observers, and the 78 determinations made by pairing the observers in all possible ways would all give the same result. Since the observations are subject to error, 78 different distances of the star from the centre of the earth are found, ranging from a value less than the radius of the earth to infinity and beyond. Both extremes are manifestly impossible. Galileo states (p. 290 of the English translation): "Then these observers being capable, and having erred for all that, and their errors needing to be corrected for us to get the best possible information from their observations, it will be appropriate for us to apply the minimum amendments and smallest corrections that we can – just enough to remove the observations from impossibility and restore them to possibility. . . ." In this statement we see the beginnings of the theory of errors, which attempts to determine the truth from inconsistent observations by minimizing various non-decreasing functions of the errors.

Roger Cotes (1722), in his last paragraph (p. 22), considers four observations p, q, r and s, which may not be equally reliable, of the position of a point. He proposes, as the most probable true position, a weighted average with weights P, Q, R and S which are inversely proportional to the spread of errors to which the respective observations are subject. This proposal represents one of the earliest attempts to determine an average which uses all the observations but does not assign equal weights to all of them.

Leonhard Euler (1749) and Johann Tobias Mayer (1750), working independently, developed what has come to be known as the Method of Averages for fitting a linear equation to observed data. In this method the observational equations are divided into as many subsets as there are coefficients to be determined, the division being made according to the values of (one of) the independent variable(s), those having the largest values of this variable being grouped together, then the next largest in another group, etc. Then the equations in each group are added together, which is equivalent to applying to each subset the condition of zero sum of residuals inherent in the method of Cotes for equal uncertainties of the observations. The resulting equations, whose number is equal to the number of coefficients to be determined, are then solved simultaneously. Mayer gives a numerical example in which he uses 27 observations on the position of a moon spot to write 27 equations each containing three unknown quantities (the coefficients in the equation to be fitted), which he divides into three groups of nine equations each. Then he adds all the equations in each group and solves the resulting three equations simultaneously to obtain the three unknown coefficients. A drawback of this method is that the results depend on the way in which the observational equations are divides into subsets and are therefore somewhat arbitrary and subjective. Euler (articles 122–123 of the cited work) is also credited with being the first to use the minimax principle (minimization of the maximum residual error) for solving a redundant system of linear equations.

Christopher Maire and Roger Joseph Boscovich (1755) report on the results of an expedition undertaken by the two authors under the auspices of Pope Benedict XIV to measure two degrees of meridian and correct the map of the Papal State. On pages 499–501, the author (Boscovich) attempts to determine the best value of the ellipticity of the earth from five measurements of degrees of meridian (the new one by Maire and himself reported earlier in the volume and four others) which he considers most reliable among a large number of available measurements. If the earth were exactly an ellipsoid of revolution and if the measurements were perfectly accurate, any two measurements of degrees of meridian made at different latitudes would determine its ellipticity exactly. But because the measurements are subject to error, each of the 10 pairs of measurements yields a different value of the ellipticity, which is directly proportional to the excess of the polar degree over the equatorial. If the ellipticity is computed from the arithmetic mean of all ten excesses, the result is 1/255, but if the two most discrepant values of the excess (one of which is actually negative) are discarded and the ellipticity is computed from the arithmetic mean of the eight remaining ones, the result is 1/195. Boscovich gives both of these results, but is not satisfied with either.

Thomas Simpson (1756) points out that the practice of taking the mean of a number of observations, while common among astronomers, has been questioned by some persons of considerable note who have maintained that a single observation, taken with due care, is as reliable as the mean of a great number. In order to refute that position, he determines the distributions of the mean errors of n independent observations from a discrete uniform (rectangular) distribution and from a discrete isosceles triangular population. He then compares these distributions with those of single observations from the same populations, and shows that the probability is less that the error of the mean of n observations equals or exceeds a given value than that the error of a single observation equals or exceeds the same value, the more so the greater the value of n.

Boscovich (1757) summarizes the measurement of a meridian arc near Rome and re-evaluates the data on this and previous measurements given by Maire and Boscovich (1755). He proposes for the first time two criteria for determining the best-fitting straight line $y = a+bx$ through three or more points: (1) the sums of the positive and negative residuals (in the y-direction) shall be numerically equal; and (2) the sum of the absolute values of the residuals shall be a minimum. His first criterion requires that the best-fitting straight line pass through the centroid $(\bar{x}, \bar{y})$ of the observations, whose coordinates are the arithmetic means of the x's and of the y's, respectively. The second criterion is then applied subject to the restriction imposed by the first. He proceeds to apply these criteria to the data of Maire and Boscovich but gives no indication of the method of solving the resulting equation for the best value of the slope b.

Simpson (1757) repeats the material of his earlier paper, with two notable additions. At the beginning he states explicitly for the first time the assumptions that the error distribution is (1) symmetric (positive and negative errors of the same magnitude are equally likely) and

(2) limited in extent (with limits depending on the goodness of the instrument and the skill of the observer). Four pages of new material at the end are devoted to extension to a continuous isosceles triangular error distribution of the results previously given for the corresponding discrete distribution.

Boscovich (1760) gives (pp. 420–425) a geometric method of solving the equations resulting from the criteria stated in his earlier paper to the problem of finding the straight line $y = a+bx$ of best fit to a number of points which are not collinear and applies this method to the same five meridian arcs, obtaining the value 1/248 for the ellipticity of the earth. This method is based on the ordered slopes $b_1 \geqslant b_2 \geqslant b_3 \geqslant b_4 \geqslant b_5$ of the lines connecting the five observational points (x_i, y_i), $i = 1, 2, \ldots, 5$ to their centroid $(\bar{x}, \bar{y})$.

According to Sheynin (1966), three works of Johann Heinrich Lambert (1760, 1765*a*, *b*), none of which the present author has seen, contain: (1) the first general outline since Galilei (1632) of the properties of errors of observations; and (2) a rule for estimating the precision of measurements by comparing the means taken with and without the most extreme observation. In the first work (1760), Lambert uses the principle of maximum likelihood for which he gives a graphical method of solution; Sheynin notes, however, that Lambert did not regard this principle as useful in practice and never returned to it. In the second work (1765*a*), Lambert states that the objectives of the theory of errors are to find the relations between errors, their consequences, the conditions of observation and the accuracy of instruments. He also undertakes a study of the errors of functions of the observations and endeavours to determine the "true value" of the observed quantity and to estimate the accuracy of the observations. He gives rules for fitting straight lines and curves by dividing the observations into groups and taking their centres of gravity instead of the original observations. In the third work (1765*b*), Lambert gives a justification for preferring an arithmetic mean to a single observation, a derivation of a semicircular probability density function for the distribution of errors, and a statement of the minimax principle (minimizing the maximum residual error), but confesses that he does not know how to use this principle in a general and straightforward manner.

James Short (1763) endeavours to determine the parallax of the sun from observations, at various points, of a transit of Venus, on the basis of comparisons of (1) observed durations, (2) least distance at centres and (3) internal contact at egress. In doing so he makes extensive use of what modern statisticians call trimmed means, i.e. means of the observations remaining after rejection of those differing from the mean of all the observations by more than a specified amount.

After solving several problems concerning averages of observations having discrete error distributions, which are reminiscent of Simpson (1756, 1757), Joseph Louis Lagrange (1774) states and solves his Problem X: "One supposes that each observation is subject to all possible errors between the two limits p and $-q$, and that the facility of each error x, that is, the number of cases in which it can occur, divided by the total number of cases, is represented by any function whatever of x designated by y; one requires the probability that the mean error of n observations shall be included between the limits r and $-s$." He applies the result to two examples: (1) $y = K$ (a constant) [uniform or rectangular distribution of error]; (2) $y = K(p^2-x^2)$, $(-p, p)$ [parabolic distribution of error]. He remarks (p. 228) that the latter appears to be "the simplest and most natural which one can imagine". He also considers a Problem XI, which is essentially a third example of Problem X with $y = K \cos x$, $(-\pi/2, \pi/2)$ [cosine distribution of error]. In each case the mean error of n observations has smaller dispersion (the more so the larger n) than the error of a single observation.

Pierre Simon Laplace (1774) considers the problem of determining the best average of three observations. He proposes two criteria: (1) the average should be such that it is equally likely to fall above or below the true value; and (2) the average should be such that the sum of the products of the errors and their respective probabilities is a minimum. He demonstrates that the two criteria lead to the same average. Let $x_1 \leqq x_2 \leqq x_3$ be the three observations, and let $p = x_2 - x_1$ and $q = x_3 - x_2$. Suppose that the true value is $x_1 + x$; then the probability [density] that the three observations (assumed to have come from a symmetric distribution) will fall at the points x_1, x_2 and x_3 will be $f(x)f(p-x)f(p+q-x)$, where $f(x)$ is the probability [density] that a single observation will fall at a distance x from the true value. Now construct a curve whose equation is $y = f(x)f(p-x)f(p+q-x)$. In order to satisfy Laplace's criteria it is necessary to find the value of x such that an ordinate erected at the abscissa x (measured from x_1) divides the area under this curve equally. The solution depends, of course, on $f(x)$. Laplace takes $f(x) = (m/2)e^{-m|x|}$ [the density function of what we now call Laplace's first distribution] and finds the solution $x = p + (1/m)\ln[1 + (1/3)e^{-mp} - (1/3)e^{-mq}]$, which approaches the arithmetic mean $(2p+q)/3$ as $m \to 0$ and the median as $m \to \infty$; for $0 < m < \infty$, it lies between the arithmetic mean and the median.

Daniel Bernoulli (1778) questions the practice common to astronomers of rejecting completely observations judged to be too wide of the truth, but assigning equal weights to all those retained. He advocates rejection of observations only if an accident occurred which rendered an observation open to question. He proposes a semicircular distribution of error, and discusses the choice of diameter. As limiting cases, the choice of an infinite diameter leads to taking the arithmetic mean as the average of the observations, while diminishing the diameter as much as possible without contradiction leads to taking the midrange. He proposes what has come to be known as the method of maximum likelihood to determine the average of a number of observations. For two observations, the result is equal to the arithmetic mean. For three observations, $x_1 \leqslant x_2 \leqslant x_3$, the result is greater than, equal to, or less than the arithmetic mean $(x_1 + x_2 + x_3)/3$ according as the median x_2 is less than, equal to, or greater than the midrange $(x_1 + x_3)/2$. For more than three observations, the method becomes unwieldy, since for n observations it requires solution of an equation of degree $(2n-1)$. In commenting on Bernoulli's paper, Euler (1778) proposes maximizing the sum of the fourth powers of the probability densities of the errors of the observations instead of maximizing their product (the likelihood function). He advances certain unconvincing arguments for the use of his criterion instead of Bernoulli's, and works out two examples based on real observations. The really vulnerable part of Bernoulli's method, as Isaac Todhunter (1865) has pointed out, is not the principle of maximum likelihood but the particular law of probability assumed.

Laplace (1781) extends the theory given in his earlier paper to any number of observations and generalizes it to the case in which each observation may have a different law of facility of error. He states that one can make infinitely many choices of an average according as one imposes various criteria, of which he enumerates four: (1) one may require that average such that the sum of the positive errors equal the sum of the negative errors [the arithmetic mean]; (2) one may require that the sum of the positive errors multiplied by their respective probabilities equal the sum of the negative errors multiplied by their respective probabilities; (3) one may require that the average be the most probable true value [Daniel Bernoulli's maximum likelihood criterion]; or (4) one may require that the error be a minimum, i.e. that the sum of the products of the errors (taken without regard to sign) and their respective probabilities be a minimum. He shows that criterion (4), which he regards as the fundamental one, is equivalent to criterion (2). He also shows that criterion (4) leads to the arithmetic mean and hence agrees with criterion (1), when the following conditions are satisfied: (1) the law of facility of error is the same for all the observations; (2) positive and negative errors of the same magnitude are equally probable; and (3) errors can be infinite, but the probability of an error x tends to zero as $|x| \to \infty$.

Jean Bernoulli III (1785), in an article on averages, refers to the methods of Boscovich (1757, 1760) and Lambert (1765*a*) and gives fuller accounts of the memoirs of Lagrange (1774) and Daniel Bernoulli (1778), the latter differing somewhat from the published version. The discrepancy is apparently accounted for by the fact that the summary given is based on a preliminary 1769 version in which a semicircular distribution of error is assumed as in the version published in 1778 but the method of maximum likelihood is not employed. Instead, the following iterative procedure is used: First take the mean of all the observations as the centre of the semicircle and determine the centre of gravity of the area corresponding to the observations; take this point as the centre of a new semicircle, and repeat the operation until the centre of gravity and the centre of the semicircle coincide.

Laplace (1786), given three or more non-collinear pairs of observations of two variables, x and y, proposes testing the adequacy of the linear relation $y = a+bx$ by first determining a and b so as to minimize the maximum absolute deviation from the fitted straight line, then deciding subjectively whether a deviation of this magnitude is consistent with the limits of the errors to which the observations are susceptible. He gives a procedure for determining the required values of a and b. In a later paper, Laplace (1793) gives a procedure which he says is much simpler. He observes that when the absolute value of the largest deviation is made a minimum, there are actually three observations whose deviations, two with one sign and one with the other, have this same absolute value. He offers another method of treating the observations, based on the criteria that (1) the sum of the deviations should be zero and (2) the sum of the absolute deviations should be a minimum. These criteria were first proposed by Boscovich (1757). Laplace develops an analytic procedure based on these criteria, while the procedure used by Boscovich (1760) was geometric. Laplace applies both his methods to data on lengths of degrees of meridian and on lengths of the seconds pendulum, both of which he uses to determine the earth's ellipticity. In the second volume of his two-volume treatise on celestial mechanics, Laplace (1799) summarizes the results of his earlier papers, again proposing the same two methods for determining the straight line $y = a+bx$ which best fits three or more points (x_i, y_i) whose coordinates are pairs of related observations: (1) minimizing the maximum residual; and (2) minimizing the sum of the absolute residuals subject to the restriction that the sums of the positive and negative residuals shall be numerically equal.

