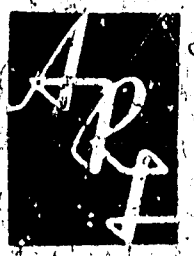
THE METHOD OF LEAST SQUARES AND SOME ALTERNATIVES

H. Leon Harter

Aerospace Research Laboratories
Wright-Patterson Air Force Base, Ohio

September 1972

ARL 72-0129
SEPTEMBER 1972



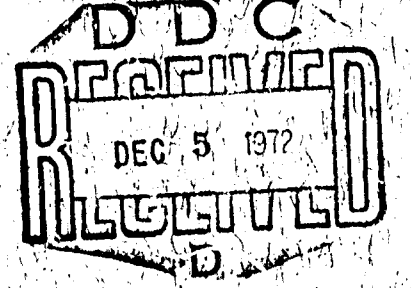
Aerospace Research Laboratories

THE METHOD OF LEAST SQUARES AND SOME ALTERNATIVES

H. LEON HARTER

APPLIED MATHEMATICS RESEARCH LABORATORY

PROJECT 7071



AIR FORCE SYSTEMS COMMAND
United States Air Force

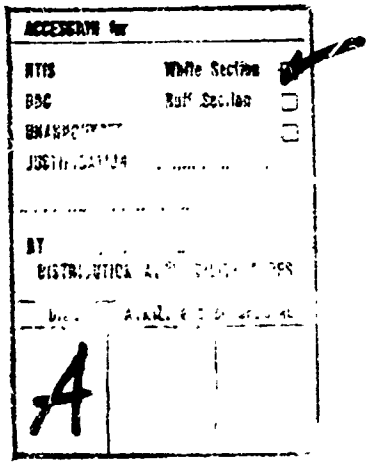
NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Agencies of the Department of Defense, qualified contractors, and other Government agencies may obtain copies from:

Defense Documentation Center
Cameron Station
Alexandria, VA 22314

This document has been released (for sale to the public) to:



National Technical Information
Services
Clearinghouse
Springfield, VA 22151

Copies of ARL Technical Reports should not be returned to the Aerospace Research Laboratories unless return is required by security considerations, contractual obligations, or notices on a specific document.

AIR FORCE/56780/26 October 1972 – 500

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)
Aerospace Research Laboratories (LB)
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio 45433

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b GROUP

3 REPORT TITLE
The Method of Least Squares and Some Alternatives

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Interim

5 AUTHOR(S) (First name, middle initial, last name)
H. Leon Harter

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
September 1972	~~238~~ 242	487

8a CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
Internal	
b. PROJECT NO 7071-02-11	ARL TR 72-0129
c DoD Element 61102F	9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. DoD Subelement 681304	

10 DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
TECH OTHER	Aerospace Research Laboratories (LB) Air Force Systems Command Wright-Patterson AFB, Ohio 45433

13 ABSTRACT

A very important problem in mathematical statistics is that of finding the best linear or nonlinear 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.

DD FORM 1473

I-H

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Observations						
Laws of error						
Averages						
Dispersion						
Regression						
Least first powers						
Least squares						
Least maximum deviation						
Outliers						
Skewness						
Kurtosis						
Robust estimation						

ARL 72-0129

THE METHOD OF LEAST SQUARES AND SOME ALTERNATIVES

H. LEON HARTER

APPLIED MATHEMATICS RESEARCH LABORATORY

SEPTEMBER 1972

PROJECT '671

Approved for public release; distribution unlimited.

AEROSPACE RESEARCH LABORATORIES
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

I-C

FOREWORD

This report was prepared by Dr. H. Leon Harter of the Applied Mathematics Research Laboratory, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio. The research was performed under Project 7071, "Research in Applied Mathematics", Work Unit 11, "Order Statistics and their Use in Testing and Estimation." The author wishes to express his appreciation to Miss Eva Brandenburg for painstakingly typing the reproducible copy and to TSgt Paul A. Dienstbier for preparing the figure.

ABSTRACT

A very important problem in mathematical statistics is that of finding the best linear or nonlinear 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. The reader who seeks a more sophisticated solution can choose one from among the

many given in the literature cited or devise one to fit the special conditions of his problem.

TABLE OF CONTENTS

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 (1972) 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 asscciated 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 nonlinear regression.

The problems of determining the best average or measure of central tendency and the best linear or nonlinear 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 thirteen 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 center 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 (page 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 the 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 twenty-seven observations on the position of a moon spot to write twenty-seven 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 divided 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 & 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 pp. 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 inversely 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 reevaluates 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 $b_1 \geq b_2 \geq b_3 \geq b_4 \geq b_5$ of the lines connecting the five observational points (x_i,y_i), i=1,2, ... , 5 to their centroid $(\bar{x},\bar{y})$.

According to Sheynin (1966), three works of Johann Heinrich Lambert (1760, 1765a, 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 (1765a), 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 endeavors 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 centers of gravity instead of the original observations. In the third work (1765b), 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.