Gaspard Clair François Marie Riche Prony (1804) gives a geometric interpretation of the two methods of Laplace (1799), applies them to actual data, and compares the results with those obtained by a third method (his own) based on the idea that the deviation to be expected should be proportional to the independent variable x, or almost so.

Jean Trembley (1804), after brief mention of the work of Lambert, Laplace and Daniel Bernoulli on the most advantageous method of taking averages of observations, turns to the work of Lagrange (1774) on the same problem. He states that his purpose is to use combinatorial theory to obtain the same results which Lagrange obtained by the use of integral calculus. He succeeds in using combinatorial theory to obtain results for discrete error distributions which Lagrange found with the aid of differential calculus and Simpson (1756, 1757) by series expansions. He does not treat the case of continuous error distributions, which is the only one for which Lagrange employed integral calculus.

3. Eighty Years of Least Squares (1805–1884)

Adrien Marie Legendre (1805), while not the first to use the method of least squares, was the first to publish it. He starts with the linear form $E = a+bx+cy+\ldots$, where $a, b, c, \ldots$ are known coefficients which vary from one equation to another and $x, y, \ldots$ are unknowns which must be determined by the condition that the value of E reduces, for each equation, to zero or a very small number. He derives the normal equations without the explicit use of calculus by multiplying the linear form in the unknowns by the coefficient of each of the unknowns and summing over all the observations, then setting the sums equal to zero. If the results, when substituted in the normal equations, produce one or more errors judged too large to be admissible, he recommends rejecting the equations which produced them, and determining the unknowns from the remaining equations. Though he offers no mathematical proof of the method of least squares, Legendre makes the following claim for its superiority: "Of all the principles which one can propose for this object, I think that none is more general, more exact, or easier to apply than the one which we have used in the preceding research, which consists in making the sum of the squares of the errors a minimum. By this means, a sort of equilibrium among the errors is established which, preventing the extremes from prevailing, is most proper to make known the state of the system nearest to the truth." [Translation by present writer of statements on pp. 72–73.]

Puissant (1805) gives a theoretical discussion of the method of least squares, followed by an application to the determination of the ellipticity of the earth from measures of degrees of meridian. He mentions the method of conditional equations [method of averages] proposed by Mayer and the [Boscovich] method (preferred, he says, by Delambre) which gives "the least errors of latitude, half positive, half negative". He also applies the method of least squares to the determination of the ellipticity of the earth from the lengths of seconds pendulums, and compares the results with those obtained by Mathieu by minimizing the maximum discrepancy between observed and fitted values, as proposed by Laplace (1799).

Svanberg (1805), in the preliminary discourse of a book describing the measurement of a meridian arc in Lapland by Svanberg and three colleagues, compares the results obtained by applying the two methods proposed by Laplace (1799) to the determination of the earth's ellipticity from 15 measurements of the length of seconds pendulums and of degrees of meridian by various observers at different latitudes. No mention is made of the method of least squares; it is reasonable to assume that, at the time of writing, the author had not heard of it. The same assumption is probably valid in the case of von Zach (1805), who expresses the opinion that little reliance can be placed on the arithmetic mean when it does not stand equally far from the extremes. He reviews the work of Lambert (1765*a*) and Daniel Bernoulli (1778), but expresses a preference for the modification of Bernoulli's procedure due to Euler (1778), which he applies to data on terrestrial refraction and barometric pressure.

Jean Baptiste Joseph Delambre (1806–10) gives a three-volume report on a vast undertaking, carried out under the auspices of the Académie des Sciences with the support of the French government, to establish the base of the metric system (1 metre = one ten-millionth of the distance from the Equator to the North Pole) by measuring the meridian arc between the parallels of Dunkerque and Barcelona (over 9°). On page 117 of the first volume Delambre places himself squarely on the side of those who never suppress an observation or assign it a smaller weight simply because it deviates from other observations of the same kind. On pages 92 and 110 of the third volume he compares values of the earth's eccentricity (ellipticity) calculated from the observations of Delambre and Méchain by Laplace (1799), by Legendre (1805) and by himself. Laplace [by minimizing the maximum deviation] obtained the value 1/150; Legendre [by the method of least squares], 1/148; and Delambre [by an unspecified method, probably that of Delambre (1813)], 1/139. However, by combining the observations of Delambre and Méchain with those made by Bouguer in Peru about 60 years earlier, the task force obtained the value 1/334, which agrees much better with results obtained from measurements of the length of a pendulum of known period and with those predicted by the theory of nutation and precession. This latter value was used in determining the length of the standard metre.

Carl Friedrich Gauss (1806) claims priority in the use (though not in the publication) of the method of least squares in the following words (p. 184): "I still have not seen Legendre's [(1805)] work. I have purposely not taken the trouble to do so, in order that the work on my method shall remain entirely my own ideas. Through a few words, method of least squares, which de Lalande let fall in the last *History of Astronomy*, 1805, I arrive at the supposition that a fundamental theorem, which I myself have already used for 12 years in many calculations, and which I will also use in my work [Gauss (1809)], whether or not it belongs essentially to my method – that this fundamental theorem is also employed by Legendre." [Translation of portion quoted by Merriman (1877), pp. 162–163.]

Bernhard August von Lindenau (1806) states Laplace's (1799) analytic form of the method of Boscovich (1760), as well as Legendre's (1805) method of least squares, and applies both in the determination of the elliptic meridian. He does not comment as to the relative merits of the two methods, but reports that, in at least one instance, they yield very nearly the same results.

Robert Adrain (1808), apparently unaware of the work of Legendre (1805) and of the (as yet unpublished) work of Gauss, independently develops the method of least squares and uses it to solve the following problems: (1) Suppose $a, b, c, d, \ldots$ to be the observed measures of any quantity x, the most probable value of x is required [Ans. the arithmetic mean of the observations]; (2) Given the observed positions of a point in space, to find the most probable position of the point [Ans. the centre of gravity of the observed positions]; (3) To correct the dead reckoning at sea, by an observation of the latitude [the answer differs from all rules previously used, which he hopes will be abandoned]; (4) To correct a survey. The author mentions that he has also used the same principle to determine the most probable value of the earth's ellipticity. These last results were not published until ten years later [Adrian (1818*a*)].

Gauss (1809) deduces the normal (Gaussian) law of error from the postulate that when any number of equally good direct observations of an unknown quantity x are given, the most probable value is their arithmetic mean. He shows that the method of least squares, used by him since 1795, but named by Legendre (1805), follows as a consequence of the Gaussian law of error. If one does not assume this law, he might minimize the sum of the $2n$th powers of the errors for $n = 1, 2, 3, \ldots$, but Gauss points out that minimizing the sum of their squares ($n = 1$) is simplest. Letting $n \to \infty$ is equivalent to minimizing, as proposed by Laplace, the maximum errors (one positive and one negative, equal in magnitude). Gauss also mentions Laplace's other principle, first proposed by Boscovich, of making the sum of the absolute values of the deviations a minimum. He was apparently unaware that Boscovich proposed to minimize the sum of the absolute values of the deviations subject to the restriction that the sums of the positive and negative deviations shall be equal, since he speaks of this restriction as one added by Laplace. He does not mention the fact, though he may have been aware of it, that this restriction results in minimizing the sum of the absolute deviations from the arithmetic mean instead of from the median. The same is true of the fact that minimizing the sum of the $2n$th powers for $n \to \infty$ results in the choice of the midrange as an average instead of the arithmetic mean.

Laplace (1810) shows that if random samples of size n are drawn from a distribution with mean μ and known dispersion, then the distribution of sample means has mean μ and dispersion $1/\sqrt{n}$ times that of the parent distribution; moreover, under very general conditions [which Laplace does not state explicitly] on the parent, the distribution of sample means tends to normality as the sample size n increases. As Eisenhart (1964) has pointed out, these results greatly strengthen the justification given by Gauss (1809) for the use of the method of least squares, especially when dealing with a large number of observations. In a supplement, Laplace shows that when the law of error is the normal law, his own "most advantageous method" [Laplace (1781)], the method of maximum likelihood [Bernoulli (1778), Euler (1778?), and Gauss (1809)], and the method of least squares [which he introduces without reference to either Legendre (1805) or Gauss (1809)] are all equivalent and lead to the choice of the arithmetic mean as the average of a number of observations.

Friedrich Wilhelm Bessel (1810) uses the method of least squares to determine the orbit of a comet and Gauss (1811) uses it to determine the orbit of the asteroid Pallas. Gauss obtains 12 equations involving six unknown corrections to the elements of the orbit. Because the nature of the observations which furnish the tenth of these equations does not inspire confidence, he discards that equation and determines the unknowns from the other 11. Merriman (1877), page 166, notes: "We find here for the first time the notation $[a\ b] = a'b'+a''b''+a'''b'''+\ldots$ and also the algorithm for the solution of normal equations by successive substitution, since universally followed in lengthy computations...."

Laplace (1811*a*) considers, in his Articles VI and VII, the problem of choosing the average to take of n observations in order to correct an element already known approximately. He finds that the normal (Gaussian) law is the only one of the form $f(x) = Ke^{-g(x^2)}$ where $g(x^2)$ is continuous, for which the arithmetic mean is the "most advantageous" in the sense of Laplace (1781). However, because of the rudimentary form the central limit theorem given by Laplace (1810), choice of the arithmetic mean is advantageous when the number of observations is large or when one is taking the average of results each based on a large number of observations, and hence in these cases one may use the method of least squares, which Gauss (1809) developed from the postulate that the arithmetic mean is the best average of a number of observations. In his Article VIII [reprinted as Laplace (1811*b*)], Laplace extends these results to the case of correcting two unknown elements [regression coefficients]. His analysis is already quite laborious for this case, but he indicates that the results hold for any number of unknown elements whatever.

Laplace (1812), in his monumental work on the analytic theory of probabilities, summarizes the results of his study spanning almost four decades. Articles 20–24 of his Book II, Chapter iv, which is entitled "Of the probability of errors of the mean results of a large number of observations, and of the most advantageous mean results", contain most of the relevant material. Articles 20 and 21, which deal respectively with the correction of one or two elements, already known approximately, by the aggregate of a large number of observations, and which contain Laplace's "proof" of the method of least squares, follow closely the treatment of Laplace (1811*a*, *b*). Article 22, which deals with the case in which the facility of positive errors is not the same as that of negative ones [the distribution of errors is not symmetric] follows Laplace (1810). Article 23, unlike the preceding ones, deals with the case in which the observations have already been made. The idea of the "most advantageous" average as the abscissa corresponding to the ordinate which divides equally the area under the [joint] probability [density] curve [likelihood curve] of the observations goes back to two of Laplace's earliest memoirs [Laplace (1774, 1781)]. The author also summarizes the results of the supplement of Laplace (1810) and gives a more straightforward proof than that of Laplace (1811*a*) of the fact that the normal law of error is the only one of the form $f(x) = Ke^{-g(x^2)}$ for which the arithmetic mean is most advantageous. In Article 24, the author mentions various other methods of averaging observations, including the one proposed by Cotes (1722) and applied by Euler (1749) and Mayer (1750), and the one based on minimizing the sum of the 2nth powers of the deviations, which for $n \to \infty$ is equivalent to minimizing the maximum deviation, as proposed by Laplace (1786, 1799). He concludes that the best choice of method depends on the law of error when the number of observations is small, but that the method of least squares proposed by Legendre (1805) and Gauss (1809) is best whenever the number of observations is large. [The latter conclusion is unwarranted; see Lejeune Dirichlet (1836).] In the second supplement (first published in 1818), Laplace examines the method proposed by Boscovich (1757, 1760) [see also Laplace (1793, 1799)] based on minimizing the sum of the absolute values of the deviations, to which he gives the name "method of situation". For an odd number of observations on a single variable, this method leads to the median as the best average, while his own "most advantageous method" leads to the arithmetic mean. By finding the respective probabilities that the two averages are in error by a given amount, he determines a condition (on the law of error) under which the median is preferable to the arithmetic mean and, given the error law, explores the possibility of finding a weighted average of the two which is more precise than either.

Delambre (1813) returns to the question [see Delambre (1806–10)] of determining the eccentricity of the earth from inconsistent observations on the lengths of meridian arcs. On page 608 he advocates a method, which is probably the one he used in his earlier work, in the following words: "It seems that one should seek neither the least sum of errors nor the least sum of squares, but the least errors, half negative, half positive." Since the least sum of absolute deviations is achieved when the deviations are taken from the median, in which case half the deviations are negative and half positive, it appears that the author, perhaps without realizing it, is advocating the Boscovich-Laplace method without the restriction that the sums of the positive and negative deviations be equal in magnitude, which requires that deviations be taken from the arithmetic mean rather than from the median. On pages 607–608, Delambre applies his method to the determination of the earth's eccentricity from the Delambre-Méchain observations.

Claude Louis Mathieu (1813–14) uses 13 measurements made by a Spanish expedition, composed of two frigates, at various points on the globe, on the length of a seconds pendulum, to compute, by the method of least squares, the eccentricity of the earth. He obtains the values 1/323·2, 1/311·5 and 1/323·3 from the 9 points in the northern hemisphere, the 7 points in the southern hemisphere, and all 16 points, respectively. He also computes the eccentricity from 15 similar observations given by Laplace (1799), from which Laplace found the values 1/321·5 and 1/335·8 by minimizing respectively the maximum error and the sum of the absolute values of the errors (subject to the restriction that the sums of positive and negative errors be equal in magnitude). Mathieu finds 1/323·3 by the former method and 1/319·0 by the method of least squares. The maximum residual of 0·132 millimetre (well within the limits of error to which measurements of the length of the pendulum are susceptible) in the former case is attained at three points (the Equator and Lapland with negative signs and the Cape of Good Hope with positive sign) as required by theory. For the method of least squares, the maximum residual is +0·174 mm at the Cape of Good Hope. Mathieu remarks that there is a very small difference between the eccentricities obtained by the two methods.

Legendre (1814) sets the stage for a quotation from pages 72–75 of his earlier work [Legendre (1805)] by stating that Laplace (1812?) has found by considerations based on the calculus of probabilities that the method of least squares should be used in preference to all others to find the most exact average value of one or of several unknown elements among all those which are given by different observations. In so doing, he overstates, as Laplace himself and many later writers have done, the generality of what Laplace actually proved about the method of least squares, which is that the method of least squares is efficient for the normal error law and is consistent for other error laws satisfying certain conditions.

Jan Frederik van Beeck Calkoen (1816) discusses the average value of a certain number of quantities or of separate observations. For several observations of a single quantity he advocates the use of the arithmetic mean. If one of the observations differs from the mean by an amount greater than the assumed limit of error, that observation is discarded, and the arithmetic mean of the remaining ones is taken. For observations on two related quantities, he proposes two methods of determining the best fitting straight line. The first, which he attributes to Lambert (1765*a*), involves dividing the points representing the pairs of observed (x, y) values into two groups (as nearly as possible equal in number), one containing the points with the smallest abscissas and the other those with the largest abscissas, and joining the centres of gravity of the two sets of points. The other method is based on the use of the Boscovich criteria, which the author attributes to Laplace (1799). By taking x^2 or $\sqrt{x}$ rather than x as the independent variable, the author obtains curvilinear regression equations of the forms

$y = \alpha+\beta x^2$ and $y = \alpha+\beta\sqrt{x}$ as well as the linear regression equation of the form $y = \alpha+\beta x$. He advocates using that power of x which gives the best fit in the sense that the sum of the absolute deviations of the observed points from the fitted curve is smallest, subject to the condition that the algebraic sum is zero. It is interesting to note that he makes no mention of the method of least squares, although the work of Legendre (1805), Gauss (1809) and Laplace (1812) was already widely known.

Gauss (1816) points out that it is not necessary to know the precision h [$= 1/\sigma\sqrt{2}$, where σ is the standard deviation] of the observations in order to apply the method of least squares, and that the relation of the precision of the results to that of the observations is independent of h, but that the value of h is itself interesting and instructive. He then proceeds to give various methods of determining h, including methods based on the nth root of the sum of the nth powers of the absolute errors (deviations from the true value) for $n = 1, 2, 3, 4, 5, 6$, and an alternate method based on the median M of the absolute values of the errors. He shows that the method based on $n = 2$ gives the greatest precision for samples from a normal population, 100 observations for $n = 2$ yielding the same precision as 114 for $n = 1$, 109 for $n = 3$, 133 for $n = 4$, 178 for $n = 5$, 251 for $n = 6$, or 249 [actually 272 – see comments below] for the alternate method based on M, but notes that the last method and the one based on $n = 1$ are arithmetically more convenient. Although he gave the correct mathematical expression for the probable error of the median absolute error M, Gauss made a mistake in calculating the value of the numerical coefficient. Several later authors, including Hauber (1830), Encke (1832–34) and Jordan (1869), have given the correct value, but it is interesting to note that the first two, writing during the lifetime of Gauss, did so without mentioning his mistake, which remains uncorrected in his collected works.

Adrain (1818*a*) calculates the earth's ellipticity by the method of least squares from data on the lengths of pendulums vibrating seconds at different latitudes given by Laplace (1799). He compares the results not only with those obtained by Laplace, based on the criteria of Boscovich (1760), but also with the results obtained by that method after correcting two errors made by Laplace. He finds that most of the discrepancy between Laplace's results and his own is due to those errors. The corrected results of applying the Boscovich-Laplace method, based on minimizing the sum of the absolute values of the residuals subject to the restriction that the algebraic sum of the residuals shall be zero, differ by less than 1 per cent from those obtained by the method of least squares. In another paper, Adrain (1818*b*) uses the method of least squares to find the diameter of the sphere (7918·7 miles) which most nearly coincides in various specified peculiarities with the actual terrestrial spheroid, given measurements of degrees of meridian.

In 1821 there appeared an anonymous paper whose authorship Czuber (1891*a*, 1899) attributes to Svanberg. The author gives a discussion, which is as much philosophical as mathematical, of the problem of finding the best average of a number of observations. He distinguishes between two cases, one in which the observations are all made on the same identical object and thus differ only because of errors of observation and the other in which observations are made on a quantity which is itself variable. He traces the history of the problem from the time when the arithmetic mean was used without question, through the period in which students of the theory of probability (among whom he mentions Boscovich, D. Bernoulli, Lambert and Lagrange) questioned its use, to the time when wide acceptance of the method of least squares developed by Legendre (1805) and Gauss (1809) led to the belief that the arithmetic mean is indeed the most probable value. He pleads for further examination of the question, raising objections to the use of the arithmetic mean when the observations are not closely bunched, especially if they are so asymmetric that there are many more on one side of the arithmetic mean than on the other, or when there is reason to believe that they are not all equally reliable. He mentions a number of other possible averages, such as the median, the midrange, and the arithmetic mean of those remaining after discarding the (one or more) largest and smallest observations. He concludes that the problem of the best average depends on the law of facility of error and hence has no general solution. Nevertheless, at the end of the paper he proposes an iterative procedure which starts from the arithmetic mean (or some other reasonable value), then takes the reciprocals of the residuals (or their squares) as weights of the corresponding observations and thus obtains a second approximation, which gives new residuals, after which the process is repeated until it converges.

Gauss (1823) compares his earlier formulation of the method of least squares [Gauss (1809)] with that of Laplace (1812), and concludes that neither is entirely satisfactory. The former is based on the assumption that the errors of observation follow a normal (Gaussian) distribution, which follows from his postulate that the best average of the observations is their arithmetic mean. Laplace asserted that the method of least squares yields a result which is best asymptotically (in the sense of minimizing the sum of the absolute values of the residuals subject to the restriction that the algebraic sum of the residuals shall be zero, when the number of observations is sufficiently large), whatever the distribution of errors [under very general conditions not stated by Laplace]. That left a gap, which the author now proposes to fill, for the case of a small or moderate number of observations whose errors are not normally distributed. Gauss begins by comparing the situation to a game in which there is no gain to hope for, but a loss to fear, the problem being how to minimize the loss, which is assumed to be the same for positive and negative errors of equal magnitude. This assumption can be met by choosing a loss function proportional to the sum of the absolute values of the errors, as Laplace did, or to the sum of their nth powers, n being a positive even integer. In the German summary, but not in the Latin text, Gauss points out, as Laplace had already done, that the larger the exponent n becomes, the nearer one comes to the situation where the most extreme errors alone serve as a measure of precision. Gauss chooses $n = 2$, which besides being the simplest of its type also possesses certain desirable properties [see Gauss (1816) for a proof, assuming a normal distribution, of these properties, which, unfortunately for his argument, do not hold for certain other common distributions]. On the basis of this choice he justifies the use of the method of least squares, whatever the number of observations and whatever the distribution of their errors. Gauss' second exposition seems to the present writer to be no more satisfactory than his first. In each case he starts from a postulate, plausible but not universally valid, which leads inexorably to the foregone conclusion. Nevertheless, his argument apparently convinced his contemporaries, since the literature of the next few decades includes many writings on least squares but only a few on rival methods.

Augustin-Louis Cauchy (1824), given a large number of observations of two variables, x and y (points in the xy-plane), seeks to determine the values of two elements (coefficients in the linear regression equation $y = a+bx$) such that the absolute value of the largest residual is a minimum. He accomplishes this minimization by means of an iterative scheme. He shows that a line in the plane may be such that one, two, or three of the given points deviate from it by the maximum amount, but that for the line which is the unique solution of the problem there are three such points with the maximum residual, two residuals of one sign and one of the other [cf. Laplace (1793)]. He proves four theorems concerning the possible system of values of the elements, and gives a geometric interpretation of each in terms of the number of faces, edges and vertices of a convex polyhedron. This paper is a condensation of a memoir presented in 1814; the entire memoir, which was published later [Cauchy (1831)], includes a generalization of the theory in two directions, considering the case in which the function of the elements which represents the errors is a power series and the number of elements exceeds two. Cauchy shows that the number of residuals whose absolute value is equal to the maximum always exceeds by at least one the number of variable elements.

Jean Baptiste Joseph Fourier (1824) considers the problems of fitting a linear equation in

n variables to a set of m observed points ($m>n$) so as to minimize (1) the maximum absolute deviation or (2) the average absolute deviation. Fourier gives, for the case $n = 2$, a geometric solution of the first problem which is equivalent to the analytic solution of a system of inequalities. The latter can easily be extended to larger values of n. He states that the second problem can be solved in an analogous manner. Both are formulated as what we would now call linear programming problems, i.e. minimization of an objective function (the largest absolute deviation or the average absolute deviation) subject to constraints in the form of linear inequalities. The method used by Fourier has come to be known as Fourier's method of descents. This method is also given in his posthumous book [Fourier (1831)].

Two memoirs by Poisson (1824, 1829), large parts of which are reproduced in a later work Poisson [(1837)] are, according to Merriman (1877), pages 175–176, a commentary on the fourth chapter of Laplace (1812). Merriman quotes Todhunter (1869) to the effect that Poisson confines himself to the case in which one element is to be determined from a large number of observations, but treats it in a more general manner than Laplace, dropping the assumptions that positive and negative errors are equally likely and that the law of facility of error is the same for every observation.

James Ivory (1825, 1826) gives four demonstrations of the method of least squares. His first paper is divided into three parts. In the first part he gives two of his demonstrations, neither of which is based on the theory of probability, which he considers irrelevant. In the second part he discusses the probability of errors, failing to recognize that the probability of any definite error for a continuous distribution must be an infinitesimal, and making no distinction between true errors and residuals. In the third part he attempts to show that the method of least squares cannot give the most advantageous or probable results unless the law of facility of error is the normal law $\phi(x) = ce^{-h^2x^2}$. On page 165, he makes the following statement concerning the demonstration of Laplace (1812), Book II, Ch. iv, Art. 20: "... whatever merit it may have in other respects, [it] is neither more nor less general than the other solutions of the problem". Later authors have regarded Ivory's demonstrations as unsatisfactory, and the present writer shares this opinion. Glaisher (1872) has analysed Ivory's criticism of Laplace, which he regarded as a result of Ivory's failure to understand the demonstration of Laplace. It appears to the present writer that Glaisher was guilty of the same fault. In modern terminology, what Laplace actually asserted in the article cited by Ivory [see also Laplace (1810, 1811*a*)] is that the method of least squares is *asymptotically* most advantageous for any error distribution which is well enough behaved so that its mean is asymptotically normally distributed. He did not claim to have shown that it is most advantageous [best] for any finite number of observations from a non-normal error distribution, but recommended it as advantageous [good] and computationally convenient whenever the number of observations is large. Ivory's second paper contains his fourth demonstration, regarded by Ellis (1844) as no more satisfactory and by Merriman (1877) as still more absurd than the previous ones.

Georg Wilhelm Muncke (1825) gives an exposition of the method of least squares based largely on the demonstration of Gauss (1823). He proposes the use of the arithmetic mean of the observations remaining after those farthest from the mean have been excluded. Gauss (1828) gives a method of solving the normal equations which arise in carrying out the method of least squares. Whittaker and Robinson (1924) state that this method is substantially equivalent to reduction of a quadratic form to a sum of squares.

Carl Friedrich Hauber (1830*a*) extends the work of Gauss (1816, 1823) on the estimation of the precision of observations to the case of s observations arising from populations having (possibly) different dispersions. The situation in which all come from the same population is included as a special case. He considers estimators based on the square root of the mean of the squares of the errors, the mean absolute error, and the median absolute error. He compares the precision of these estimators when the law of the facility of error is the normal (Gaussian) that the best method of correcting one or more elements on the basis of linear observational equations greater in number than the elements depends, for a small number of observations, upon the underlying law of error, but that this dependence disappears when the number of observations is very large, in which case the method of least squares is to be preferred to all others. He points out that Laplace's conclusion concerning the superiority of the method of least squares for very large numbers of observations is not only unwarranted by the evidence presented by Laplace, but is actually incorrect. In the case of a single element, for example, Laplace's conclusion does not follow from the fact that the arithmetic mean tends in probability to the true value as the number of observations increases, since the same is true of other measures of central tendency, e.g. the median. Whatever the sample size, Lejeune Dirichlet points out, the question of whether or not the arithmetic mean is superior to the median depends upon the ratio of two constants [their standard errors], one of which [the standard error of the arithmetic mean] depends on an integral over the whole range of the error curve, while the other [the standard error of the median] depends only on the maximum ordinate of the curve.

Cauchy (1837) states the following problem (pp. 460–461 of the English translation): "... I suppose that a function of x represented by y is developed in a converging series arranged according to the ascending or descending powers of x, or according to the sines and cosines of an arc x, or, more generally, according to other functions of x which I shall represent by $\phi(x) = u, \chi(x) = v, \psi(x) = w$; so that we have (1) $y = au+bv+cw+\ldots$ where a, b, c,... are constant coefficients. Now the question is, first, how many terms of the second member of the equation (1) are to be employed, in order that the difference between it and the exact value may be very small, and capable of being compared with the errors to which the observations are liable; secondly, to determine in numbers the coefficients of the terms retained, or, in other words, to find the approximate value just mentioned." The data consist of n values of y represented by y_i ($i = 1, \ldots, n$) and the corresponding values of x_i (and hence of u_i, v_i, w_i, ...) related by n equations (2) $y_i = a_iu_i+b_iv_i+c_iw_i+\ldots$. The author proposes successive approximations based on neglecting all but one, two, ... terms on the right-hand side of equations (2), the process continuing until the residuals are comparable to the inevitable errors of observation. Cauchy's method must be considered as one of the alternatives to the method of least squares.

Gotthilf Heinrich Ludwig Hagen (1837) advocates the use of the method of least squares [Legendre (1806), Gauss (1809, 1823)], which he explains in considerable detail. He does mention, however, the use by Prony (1804), before the method of least squares was known, of the method of Laplace (1799) based on the criteria of Boscovich, as well as the work of Lambert (1765*a*). He also discusses the suppression of outlying observations, which he strongly opposes unless there is some reason other than the fact that they deviate considerably from the remaining ones, and the assignment of weights to individual observations. Friedrich Wilhelm Bessel (1838) discusses the probability of errors of observation and the method of least squares as developed by Laplace (1812), Gauss (1823) and others. He shows that the normal law of error is not to be regarded as an *a priori* rule, free from exception, and throws new light on the conditions under which it holds. Bessel and Johann Jakob Baeyer (1838) join Hagen (1837) in taking a firm stand against the rejection of outlying observations, which had been advocated and practised by such earlier authors as Boscovich (1755, 1757), Lambert (1760, 1765*a*) and Legendre (1805). S. Stampfer (1839), on the other hand, has no qualms about rejecting observations. Of nine determinations of the ratio of the lengths of the Vienna fathom and the metre by various methods, he rejects the two smallest on the grounds that both were obtained by comparisons with the French standard half toise, which leads him to suspect a constant error in the standard half toise. He is still not satisfied with the result, and proceeds to discard also the smallest of the remaining values, apparently for no other reason than its discrepancy from the six still remaining. Christian Ludwig Gerling (1843) gives an excellent treatment of the method of least squares. He recommends great caution in discarding observations, but says even so that "there remain observations which we must discard after the fact, because we hold it to be more probable that a gross blunder has occurred than that an unavoidable error can produce such a large deviation".

William Fishburn Donkin (1844) starts from the assumption that the weight of an observation is proportional to the square of its precision (inversely proportional to its variance) and, as one would expect, he reaches the same conclusion as the one Gauss (1823) reached by assuming a squared error loss function, namely that the method of least squares should be used, independently of the law of facility of error. Robert Leslie Ellis (1844) examines in detail the demonstrations of the method of least squares by Gauss (1809, 1823), Laplace (1812) and Ivory (1825, 1826). He concludes that Laplace's objection to Gauss' first demonstration, based on the postulate that the arithmetic mean is the best average to take of a number of observations, is justified. He regards Laplace's demonstration and Gauss' second as somewhat more satisfactory, but endeavours to show that none of the three tends to prove that the results of the method of least squares are the most probable of all possible results. He finds Ivory's demonstrations, which are not based on the theory of probabilities, not at all conclusive.

Lambert Adolphe Jacques Quetelet (1846) gives an elementary exposition of the theory of means and of the laws of error, in which he advocates use of the interquartile distance as a measure of the probable error. He uses this method of estimating the probable error of the right ascension of the North Star [repeated measurements of the same quantity] and of the chest measures of Scottish soldiers [measurements of related quantities].

Augustus De Morgan (1847) gives an extensive treatment of the method of least squares which consists largely of a translation of and comments on the treatment of Laplace (1812). In cases in which the relative precision of the observations is in doubt, he proposes an iterative procedure in which one makes the best possible initial estimate of the weights, finds the most probable result, then adjusts the weight accordingly, and repeats the process until assumed and deduced weights agree.

Sir John Frederick William Herschel (1850) gives a demonstration of the method of least squares, similar to that of Adrain (1808), based on the assumption that the components of error in two orthogonal directions are independent. In commenting on the work of Quetelet (1846), he questions by what numerical process the latter obtained his averages of the chest measures of Scotch soldiers and the heights of French conscripts, pointing out that his values do not agree with those of either the arithmetic mean or the median. Ellis (1850) discusses Herschel's proof of the method of least squares, which he regards as unsatisfactory, and explains and defends Laplace's method. Donkin (1851) offers some critical remarks on the theory of least squares, and especially on the remarks of Ellis. Donkin says that Herschel's proof "should be treated with respect" and that the method of least squares may be used, if for no other reason, because "it is a very good method", as shown by Gauss (1823).

Jules Bienaymé (1852) reviews the development of the theory of least squares from the early work of Legendre (1805), Gauss (1809, 1823) and Laplace (1811*a*, 1812). He considers the modifications and generalizations required when the observations are not all equally precise and when not one but several variables are to be estimated, with particular emphasis on the precision of the results of applying the method of least squares to these cases. A later paper of Bienaymé (1858) is practically identical with this one.

Benjamin Peirce (1852) proposes the first objective criterion for the rejection of observations, based on the principle that observations in question should be rejected when the probability of the system of errors when they are retained is less than that of the system of errors obtained by their rejection multiplied by the probability of making exactly so many abnormal observations. The details of this criterion and others proposed by later authors will be omitted. Rider (1933) gives an excellent summary of those proposed up to that time.

Cauchy (1853*a*) maintains that his method of interpolation [Cauchy (1837)] can be used to determine several unknown quantities from a redundant system of equations, with results nearly as accurate as by the method of least squares. Bienaymé (1853*a*) argues, however, that the two methods are completely different and even that a contradiction exists. Cauchy (1853*b*) maintains that in many investigations his method of interpolation is preferable to the method of least squares. Cauchy (1853*c*) claims that his method is the shortest, and that the method of least squares gives the most probable results only under certain conditions which are, according to Cauchy (1853*d*, *e*), that the law of facility of error is the same for all the errors, that no limits can be assigned to the magnitude of an error, and that the probability of an error is proportional to $e^{-h^2x^2}$. Cauchy (1853*f*) shows that the most probable values may sometimes differ from those found by the method of least squares. Bienaymé (1853*b*) reviews some of Cauchy's articles [Cauchy (1853*d*, *e*, *f*)] and maintains that the mean sum of squares of errors is under all circumstances a measure of precision of the observations. Cauchy (1853*g*) shows that the system of weights which makes the largest error to be feared in a mean as small as possible often differs considerably from that given by the method of least squares.

Pafnutiĭ L'vovich Chebyshev (1854) observes that if one wants the best polynomial approximation (of a given degree) to a continuous function $f(x)$ in the neighbourhood of $x = a$, one should use the sum of the appropriate terms in the Taylor series expansion in powers of $x-a$, but that if one wishes to find the best such approximation in the interval $(a-h, a+h)$, one should prefer another polynomial, whose maximum deviation from $f(x)$ in the given interval is less than for any other polynomial of the same degree, as proposed by Poncelet (1835). If U is a polynomial of degree n with $[n+1]$ arbitrary coefficients, and if one chooses these coefficients so that the difference $f(x)-U$, from $x = a-h$ to $x = a+h$, remains within these limits the closest to 0, he shows that the difference $f(x)-U$ has the property that, among the largest and smallest values of the difference $f(x)-U$ between the limits $x = a-h$ and $x = a+h$, one finds the same numerical value at least $n+2$ times. The analogous property for the case of fitting a polynomial to a finite number of points was known already to Laplace. For the continuous case, Chebyshev gives a method of finding the abscissas of the $n+2$ points where the absolute error should take its maximum value. This method and the theory of approximation based upon it are, for reasons that should be obvious, not applicable to the case of a finite number of points and hence not directly relevant to the present study.

Joseph Bertrand (1855) offers certain historical and critical remarks on presenting a copy of his translation into French of the Latin memoirs of Gauss. An English translation by Hale F. Trotter (1957) has since been prepared from Bertrand's French translation.

Benjamin Apthorp Gould, Jr. (1855) gives tables for Peirce's criterion which are more extensive than those of Peirce (1852) and include two more significant figures. He reworks Peirce's two examples, and in one case concludes that only one observation should be rejected, whereas Peirce rejected two; however, Rider (1933) has pointed out that neither result is trustworthy, since Gould uses Peirce's incorrect value of the standard deviation.

Humphrey Lloyd (1855) advocates, in effect, the use of the midrange in averaging meteorological observations. This average is still used by meteorologists today, the so-called mean daily temperature being the arithmetic mean of the highest and lowest temperatures recorded during a 24-hour period.

George Bidwell Airy (1856), after studying the papers of Peirce (1852) and Gould (1855) on Peirce's criterion for the rejection of doubtful observations, summarizes his conclusions as follows: "(1) The mathematical theory of probabilities fails in all questions applying to errors of extreme magnitude. (2) No considerations of the magnitude of residual errors *per se* will justify us in rejecting a result. (3) We are justified in rejecting a result only when, from the best estimate that we can form of the extent of action of the various causes which can produce error, we find that the combination of those causes of error cannot possibly produce the discordance in question. (4) And when we perceive that other causes may have intervened, whose nature is such that they cannot be recognized as occurring in the ordinary series of observations." Joseph Winlock (1856) answers Airy's criticisms of Peirce's criterion, summing up the case in its favour as follows: "Regarding the probability of an error as a function of its magnitude, we are enabled to find the probability of any system of residual errors, and by the comparison of the system of errors before and after rejection in accordance with the rule of the Criterion, we can decide, within safe limits, whether the probability of our final result is lessened by retaining the doubtful observations."

Joseph Petzval (1857) holds that the method of least squares is not applicable in optics, because of the failure of the underlying assumptions that positive and negative errors of equal magnitude are equally probable and that all observations are made under equally favourable conditions by equally skilled observers with equally good instruments. He questions especially the second of these assumptions. He proposes instead what he calls "the method of numerically equal maxima and minima" in which the maximum of the absolute values of the residuals is minimized. He points out that this is equivalent to minimizing the sum of the $2m$th powers of the residuals, where m is an integer which tends to infinity as a limit.

C. G. von Andrae (1860) studies the problem of choosing one, two, three,... of a series of n equally reliable observations to be used instead of all n observations in determining the most advantageous value of the measured quantity. In choosing a single observation, he uses the principle [Laplace (1799)] of minimizing the sum of the absolute values of the errors, dropping the Boscovich condition that the sums of positive and negative errors be equal in magnitude. Hence he chooses the median, which he defines, for a sample of size n, as the mth ordered observation, where $m = n/2$. By analogy, if s observations are to be used, he chooses the m_ith ordered observations ($i = 1, 2, \ldots, s$), where $m_i = ni/(s+1)$.

Airy (1861) gives a discussion of the method of least squares, without mentioning any rival methods, and makes further comments along the lines of those in his earlier paper [Airy (1856)] on the rejection of doubtful observations.

Charles A. Schott (1861) gives a free translation into English of the paper of Cauchy (1837) on Cauchy's method of interpolation. William Pitt Greenwood Bartlett (1862) applies this method to actual observations in the fields of physics and chemistry.

William Chauvenet (1863), in an appendix to the second volume of his treatise on astronomy, gives a detailed discussion of the method of least squares. The author discusses Peirce's criterion for the rejection of doubtful observations and proposes his own criterion for rejecting a single observation. The latter is based on the principle that, since the number of errors numerically greater than $\kappa\sigma$ that may be expected to occur in n observations is

$$2n \int_{\kappa}^{\infty} \phi(t)\, dt = n\psi(\kappa),$$

where $\phi(t) = e^{-t^2/2}/\sqrt{2\pi}$, an observation deviating from the mean by an amount greater than $\kappa\sigma$ should be rejected if the quantity $n\psi(\kappa)$ exceeds 1/2, since such an error "will have a greater probability against it than for it". The appendix and related tables were reprinted separately in 1868.

Augustus De Morgan (1864) declares that the arithmetic mean is the best average of a series of observations because the most probable result is the arithmetic mean plus corrections of which we have no knowledge, either as to sign or value and no means of getting any, so that there is no reason for supposing that the true value lies on one side of the arithmetic mean rather than the other, so long as we know nothing of the law of facility of error.

Isaac Todhunter (1865), in his history of probability, summarizes the work of various writers on the theory of errors, including Simpson (1757), Lagrange (1774), D. Bernoulli (1778), Euler (1778), J. Bernoulli (1785) and Svanberg (1805), as well as numerous writings of Laplace (1774, 1781, 1786, 1799, 1810, 1811*a*, *b*, 1812). The last four of these deal primarily with the method of least squares but Todhunter makes it clear that Laplace never entirely abandoned some of his earlier methods.

Edward James Stone (1868) defines what he calls a modulus of carelessness, m, which, for a given observer and a given class of observations, expresses the average number of observations which that person makes with one mistake. Stone proposes a criterion for rejection of observations which, with $m = 2n$, where n is the number of observations, is equivalent to Chauvenet's.

Wilhelm Jordan (1869) extends Gauss' table of factors for computing the probable error and its probable uncertainty from the nth root of the mean of the nth powers of the absolute values of deviations from the true value up through $n = 10$ and corrects Gauss' factors for the median, which he shows to give a slightly less (rather than more) precise estimate than the other method for $n = 6$.

Todhunter (1869) develops Laplace's treatment of the method of least squares and demonstrates that some of the results which Laplace obtained for the case of two elements hold for the case of any number of elements.

Cleveland Abbe (1871) gives a historical note on the method of least squares in which he points out that although Legendre (1805) was the first to publish the method and Gauss had used it since 1795 (though he did not publish it until 1809), it was independently developed by Adrain (1808) in America. The author reprints a portion of Adrain's original investigation, gives interesting biographical notes on Adrain, and summarizes the results of two of his later papers [Adrain (1818*a*, *b*)] in which he applies the method of least squares. G. Zachariae (1871) gives an excellent textbook treatment of the method of least squares.

James Whitbread Lee Glaisher (1872) gives a history of least squares, including an account and a critical evaluation of the contributions of Legendre, Adrain, Gauss, Laplace, Ivory, Ellis, De Morgan and others. He offers an alternative to the rejection of observations in the form of an iterative procedure in which the weights of the observations are adjusted after each iteration as proposed by De Morgan (1847). Last, but not least, he proves that if errors are distributed according to Laplace's first law [$f(x) = (m/2)\,e^{-m|x|}$], the median of the observations is the most probable true value. He does this by showing that the probability [density] of the true value x is proportional to exp [$-m$ (the sum of the absolute values of the deviations of the observations from x)] and that that sum is a minimum when taken about the median [the middle one of an odd number of observations or any value between the middle two of an even number of observations].

Friedrich Robert Helmert (1872), in the first edition of a book on the adjustment computation by the method of least squares, gives a proof, following that of Gauss (1816), that the probable error can be determined more precisely from the mean of the squares of a number of observations (assumed to have come from a normal distribution) than from the mean of the absolute values of the errors. In later editions, he adds a section on the theory of the maximum error and its use in the exclusion of observations.

Glaisher (1873) elaborates on the alternative to the rejection of observations proposed in his earlier paper [Glaisher (1872)]. He also examines the criterion proposed by Stone (1868) for the rejection of outlying observations, and criticizes it on two grounds: (1) any rejection criterion based on the supposition of the validity of the arithmetic mean is inconsistent; (2) even among such criteria Stone's is not the most desirable one and is impractical because of the practical impossibility of determining the value of n, it being assumed that the observer makes one mistake in n observations. In two papers published the same year, Stone (1873*a*, *b*) justifies the use of the arithmetic mean and the normal law of error on the basis of the axiom that all direct measures are of equal value and examines in detail the objections raised by

Glaisher (1873) to the author's criterion [Stone (1868)]. He points out that his criterion is relatively insensitive [robust, as modern statisticians would say] to moderately large variations in n. He insists that even if Glaisher's assumptions are granted, Glaisher has not maximized the right expression, and hence has not found the correct weights for the observations. Further notes by Glaisher (1874) and Stone (1874) appear to have generated more heat than light.

Charles Sanders Peirce (1873) shows that the method of least squares, with equal weights for all observations, is valid only when the precision constant h ($= 1/\sigma\sqrt{2}$, where σ is the standard deviation) is the same for all the observations; otherwise the weights assigned to the observations must be proportional to the squares of their respective precision constants. But since the true precision constants are almost always unknown, the weights must be estimated from the observations themselves. Since the method of least squares also depends on the errors of observation being normally distributed, Peirce reports the results of an experiment performed to test the validity of this assumption. In this experiment the time required for a previously untrained observer to respond to a signal by pressing a telegraph key was measured (in thousandths of a second) about 500 times each day for 24 days. Peirce concludes that the response times on a given day are approximately normally distributed, though there is evidence of an increase in precision from day to day, especially at the outset (learning effect). Peirce also discusses the rejection of discordant observations, which he says is entirely in accordance with the method of least squares. He advocates use of the criterion proposed by his father, Benjamin Peirce (1852).

Todhunter (1873) reviews the work of various authors, especially Boscovich and Laplace, on methods used to find the equation $y = a+bx$ of the best-fitting straight line involved in the determination of the ellipticity of the earth from measurements of degrees of meridian and lengths of a seconds pendulum at widely separated points on the earth's surface. He writes: "I presume that neither of the methods which Laplace [(1799)] discusses would now be practically used in such calculations, but the method of least squares".

Gustav Theodor Fechner (1874) shows that, while the sum of squares of deviations is a minimum when taken from the arithmetic mean, the sum of the absolute deviations is a minimum when taken from the median. He makes a remark which leads to the conclusion that he was unaware that the latter fact was known to earlier writers [including Laplace (1812–1818 suppl.), von Andrae (1860) and Glaisher (1872)]. He also discusses power means, which he defines as values such that the sums of powers of deviations are minimal when taken from them, and probability laws under which such power means are valid averages.

William Stanley Jevons (1874) notes that Boethius stated ten kinds of means or averages and Jordanus added an eleventh, but he discusses only three of them (arithmetic, geometric, and harmonic means). He notes that Quetelet (1846) distinguished between means of observations of a quantity which is itself constant, but subject to variable measurement error, and of a quantity which is itself variable, and that Herschel (1850) pointed out the importance of this distinction. Jevons gives demonstrations of the normal law of error by Gauss, by Herschel, and by Laplace and Quetelet. He states that there is no justification for believing that the errors committed in all classes of observations should follow the same law, but does not suggest any other. He also discusses in some detail the method of least squares and the rejection of outlying observations. He takes a position against arbitrary exclusion of discrepant observations, in the following words (page 393 of the second edition): "The mere fact of divergence ought not to be taken as conclusive against a result, and the exertion of arbitrary choice would open the way to the fatal influence of bias. . . . The apparently divergent number may prove in time to be the true one. . . . To neglect a divergent result is to neglect the possible clue to a great discovery."

Hervé Auguste Étienne Albans Faye (1875) discusses various justifications of the method of least squares. He points out that Gauss and Legendre deduced it from the accepted opinion that the most probable value of a quantity of which a number of observations have been made is their arithmetic mean, while Laplace and others justified it on the basis that the errors are due to a large number of causes each contributing only a small part of the resultant error. He insists that the law of probability of errors cannot be established *a priori*, on the basis of a hypothesis or of a generally accepted opinion, in spite of the extreme elegance of the proof of Gauss, but must be established *a posteriori*, from a direct study of the facts. He gives an example in which, because of a systematic error, the method of least squares gives an extremely misleading result; quite rightly, however, he does not blame this result on the method but on the observations. Hermann Laurent (1875), commenting on the same question, says that the Gaussian law of error should never be accepted *a priori*; on the contrary, one ought to reject it, because it assigns positive probabilities to impossibly large errors. "Who is the astronomer", he inquires, "who makes an error of 361 degrees in measuring an angle?" He makes a study of 1,444 measurements of an angle of approximately 16°, and concludes that the observations cast doubt on the exactness of the Gaussian law, and that therefore one ought to reject the method of least squares when one has only a small number of observations.

Francis Galton (1875) proposes the use of the median as a measure of central tendency and of the difference between the median and one of the quartiles, or the average distance between the median and the two quartiles, as a measure of dispersion (probable error).

Truman Henry Safford (1876) gives rules for good observation based on the method of least squares, and hints for abbreviating computations. Mansfield Merriman (1877) gives a chronological bibliography, containing 408 titles and covering the period 1722–1876, on the method of least squares and rival methods, with valuable historical and critical notes.

Benjamin Peirce (1878) gives a fuller explanation of the criterion which he proposed over a quarter of a century earlier [Peirce (1852)]. Charles A. Schott (1878) makes favourable remarks on Peirce's criterion, based on 20 years of use in various investigations.

Francis Ysidro Edgeworth (1883*a*) questions the universal and indiscriminate use of the normal (Gaussian) law of error in the following words: "The Law of Error is deducible from several hypotheses, of which the most important is that every measurable (physical observation, statistical number, &c.) may be regarded as a function of an indefinite number of elements, each element being subject to a determinate, although not in general the same, law of facility. Starting from this hypothesis, I attempt first, to reach the usual conclusion by a path which, slightly diverging from the beaten road, may afford some interesting views; secondly, to show that the exceptional cases in which that conclusion is not reached are more important than is commonly supposed" (pp. 300–301). Later in the same paper (pp. 305–306), he writes: "I submit, in the absence of evidence to the contrary, that non-exponential [non-Gaussian] laws . . . do occur in *rerum naturâ*, that the 'ancient solitary reign' of the exponential [Gaussian] law of error should come to an end". Edgeworth (1883*b*) begins a paper on the method of least squares with a philosophical discussion of the differences between the approaches of Gauss and Laplace, between most probable results and most advantageous results, and between minimizing mean square errors and mean absolute errors. He proceeds to the question of how to treat outlying observations. He proposes a method of weighting the observations which is the same as that proposed by Stone (1873*b*). In a later paper [Edgeworth (1887*a*), p. 373 (footnote)], he acknowledges Stone's priority, of which he was unaware at the time he wrote this paper.

M. H. Doolittle (1884) in discussing the rejection of doubtful observations, divides errors into two classes, *instructive errors* (those that indicate error in other observations) and *uninstructive errors* (blunders in recording, pointing on wrong objects, etc.). He asserts that the larger an instructive error is the more instructive it is, and the more important it is that the observation containing it *should not* be rejected; on the other hand, the larger an uninstructive error is, the more important it is that the observation *should* be rejected. Intelligent rejection of an observation is therefore, he says, based upon a comparison of the antecedent probability of the occurrence of an instructive error of the magnitude involved and that of the occurrence of an uninstructive error of the same magnitude. He criticizes Peirce's criterion on the ground that Peirce takes two probabilities, both functions of probabilities of instructive error, and balances them against each other.

The year 1884 saw the publication of two books on the adjustment of observations by the method of least squares. Both authors also consider the question of the rejection of outlying observations. Merriman (1884) advocates the use of Chauvenet's criterion, but he also discusses two other criteria – Peirce's and a new one based on Hagen's deduction of the law of error. Moreover, he states (p. 169): "In general, it should be borne in mind that the rejection of measurements for the single reason of discordance with others is not usually justifiable unless that discordance is considerably more than indicated by the criterions. A mistake is to be rejected, and an observation giving a residual greater than $4r$ or $5r$ [r = probable error] is to be regarded with suspicion, and be certainly rejected if the notebook shows any thing unfavourable in the circumstances under which it was taken." Thomas Wallace Wright (1884) advocates rejecting an observation whose residual is greater than five times the probable error (or three times the mean square error); in the second edition [Wright & John Fillmore Hayford (1906)], this rule is restated in slightly modified form.

References (Glossary of Code Letters, page 173)

Galilei, Galileo (1632). *Dialogo sopra i due massimi sistemi del mondo: Tolemaico, e Copernicano*. Landini, Florence. [English translation, *Dialogue concerning the two chief world systems, Ptolemaic and Copernican*, by Stillman Drake (with foreword by A. Einstein). Univ. of Calif. Press, Berkeley, 1953.] (TE)

Cotes, Roger (1722). Aestimatio errorum in mixta mathesi, per variationes partium trianguli plani et sphaerici. *Opera Miscellanea*, pp. 1–22. (Appended to his *Harmonia Mensurarum*.) Cantabrigiae. (TE, AV, WA)

Euler, Leonhard (1749). *Pièce qui a Remporté le Prix de l'Academie Royale des Sciences en 1748, sur les Inegalités du Mouvement de Saturn et de Jupiter*. Paris. (TE, LR, MA, MM)

Mayer, Johann Tobias (1750). Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmographische Nachrichten und Sammlungen* for 1748, **1**, 52–183. (TE, LR, MA)

Maire, Christopher; Boscovich, Roger Joseph (1755). *De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani gradus, et corrigendam mappam geographicam, jussu, et auspiciis Benedicti XIV Pont. Max. suscepta*. Romae. [French translation, *Voyage Astronomique et Géographique dans l'État de l'Église, entrepris par l'Ordre et sous les Auspices du Pape Benoit XIV, pour mesurer deux degrés du méridien, et corriger la Carte de l'État ecclesiastique*. Paris, 1770.] (TE, LR, TO)

Simpson, Thomas (1756). A letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the advantage of taking the mean of a number of observations, in practical astronomy. *Philosophical Transactions of the Royal Society of London* for 1755, **49** (1), 82–93. (TE, AV, AM)

Boscovich, Roger Joseph (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientiarum et Artum Instituto Atque Academia Commentarii*, **4**, 353–396. (TE, AV, LR, LF, TO, AM, [MD])

Simpson, Thomas (1757). An attempt to shew the advantage arising by taking the mean of a number of observations, in practical astronomy. *Miscellaneous Tracts on Some Curious, and Very Interesting Subjects in Mechanics, Physical-Astronomy, and Speculative Mathematics*, pp. 64–75. J. Nourse, London. (TE, AV, AM)

Boscovich, R. J. (1760). De recentissimis graduum dimensionibus, et figura, ac magnitudine terrae inde derivanda. *Philosophiae Recentioris, a Benedicto Stay in Romano Archigynasis Publico Eloquentare Professore, versibus traditae, Libri X, cum adnotianibus et Supplementis P. Rogerii Joseph Boscovich, S. J.*, Tomus II, pp. 406–426, esp. 420–425. Romae. [French translation included in note appended to 1770 French edition of Maire and Boscovich (1755).] (TE, LR, LF, OS)

Lambert, J. H. (1760). *Photometria, sive di Mensura et Gradibus Luminis, Colorum et Umbrae* (esp. Arts. 271–306). Augustae Vindelicorum, Augsburg. [Review, *Nova Acta Eruditorum* (1760), 564–578.] [TE, AV, ML, AM, TO)

Short, James (1763). Second paper concerning the parallax of the sun determined from the observations of the late transit of Venus, in which this subject is treated of more at length, and the quantity of the parallax more fully ascertained. *Philosophical Transactions of the Royal Society of London*, **53**, 300–345. (TE, AV, AM, DA, EX, TO)

Lambert, J. H. (1765*a*). Theorie der Zuverlässigkeit der Beobachtungen und Versuche. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Vol. 1, pp. 424–488. Berlin. (Second edition, 1792.) (TE, AV, LR, NR, AM, TO)

Lambert, J. H. (1765*b*). Anmerkungen und Zusätze zur practischen Geometrie. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Vol. 1, page nos. unknown. Berlin. (Second edition, 1792.) (TE, AV, MM, AM, [MR])

Lagrange, J. L. (1774). Mémoire sur l'utilité de prendre le milieu entre les résultats de plusieurs observations; dans lequel on examine les avantages de cette méthode par le calcul des probabilités; et où l'on resoud differents problèmes relatifs à cette matière. *Miscellanea Taurinensia* for 1770–73, **5**, 167–232 (esp. Problème X, 225–229). [Reprinted in *Œuvres de Lagrange*, Vol. 2, pp. 173–234. Gauthier-Villars, Paris, 1868.] (TE, AV, AM)

Laplace, P. S. (1774). Memoire sur la probabilité des causes par les événemens. *Mémoires de Mathématique et de Physique Presentés . . . par Divers Savans*, **6**, 621–657 (esp. Problème III: Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène, 634–644). [Reprinted in *Œuvres Complètes de Laplace*, Vol. 8, pp. 27–65. Gauthier-Villars, Paris, 1891.] (TE, AV, EA, AM, MD, MR, OS)

Bernoulli, Daniel (1778). Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. *Acta Academiae Scientiorum Petropolitanae*, **1** (1), 3–23 (Memoirs). [English translation by C. G. Allen, *Biometrika*, **48** (1961), 3–13.] (TE, AV, AM, MD, MR, ML, TO, OS)

Euler, L. (1778). Observationes in praecedentem dissertationem. *Acta Academiae Scientiorum Petropolitanae*, **1** (1), 24–33 (Memoirs). [English translation by C. G. Allen, *Biometrika*, **48** (1961), 13–18.] (TE, AV, M4)

Laplace, P. S. (1781). Mémoire sur les probabilités. *Mémoires de l'Académie royale des Sciences de Paris*, **Année 1778**, 227–332 (esp. 322–332). [Reprinted in *Œuvres Complètes de Laplace*, Vol. 9, pp. 383–485. Gauthier-Villars, Paris, 1893.] (TE, AV, AM, ML, EA)

Bernoulli, Jean III (1785). Milieu. *Encyclopédie Méthodique*, Vol. II, pp. 404–409. Paris. (Second edition, 1789, entitled *Dictionaire Encycl. des Mathématiques*.) (TE, AV, LR, NR, AM, MD, MR, OS, ML, LF, TO)

Laplace, Pierre Simon (1786). Mémoire sur la figure de la terre. *Mémoires de l'Académie royale des Sciences de Paris*, **Année 1783**, 17–46. [Reprinted in *Œuvres Complètes de Laplace*, Vol. 11, pp. 3–32. Gauthier-Villars, Paris, 1895.] (TE, [AV], LR, MM, [MR])

Laplace, P. S. (1793). Sur quelques points du système du monde. *Mémoires de l'Académie royale des Sciences de Paris*, **Année 1789**, 1–87 (esp. Sur le degrés mesurés des méridiens, et sur les longeurs observées du pendule, 18–43). [Reprinted in *Œuvres Complètes de Laplace*, Vol. 11, pp. 477–558. Gauthier-Villars, Paris, 1895.] (TE, LR, LF, MM, OS)

Laplace, P. S. (1799). *Traité de Mécanique Céleste*, Vol. 2. J. B. M. Duprat, Paris. [Reprinted as Vol. 2 of *Œuvres Complètes de Laplace*. Gauthier-Villars, Paris, 1878.] Part I, Book III, ch. v., pp. 116–165 (esp. 134–165).] (TE, [AV], LR, LF, MM, [AM], [MR], [MD])

Prony, R. (1804). *Recherches Physico-Mathématiques sur la Théorie des Eaux Courantes* (esp. pp. xvii–xxxii). L'Imprimerie Imperiale, Paris. (TE, LR, LF, MM, OS)

Trembley, Jean (1804). Observations sur la méthode de prendre les milieux entre les observations. *Mémoires de l'Academie Royale des Sciences et Belles Lettres de Berlin, Classe de Mathématique*, **Année 1801**, 29–58. (TE, AV, AM)

Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Courcier, Paris (esp. Appendice sur la méthode des moindres quarrés, pp. 72–80). [Second edition, with supplement, 1806.] (TE, LR, LS, MM, TO)

Puissant, Louis, (1805). *Traité de Géodésie, ou Exposition des Méthodes Astronomiques et Trigonométriques*, Vol. 1. Paris. [Volume 2, 1819; Second edition (2 vols.), 1842.] (TE, LR, LS)

Svanberg, Jons (1805). *Exposition des opérations faites en Lapponie, pour la détermination d'un arc du méridien en 1801, 1802 et 1803*, ... Stockholm. (TE, LR, LF, MM)

von Zach, Franz Xaver (1805). Versuch einer auf Erfahrung gegrundeten Bestimmung terrestrischer Refractionen. *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, **11**, 389–415, 485–504. (TE, AV, LR, AM, M4, TO)

Delambre, J. B. (editor) (1806–10). *Base du Systéme Métrique Décimal, ou Mesure de l'Arc du Méridien Compris entre les Parallèles de Dunkerque et Barcelone, Executée en 1792 et Années Suivantes, par MM. Méchain et Delambre*. Baudouin, Paris. Vol. 1, 1806; Vol. 2, 1807; Vol. 3, 1810. (TE, AV, LR, LS, MM, [LF], TO)

Gauss, C. F. (1806). II Comet vom Jahr 1805. *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, **14**, 181–186. (TE, LR, LS)

von Lindenau, B. A. (1806). Über den Gebrauch der Gradmessungen zur Bestimmung der Gestalt der Erde. *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, **14**, 113–158. (TE, LR, LS, LF, OS)

Adrain, Robert (1808). Research concerning the probabilities of the errors which happen in making observations. *Analyst*, **1**, 93–109. (TE, AV, LR, LS)

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Frid. Perthes et I. H. Besser, Hamburgi. [Reprinted 1906 in *Werke*, Band VII, Königlichen Gesellschaft der Wissenschaften, Göttingen, pp. 1–280; partial English translation by Hale F. Trotter in TR #5, Statistical Techniques Research Group, Princeton University, 1957, pp. 127–147.] (TE, AV, LR, AM, MD, MR, LS, MM, LF)

Bessel, F. W. (1810). *Untersuchungen über die scheinbare und wahre Bahn der grossen Cometen von 1807*. Königsberg. [Review, *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, **22** (1810), 205–212]. (TE, LR, LS)

Laplace, P. S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et leur application aux probabilités. *Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France*, **Année 1809**, 353–415; supplement, 559–565. [Reprinted in *Œuvres Complètes de Laplace*, Vol. 12, pp. 301–353. Gauthier-Villars, Paris, 1898.] (TE, AV, LS, AM, ML, EA)

Gauss, C. F. (1811). Disquisito de elementis ellipticis Palladis ex oppositionibus 1803, 1804, 1805, 1807, 1808, 1809. *Commentationes Societatis Gottingensis*, **1**, 26 pp. [Partial English translation by Hale F. Trotter in TR #5, Statistical Techniques Research Group, Princeton University, 1957, pp. 148–156.] (TE, LR, LS)

Laplace, P. S. (1811*a*). Mémoire sur les intégrales définies, et leur application aux probabilités, et specialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France*, **Année 1810**, 279–347. [Reprinted in *Œuvres Complètes de Laplace*, Vol. 12, pp. 357–412. Gauthier-Villars, Paris, 1898.] (TE, AV, LR, AM, LS, EA)

Laplace, P. S. (1811*b*). Du milieu qu'il faut choisir entre les résultats d'un grand nombre des observations. *Connaissance des Tems*, **Ann. 1813**, 213–223. [German translation, *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, **25** (1812), 105–120. (TE, AV, LR, AM, LS)

Laplace, P. S. (1812). *Théorie Analytique des Probabilités*. Courcier, Paris. Book II, ch. iv, pp. 304–348; second supplement (Application du calcul des probabilités aux opérations géodésiques), 1818. [Third edition, 1820, containing three supplements, reprinted 1886, together with a fourth supplement added in 1825, as *Œuvres Complètes de Laplace*, Vol. 7. Gauthier-Villars, Paris. Book II, ch. iv, pp. 309–354; second supplement, pp. 531–580.] (TE, AV, LR, AM, MM, LF, LS, EA, OS)

Delambre, J. B. (1813). *Abrégé d'Astronomie, ou Leçons Élémentaire d'Astronomie Théorique et Pratique*. Courcier, Paris. (TE, AV, LR, MD, LF)

Mathieu, Claude Louis (1813–14). Sur les expériences du pendule, faites par les navigateurs espagnols, en différens points du globe. *Connaissance des Tems*, **Année 1816**, 314–332. (TE, LR, LF, LS, MM, EX, OS)

Legendre, A. M. (1814). Méthode des moindres quarrés, pour trouver le milieu le plus probable entre les résultats de différentes observations. *Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France*, **Année 1810** (2), 149–154. (TE, AV, LR, AM, LS, TO)

van Beeck-Calkoen, J. F. (1816). Over de Theorie der Gemiddelde Waardij. *Verhandlingen der K. Nederlandandsch Instituut van Wetenschappen*, **2**, 1–19. (TE, AV, LR, AM, LF)

Gauss, C. F. (1816). Bestimmung der Genaugkeit der Beobachtungen. *Zeitschrift für Astronomie und verwandte Wissenschaften*, **1**, 185–196. [Reprinted 1880 in *Werke*, Band IV, Königlichen Gesellschaft der Wissenschaften, Göttingen, pp. 109–117; English translation by Hale F. Trotter in TR #5, Statistical Techniques Research Group, Princeton Univ., 1957, pp. 156–167.] (TE, DI, MD, AD, SD)

Adrain, Robert (1818*a*). Investigation of the figure of the earth, and of the gravity in different latitudes. *Transactions of the American Philosophical Society* (New Series), **1**, 119–135. (TE, [AV], LR, LS, LF, OS)

Adrain, Robert (1818*b*). Research concerning the mean diameter of the earth. *Transactions of the American Philosophical Society* (NS), **1**, 353–366. (TE, LR, LS)

Anonymous (1821). Dissertation sur la recherche du milieu le plus probable, entre les résultats de plusieurs observations ou expériences. *Annales de Mathématiques Pures et Appliquées*, **12**, 181–204. (TE, AV, AM, MD, MR, LS, OS, TO)

Gauss, C. F. (1823). Theoria combinationis observationum erroribus minimis obnoxiae. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, **5**; German summary, *Göttingische Gelehrte Anzeigen* (1821), 321–327 (pars prior) and (1823), 313–318 (pars posterior). [Reprinted 1880 in *Werke*, Band IV, Königlichen Gesellschaft der Wissenschaften, Göttingen, pp. 3–53, 95–104; English translation by Hale F. Trotter in TR #5, Statistical Techniques Research Group, Princeton University, 1957, pp. 1–82.] (TE, LR, LS, LF, MM)

Cauchy, Augustin-Louis (1824). Sur le système des valeurs qu'il faut attribuer a deux éléments déterminés par un grand nombre d'observations, pour que la plus grande de toutes les erreurs, abstraction faite du signe, devienne un minimum. *Bulletin de la Société Philomathique* (Paris), **1824**, 92–99. (TE, AV, LR, MM)

Fourier, J. B. J. (1824). Solution d'une question particulière au calcul des inégalités, second extrait. *Histoire de l'Académie des Sciences pour 1824*, pp. xlvii–lv. Reprinted in *Œuvres de Fourier* (ed. Gaston Darboux), Vol. 2, pp. 325–328. Gauthier-Villars, Paris, 1890. (TE, LR, LF, MM, EX, OS)

Poisson, S. D. (1824, 1829). Sur la probabilité des résultats moyens des observations. *Connaissance des Tems*, **Année 1827**, 273–302; **Année 1832**, 3–22. (TE, AV, LS)

Ivory, James (1825). On the method of the least squares. *Philosophical Magazine*, **65**, 1–10, 81–88, 161–168. (TE, AV, LR, LS, EA)

Muncke, G. W. (1825). Beobachtung. Gehler's *Physikalisches Wörterbuch*, Second edition. Leipzig. Vol. I, pp. 884–912. (TE, AV, LR, LS, TO)

Ivory, James (1826). On the method of the least squares. *Philosophical Magazine* **68**, 161–165. (TE, LR, LS)

Gauss, C. F. (1828). Supplementum theoriae combinationis observationum erroribus minimis obnoxiae. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores* **6**; German summary, *Göttingische Gelehrte Anzeigen* (1826), 1521–1527. [Reprinted 1880 in *Werke*, Band IV, Königlichen Gesellschaft der Wissenschaften, Göttingen, pp. 57–93, 104–108; English translation by Hale F. Trotter in TR #5, Statistical Techniques Research Group, Princeton University, 1957, pp. 83–126.] (TE, LR, LS)

Hauber, C. F. (1830*a*). Über die Bestimmung der Genauigkeit der Beobachtungen. *Zeitschrift für Physik und Mathematik*, **7**, 286–314. (TE, AV, DI, AM, MD, SD, AD)

Hauber, C. F. (1830*b*). Verallgemeinerung der Poisson'schen Untersuchungen über die Wahrscheinlichkeit der mittleren Resultate der Beobachtungen in den "Additions à la *Connaissance des Temps de 1827*". *Zeitschrift für Physik und Mathematik*, **7**, 406–429. (TE, AV, LR, LS)

Hauber, C. F. (1830–32). Theorie der mittleren Werthe. *Zeitschrift für Physik und Mathematik*, **8**, 25–26, 147–149, 295–315, 443–445; **9**, 302–322; **10**, 425–457. (TE, AV, DI, LR, AM, SD, MA, LS)

von Riese, C. (1830) Eine Recension. *Jahrbucher für Wissenschaftliche Kritik*, **1** (1), 269–284. (TE, LR, LS)

Cauchy, A. L. (1831). Mémoire sur le système de valeurs qu'il faut attribuer à divers élémens déterminés par un grand nombre d'observations pour que la plus grande de toute les erreurs, abstraction faite du signe, soit un minimum. *Journal de l'École Polytechnique*, **13** (20), 175–221. [Lith. MS, 1814]. (TE, AV, LR, NR, MM)

Fourier, J. B. J. (1831). *Analyse des Équations Déterminées, Première Partie*. Didot Frères, Paris. (TE, LR, LF, MM, EX, OS)

Encke, J. F. (1832–34). Über die Methode der kleinsten Quadrate. *Berliner Astronomisches Jahrbuch* for **1834**, 249–304; for **1835**, 255–320; for **1836**, 253–309. (TE, LR, LS)

Poncelet, J. V. (1835). Sur la valeur approchée linéaire et rationnelle des radicaux de la forme $\sqrt{(a^2+b^2)}$, $\sqrt{(a^2-b^2)}$ etc. *Journal für die Reine und Angewandte Mathematik*, **13**, 277–291. (TE, LR, MM, EX)

Lejeune Dirichlet, Peter Gustav (1836). Über die Frage, in wiefern die Methode der kleinsten Quadrate bei sehr zahlreicher Beobachtungen unter allen linearen Verbindungen der Bedingungsgleichen als das vorteilhafteste Mittel zur Bestimmung unbekannter Elemente zu betrachten sei (extract). *Abhandlungen der Akademie der Wissenschaften* (Berlin), **33**, 67–68. Reprinted in Lejeune Dirichlet's *Werke*, Vol. I, pp. 281–282 (extract); Vol. II, pp. 347–351 (complete). G. Reimer, Berlin, 1889 and 1897; reprinted by Chelsea Publishing Co., New York, 1969. (TE, AV, AM, MD, DI, SD, LR, LS, LF)

Cauchy, Augustin (1837). Mémoire sur l'interpolation. *Journal de Mathematiques Pures et Appliquees* (1), **2**, 193–205. [Lith. MS, 1835; English translation, *Philosophical Magazine* (3), **8**, 459–468.] (TE, LR, NR, CM)

Hagen, G. H. L. (1837). *Grundzüge der Wahrscheinlichkeitsrechnung*. Ernst & Korn, Berlin. [Third edition, 1882.] (TE, AV, LR, LS, LF, OS, TO)

Poisson, S. D. (1837) *Recherches sur la Probabilité des Jugements*. Bachelier, Paris (esp. Chapter 4). (TE, AV, LS)

Bessel, F. W. (1838). Untersuchungen uber die Wahrscheinlichkeit der Beobachtungsfehler. *Astronomische Nachrichten*, **15**, 369–404. (TE, LR, LS)

Bessel, F. W., Baeyer, J. J. (1838). *Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreieckskelten*. Berlin. [Reprinted in part in *Abhandlungen von Freidrich Wilhelm Bessel* (edited by Rudolf Engleman), Vol. 3, pp. 62–138. Wilhelm Engelmann, Leipzig, 1876.] (TE, LR, LS, TO)

Stampfer, S. (1839). Ueber das Verhältniss der Wiener Klafter zum Meter. *Jahrbucher des K. K. Polytechnisches Institutes* (Vienna), **20**, 145–176. (TE, AV, DI, AM, TO)

Gerling, C. L. (1843). *Die Ausgleichungs-Rechnungen des Practischen Geometrie, oder die Methode der Kleinsten Quadrate mit ihren Anwendungen für Geodätische Aufgaben*. Friedrich und Andreas Perthes, Hamburg und Gotha. (TE, AV, LR, LS, TO)

Donkin, W. F. (1844). An Essay on the Theory of the Combination of Observations. *Transactions of the Ashmolean Society*, Oxford. [French abridgement, *Journal de Mathématiques Pures et Appliquées* (1), **15** (1850), 297–332.] (TE, AV, LR, LS)

Ellis, R. L. (1844). On the method of least squares. *Transactions of the Cambridge Philosophical Society*, **8**, 204–219; summary in *Proceedings of the Cambridge Philosophical Society*, **1** (1), 5–6. (TE, AV, LR, LS)

Quetelet, L. A. (1846). *Lettres à S. A. R. le Duc Régnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités Appliquée aux Sciences Morales et Politiques*. Hayez, Bruxelles. (English translation by O. G. Downes, London, 1849.] (TE, AV, DI, [MD], QD)

De Morgan, Augustus (1847). Theory of probabilities. *Encyclopaedia of Pure Mathematics* (*Encyclopaedia Metropolitana*), Pt. II, pp. 393–490. (TE, LR, LS)

Ellis, R. L. (1850). Remarks of an alleged proof of the "Method of Least Squares", contained in a late number of the *Edinburgh Review*. *Philosophical Magazine* (3), **37**, 321–328, 462. [Reprinted in *Ellis's Mathematical Writings* (Cambridge, 1863), pp. 51–62.] (TE, LR, LS)

Herschel, Sir John F. W. (1850). Quetelet's *On Probabilities*. *Edinburgh Review*, **92** (185), 1–30. [Reprinted in *Herschel's Essays from the Edinburgh and Quarterly Reviews* (Longman, Brown and Co., London, 1857), pp. 365–465.] (TE, AV, LR, LS, MD)

Donkin, W. F. (1851). On certain questions relating to the theory of probabilities. *Philosophical Magazine* (4), **1**, 353–368, 458–466; **2**, 55–66. (TE, LR, LS)

Bienaymé, J. (1852). Sur la probabilité des erreurs d'après la méthode des moindres carrés. *Journal de Mathématiques Pures et Appliquées* (1), **17**, 33–78. (TE, LR, LS)

Peirce, Benjamin (1852). Criterion for the rejection of doubtful observations. *Astronomical Journal*, **2**, 161–163. (TE, TO, PC)

Bienaymé, Jules (1853*a*). Remarques sur les différences qui distinguent l'interpolation de M. Cauchy de la méthode des moindres carrés, et qui assurent la superiorité de cette methodé. *Comptes Rendus de l'Académie des Sciences de Paris*, **37**, 5–13. (TE, LR, LS, CM)

Bienaymé, Jules (1853*b*). Considérations a l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C. R. Acad. Sci. Paris*, **37**, 309–324; discussion, 324–326. [Reprinted 1867 (without discussion) in *Journal de Mathématiques Pures et Appliquées* (2), **12**, 158–176.] (TE, LR, LS, CM)

Cauchy, Augustin (1853*a*). Mémoire sur l'évaluation d'inconnues déterminées par un grand nombre d'équations approximatives du premier degré. *C. R. Acad. Sci. Paris*, **36**, 1114–1122. (TE, LR, LS, CM)

Cauchy, Augustin (1853*b*). Mémoire sur l'interpolation, ou Remarques sur les Remarques de M. Jules Bienaymé. *C. R. Acad. Sci. Paris*, **37**, 64–68; further remarks by Bienaymé, 68–69. (TE, LR, LS, CM)

Cauchy, Augustin (1853*c*). Sur la nouvelle méthode d'interpolation comparée à la méthode des moindres carrés. *C. R. Acad. Sci. Paris*, **37**, 100–109. (TE, LR, LS, CM)

Cauchy, Augustin (1853*d*). Mémoire sur les coefficients limitateurs ou restricteurs. *C. R. Acad. Sci. Paris*, **37**, 150–162. (TE, LR, LS)

Cauchy, Augustin (1853*e*). Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables. *C. R. Acad. Sci. Paris*, **37**, 198–206; discussion (remarks by Bienaymé), 197–198, 206. (TE, AV, LR, LS, CM)

Cauchy, Augustin (1853*f*). Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature. *C. R. Acad. Sci. Paris*, **37**, 264–272. (TE, AV, LR, LS, CM, MM)

Cauchy, Augustin (1853*g*). Sur la plus grande erreur à craindre dans un résultat moyen, et sur le système de facteurs qui rend cette plus grande erreur un minimum. *C. R. Acad. Sci. Paris*, **37**, 326–334. (TE, AV, LR, LS, CM, MM)

Tchebychef, P. L. [Chebyshev, P. L.] (1854). Théorie des mécanismes connus sous le nom de parallélogrammes. *Mémoires Présentés a l'Académie Impériale des Sciences de St. Pétersbourg par Divers Savants*, **7**, 539–568. Reprinted in *Œuvres de P. L. Tchebychef* (ed. A. Markoff & N. Sonin), Vol. I (1899), pp. 111–143. Imprimerie de l'Académie Impériale des Sciences, St. Pétersbourg. (TE, LR, NR, MM, EX)

Bertrand, J. (1855). Sur la méthode des moindres carrés. (Lettre à M. Elie de Beaumont, accompagnant l'envoi d'un exemplaire de la traduction des Mémoires de Gauss.) *C. R. Acad. Sci. Paris*, **40**, 1190–1192. (TE, AV, LR, AM, LS)

Gould, B. A., Jr. (1855). On Peirce's criterion for the rejection of doubtful observations, with tables for facilitating its application. *Astronomical Journal*, **4**, 81–87. (TE, TO, PC)

Lloyd, Humphrey (1855). On the mean results of observations. *Transactions of the Royal Irish Academy*, **22**, 61–73; abstract, *Proceedings of the Royal Irish Academy*, **4**, 180–183. (TE, AV, AM, DI, RA)

Airy, G. B. (1856). Letter from Professor Airy, Astronomer Royal, to the Editor. *Astronomical Journal*, **4**, 137–138. (TE, TO, PC)

Winlock, Joseph (1856). On Professor Airy's objections to Peirce's criterion. *Astronomical Journal*, **4**, 145–147. (TE, TO, PC)

Petzval, Joseph (1857). Fortsetzung des Berichtes über optische Untersuchungen. *Sitzungsberichte der Mathematisch-Naturwissenschaftliche Klasse Königliche Akademie der Wissenschaften, Vienna*, **24**, 129–144. (TE, LR, LS, MM)

Bienaymé, I. J. (1858). Mémoire sur la probabilité des erreurs d'après la méthode des moindres carrés. *Académie des Sciences de Paris, Mémoires Presentés par Divers Savants*, **15**, 615–663. (TE, LR, LS)

von Andrae, C. G. (1860). Udvidelse af en af Laplace i *Mécanique celeste* angivet Methode for Bestemmelsen af en unbekjendt Störrelse ved givne unmiddelbare Jagttagelser. *Oversight af Kgl. Danske Videnskabernes Selskabs over Forhandlingen* (Copenhagen), **1860**, 198–225. (TE, LR, LF)

Airy, George Biddell (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*. Macmillan and Co., London. (TE, LR, LS, TO)

Schott, Charles A. (1861). Account of Cauchy's interpolation formula. *Report of the U.S. Coast Survey* for **1860**, 392–396. (TE, LR, NR, CM)

Bartlett, W. P. G. (1862). On the empirical interpolation of observations in physics and chemistry. *American Journal of Science*, **34**, 27–33. (TE, LR, NR, LS, CM)

Chauvenet, William (1863). Method of least squares. Appendix to *Manual of Spherical and Practical Astronomy*, Volume 2 (J. B. Lippincott, Philadelphia), pp. 469–566; tables, 593–599. [Appendix and related tables reprinted 1868 as *A Treatise on the Method of Least Squares*.] (TE, LR, LS, TO, PC, CC)

De Morgan, Augustus (1864). On the theory of the errors of observation. *Transactions of the Cambridge Philosophical Society*, **10**, 409–427. (TE, AV, LR, AM, LS)

Todhunter, Isaac (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*. Macmillan and Co., London. [Reprinted 1949, Chelsea Publishing Co., New York.] (TE, AV, LR, AM, LF, EA, MM, LS, ML, M4, OS)

Stone, E. J. (1868). On the rejection of discordant observations. *Monthly Notices of the Royal Astronomical Society*, **28**, 165–168. (TE, TO, SC)

Jordan, W. (1869). Ueber die Bestimmung der Genaugkeit mehrfach wiederholter Beobachtungen einer Unbekannten. *Astronomische Nachrichten*, **74**, 209–226. (TE, AV, DI, AM, MD, SD, AD)

Todhunter, Isaac (1869). On the method of least squares. *Transactions of the Cambridge Philosophical Society*, **9**, 219–238. (TE, LR, LS)

Abbe, Cleveland (1871). A historical note on the method of least squares. *American Journal of Science and Arts* (3), **1**, 411–415. (TE, AV, LR, AM, LS)

Zachariae, G. (1871). *De Mindste Qvadraters Methode*. Nyborg. (TE, LR, LS)

Glaisher, J. W. L. (1872). On the law of facility of errors of observation and on the method of least squares. *Memoirs of the Royal Astronomical Society*, **39**, 75–124. (TE, AV, LR, AM, MD, LS, LF, EA, TO, PC)

Helmert, F. R. (1872). *Ausgleichungsrechnung nach der Methode der Kleinsten Quadrate*. B. G. Teubner, Leipzig-Berlin; second ed., 1907; third ed. (with supplement by H. Hohenner), 1924. [English transl., USAF Aeronautical Chart and Information Center, St Louis, 1958.] (TE, AV, LR, LS)

Glaisher, J. W. L. (1873). On the rejection of discordant observations. *Monthly Notices of the Royal Astronomical Society*, **33**, 391–402. (TE, TO, GC)

Peirce, C. S. (1873). On the theory of errors of observations. *Report of the Superintendent of the United States Coast Survey* (for the year ending November 1, 1870), Appendix No. 21, pp. 200–224 and Plate No. 27. U.S. Government Printing Office, Washington. (TE, AV, AM, WA, DA, DI, SD, LR, LS, EX, TO, PC)

Stone, E. J. (1873*a*). On the most probable result which can be derived from a number of direct determinations of assumed equal value. *Monthly Notices of the Royal Astronomical Society*, **33**, 570–573. (TE, AV, AM)

Stone, E. J. (1873*b*). On the rejection of discordant observations. *Monthly Notices of the Royal Astronomical Society*, **34**, 9–15. (TE, TO, CC, SC, GC, S2)

Todhunter, Isaac (1873). *A History of the Mathematical Theories of Attraction and the Figure of the Earth, from the time of Newton to that of Laplace*, Volumes I and II. Macmillan and Company, London. [Reprinted in two volumes bound as one by Dover Publications, Inc., New York, 1962.] (TE, LR, LF, MM, LS)

Fechner, G. Th. (1874). Ueber den Ausgangswerth der kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinerung. *Abhandlungen der Königliche Sächsische Gesellschaft der Wissenschaft zu Leipzig*, **18** [*Math. Phys. Classe* **11**], 1–76. (TE, AV, AM, MD, PM)

Glaisher, J. W. L. (1874). Note on a paper by Mr Stone, "On the rejection of discordant observations". *Monthly Notices of the Royal Astronomical Society*, **34**, 251. (TE, TO, SC, GC)

Jevons, W. Stanley (1874). *The Principles of Science*. Macmillan and Co., London–New York. Second edition, 1877; other editions, 1887, 1924, 1958. (TE, AV, AM, GM, HM, LR, LS, EX, TO, PC, CC)

Stone, E. J. (1874). Note on a discussion relating to the rejection of discordant observations. *Monthly Notices of the Royal Astronomical Society*, **35**, 107–108. (TE, TO, SC, GC)

Faye, H. E. (1875). Note accompagnant la présentation d'une notice autographiée sur la méthode des moindres carrés. *C. R. Acad. Sci. Paris*, **80**, 352–356. (TE, AV, LR, AM, LS)

Galton, Francis (1875). Statistics by intercomparison with remarks on the law of frequency of error. *Philosophical Magazine* (4), **49**, 33–46. (TE, AV, DI, MD, QD)

Laurent, H. (1875). Sur la méthode des moindres carrés. *Journal de Mathématiques Pures et Appliquées* (3), **1**, 75–80. (TE, AV, LR, LS)

Safford, T. H. (1876). On the method of least squares. *Proceedings of the American Academy of Arts and Sciences*, **11**, 193–201. (TE, LR, LS)

Merriman, Mansfield (1877). List of writings relating to the method of least squares with historical and critical notes. *Transactions of the Connecticut Academy of Arts and Sciences*, **4** (1), 151–232. (All)

Peirce, Benjamin (1878). On Peirce's criterion. *Proceedings of the American Academy of Arts and Sciences* **13** [N.S. **5**], 348–349; remarks by Charles A. Schott, 350–351. (TE, TO, PC)

Edgeworth, F. Y. (1883*a*). The law of error. *Philosophical Magazine* (5), **16**, 300–309. (TE, AV, LR, LS)

Edgeworth, F. Y. (1883*b*). The method of least squares. *Philosophical Magazine* (5), **16**, 360–375. (TE, AV, LR, LS, PC)

Doolittle, H. M. (1884). The rejection of doubtful observations (abstract). *Bulletin of the Philosophical Society of Washington* (*Mathematical Section*), **6**, 153–156. (TE, AV, DI, LS, EX, PC, SC, GC)

Merriman, Mansfield (1884). *A Textbook on the Method of Least Squares*. John Wiley & Sons, New York; sixth edition, 1892; second printing, 1893. (TE, AV, LR, AM, LS, TO, PC, CC, MC)

Wright, T. W. (1884). *A Treatise on the Adjustment of Observations by the Method of Least Squares*. D. van Nostrand, New York. (TE, TO, PC, CC, SC, GC, WC)

Glossary of Code Letters

AC Arley's criterion (for rejection of outliers)
AD average (absolute) deviation
AM arithmetic mean
AR Anscombe's rules (for rejection of outliers)
AS average slope (of regression line)
AV average (all types)
BC Bertrand's criterion (for rejection of outliers)
BM Brown-Mood estimators (of regression parameters)
BT best two (out of three)
CC Chauvenet's criterion (for rejection of outliers)
CM Cauchy's method (of interpolation)
CT (Bliss)-Cochran Tukey criterion (for rejection of outliers)
CU Cucconi's criterion (for rejection of outliers)
DA discard averages [trimmed means]
DC Dixon's criterion (for rejection of outliers)
DD discard deviation
DH differences at half range
DI dispersion (measures of)
EA equal areas (under joint p.d. curve) [Laplace's "most advantageous method"]
EM Edgeworth's modification (of Stone's second criterion)
EX extremes (largest and smallest values in sample)
FC Ferguson's criterion (for rejection of outliers)
GA Gastwirth estimators
GC Glaisher's criterion (for rejection of outliers)
GE geometric midrange
GG geometric range
GM geometric mean
GR Goodwin's rule (for rejection of outliers)
GS Grubbs' criterion (for rejection of outliers)
HA Hodges' alternative (to Hodges-Lehmann estimator)
HC Heydenreich's criterion (for rejection of outliers)
HL Hodges-Lehmann estimator
HM harmonic mean
HO Hogg's estimator

HS Hulme-Symms alternative (to the rejection of outliers)
HU Huber's estimator
IC Irwin's criterion (for rejection of outliers)
IR interquartile range
JA Jeffreys' alternative (to the rejection of outliers)
KC Kudô's criterion (for rejection of outliers)
LA Laurent's analogue (of Thompson's criterion)
LD largest (absolute) deviation
LF least (absolute sum of) first (powers) [Laplace's "method of situation"]
LN least number of deviations (least sum of zero powers)
LR linear regression
LS least squares
LW linearly weighted means
MA method of averages
MC Merriman's criterion (for rejection of outliers)
MD median
MG method of group averages
MK McKay's criterion (for rejection of outliers)
ML maximum likelihood
MM minimax method [minimize maximum residual]
MO mode
MQ median-quartile average
MR midrange
MS method of successive differences
MT median and two other order statistics
MW multivariate Wilks' criterion (for rejection of outliers)
MZ Mazzuoli's criterion (for rejection of outliers)
M4 maximum (sum of) fourth (powers of p.d.f. of errors)
NC Nair's criterion (for rejection of outliers)
NM Newcomb's method (of treating outliers)
NR nonlinear regression
NS Nair-Shrivastava method (of curve fitting)
OM Ogrodnikoff's method (of treating outliers)
OS order statistics
PA plus approximative méthode [most approximative method]
PC Peirce's criterion (for rejection of outliers)
PM power means
QA quadratic average (mean)
QD quartile deviation [semi-interquartile range]
QM quasi-midrange [quasi-median]
QN quantiles
QR quasi-range
RA range
RC Rohne's criterion (for rejection of outliers)
RL robust estimators of location
RM range method
SC Stone's (first) criterion (for rejection of outliers)
SD standard deviation [or variance $\equiv (SD)^2$]
SM Stewart's method (criterion) (for rejection of outliers)
SR semirange
ST Student's rule (for rejection of outliers)
SW Switzer's estimator
S2 Stone's second criterion (for rejection of outliers)
TC Tippett's criterion (for rejection of outliers)
TE theory (of) errors
TF Tukey's FUNOR-FUNOM procedure
TJ Topsoe-Jensen criterion (for rejection of outliers)
TM Thompson's method (criterion) (for rejection of outliers)
TO treatment of outlying observations
VC Vallier's criterion (for rejection of outliers)
WA weighted average
WC Wright's criterion (for rejection of outliers)
WH Wright-Hayford (criterion) (for rejection of outliers)
WI Winsorization
WM Winsorized means
WR Walsh's rule (criterion) (for rejection of outliers)
YE Yanagawa's estimator

Résumé

Un problème très important en statistique mathématique est celui de trouver la meilleure équation linéaire ou non linéaire permettant d'exprimer la relation entre un variable dépendante et une ou plusieurs variables indépendantes. Les données sont des observations, sujettes à des erreurs aléatoires, en nombre supérieur à celui des paramètres de l'équation de régression, observations sur la variable dépendante et les valeurs correspondantes de la ou des variables indépendantes, qui peuvent être connues exactement ou également sujettes à des erreurs aléatoires. Les problèmes portent aussi sur le choix des meilleures mesures de tendance centrale et de dispersion des observations. Les solutions optimales de ces trois problèmes dépendent de la distribution des erreurs aléatoires. Si l'on suppose que les valeurs de la ou des variables indépendantes sont connues exactement, et que les erreurs sur les observations de la variable dépendante sont distribuées normalement, il est bien connu que la moyenne est la meilleure mesure de la tendance centrale, l'écart-type la meilleure mesure de la dispersion, et la méthode des moindres carrés la meilleure méthode d'ajustement d'une équation de régression. Des hypothèses différentes conduisent à des choix différents. La plupart des praticiens ont tendance à admettre l'hypothèse de normalité et à ne pas se soucier des conséquences que cela peut entraîner lorsqu'elle n'est pas justifiée. Un autre problème se présente lorsque les données contiennent des observations anormales (aberrantes) provenant de distributions ayant une autre moyenne et (ou) un écart-type plus grand. De nombreuses méthodes ont été proposées pour le rejet ou la correction des valeurs aberrantes. L'auteur résume (chronologiquement) la volumineuse littérature sur la mesure de la tendance centrale et de la dispersion, la méthode des moindres carrés et ses nombreuses alternatives, le traitement des données aberrantes et la robustesse des estimations, puis recommande un ensemble de procédures simples et raisonnablement robustes.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Method of Least Squares and Some Alternatives - Part I-VI		5. TYPE OF REPORT & PERIOD COVERED Journal Article
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H. LEON HARTER		8. CONTRACT OR GRANT NUMBER(s) In-House
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Flight Dynamics Laboratory/FBRD Air Force Wright Aeronautical Laboratories (AFSC) Wright-Patterson AFB, OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F AF-7071 70710202
11. CONTROLLING OFFICE NAME AND ADDRESS 1974		12. REPORT DATE Aug 1974 - Apr 1976
		13. NUMBER OF PAGES 226
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Reprinted from International Statistical Review, Vol. 42, No. 2, pp. 147-174 (Aug. 1974), No. 3, pp. 235-264, 282 (Dec. 1974); Vol. 43, No. 1, pp. 1-44 (Apr. 1975), No. 2, pp. 125-190 (Aug. 1975), No. 3, pp. 269-278 (Dec. 1975); Vol. 44, No. 1, pp. 113-159 (Apr. 1976)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Central tendency	Midrange	Least absolute values
Dispersion	Standard Deviation	Minimax estimates
Regression	Mean deviation	Random errors
Arithmetic mean	Semirange	Normal distribution
Median	Least squares	Double exponential distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A very important problem in mathematical statistics is that of finding the best linear or non-linear regression equation to express the relation between a dependent variable and one or more independent variables. Given are observations, each subject to random error, greater in number than the parameters in the regression equation, on the dependent variable and the related values of the independent variable(s), which may be known exactly or may also be subject to random error. Related problems are those of choosing the best measures of central tendency and dispersion of the observation. The best solutions of all

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

012 079

19. Uniform distribution
    Outliers
    Robust estimation
    Adaptive procedures

20. three problems depend upon the distribution of the random errors. If one assumes that the values of the independent variable(s) are known exactly and that the errors in the observations on the dependent variable are normally distributed, then it is well known that the mean is the best measure of central tendency, the standard deviation is the best measure of dispersion and the method of least squares is the best method of fitting a regression equation. Other assumptions lead to different choices. Most practitioners have tended to make the assumption of normality and not to worry about the consequences when it is not justified. Another problem arises when the data are contaminated by spurious observations (outliers) which come from distributions with different means and/or larger standard deviations. Many methods have been proposed for rejecting outliers or modifying them (or their weights). After summarizing (chronologically) the voluminous literature on measures of central tendency and dispersion, the method of least squares and numerous alternatives, the treatment of outliers and robust estimation, the author recommends a simple and reasonably robust adaptive procedure. Parts I-IV cover the time periods 1632-1884, 1885-1945, 1946-1964 and 1965-1974, respectively. Part V gives conclusions and recommendations and Part VI gives subject and author indexes.