After solving several problems concerning averages of observations having discrete error distributions, which are reminiscent of Simpson (1755, 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 $x_1<x_2<x_3$ 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 $x=p+(1/m)\log[1+(1/3)e^{-mp}-(1/3)e^{-mq}]$, which approaches the arithmetic mean (2p+q)/3 as m→0 and the median as m→∞; for 0<m<∞, 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, $x_1<x_2<x_3$, 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 (1765a), 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 center of the semicircle and determine the center of gravity of the area corresponding to the observations; take this point as the center of a new semicircle, and repeat the operation until the center of gravity and the center 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 $E = a+bx+cy+ \cdots$, where a,b,c, $\cdots$ are known coefficients which vary from one equation to another and x,y,$\cdots$ 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 fifteen measurements of the lengths 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 (1765a) 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 meter = 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)], 1/135. 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 meter.

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 fundemental theorem, which I myself have already used for twelve 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, ··· 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 center 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 [Adrain (1818a)].

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 $2n^{\text{th}}$ powers of the errors for n=1,2,3, ··· , but Gauss points out that

minimizing the sum of their squares (n=1) is simplest. Letting $n \to \infty$ [Laplace's method of situation] is equivalent to minimizing 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 $2n^{\underline{th}}$ 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 distribution, 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 twelve 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 eleven. Merriman (1877), p. 166, notes: "We find here for the first time the notation $[a\ b] = a'b' + a''b'' + a'''b''' + \cdots$ and also the algorithm for the solution of normal equations by successive substitution, since universally followed in lengthy computations ···."

Laplace (1811a) 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 of 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 (1811b)], 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 (1811a,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 (1811a) 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 $2n^{\text{th}}$ 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 recommends the method of least squares proposed by Legendre (1805) and Gauss (1809) for use whenever the number of observations is large. In the second supplement (first published in 1818), Laplace explains the method proposed by Boscovich (1757, 1760) [see also Laplace (1793, 1799)] based on minimizing the sum of the absolute values of the deviations subject to the restriction that their algebraic sum be zero, to which he gives the name "method of situation." He determines a condition under which this method is preferable to his own "most advantageous method," and 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 pp. 607-608, Delambre applies his method to the determination of the earth's eccentricity from the Delambre-Méchain observations.

Legendre (1814) sets the stage for a quotation from pp. 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 many later writers do, the generality of what Laplace actually proved about the method of least squares, which is that the method of least squares is best for the normal error law and asymptotically best (as the number of observations $n \to \infty$) 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 (1765a), 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 centers 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 $y=a+bx^2$ and $y=a+b\sqrt{x}$ as well as the linear regression equation of the form $y=a+bx$. 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 n^{th} root of the sum of the n^{th} 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 (1838), Foche (1852-54), 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 (1818a) 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% from those obtained by the method of least squares. In another paper, Adrain (1818b) 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 spheriod, given measurements of degrees of meridian.

In 1821 there appeared an anonymous paper whose authorship Czuber (1891a, 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 showed 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 n^{th} 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.

Two memoirs by Poisson (1824,1829), large parts of which are reproduced in a later work [Poisson (1837)] are, according to Merriman (1877), pp. 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 $\phi(x)=c\,e^{-h^2x^2}$. 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 analyzed Ivory's criticism of Laplace, which is 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 proved in the article cited by Ivory [see also Laplace (1810,1811a)] 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 & Robinson (1924) state that this method is substantially equivalent to reduction of a quadratic form to a sum of squares.

Carl Friedrich Hauber (1830a) 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) law. The mathematical expressions which he obtains agree with those given by Gauss (1816), as do the numerical results for the root-mean-square error and the mean absolute error. For the probable error of the median absolute error M he gives $0.78671\ w/\sqrt{s}$ (where w is the true value of M), which he approximates by $0.78671M/\sqrt{s}$, without mentioning that

Gauss incorrectly calculated the numerical coefficient (for which he gave the correct mathematical expression $\sqrt{\pi/8}e^{\rho^2}$, where ρ=0.4769363) to be 0.7520974. [Hauber's value is correct to within a unit in the fifth decimal place.] Hauber states as an advantage of the median absolute error that it is independent of the law of facility of error. This is true in the sense that it is unbiased for any law of error. As for most so-called "distribution-free" estimators, however, its precision and its efficiency relative to other estimators do depend on the law of error. Hauber (1830b) extends the results of Poisson (1824,1829) to the case of two or more quantities to be determined.

Hauber (1830-32), in a six-part article, discusses the theory of averages. In the first three parts he deals with arithmetic means, both population means (expected values) and sample means, along with the root-mean-square error and probable error of the latter and various applications. In the fourth he discusses cases in which the values of the quantities of interest are not given by the observations, but functions of these quantities. In the fifth and sixth parts he deals with methods of solving the set of observational equations, greater in number than the number of quantities to be determined. He gives three methods: (1) the method of averages, which he attributes to Tobias Mayer; (2) the method of least squares; and (3) a hybrid method in which the number of equations is reduced by Mayer's method to a manageable number (still greater than the number of quantities to be determined), which are then solved by the method of least squares.

C. von Riese (1830) summarizes the results of the paper by Gauss (1823), which contains Gauss' second "proof" of the method of least squares, and the 1828 supplement thereto, which contains his algorithm for solving the normal

equations. The author mentions the earlier "proof" by Gauss (1809) based on the postulate, which von Riese attributes to Cotes (1722?) that the arithmetic mean is the best average of a number of observations. He also refers to the work of Laplace (1810,1811a,1812) on the method of least squares and on the rival method based on minimizing the sum of the absolute errors, as well as the earlier work of Boscovich (1757?) on the latter method and the articles of J. Bernoulli (1785) on various averages and of Muncke (1825) on the method of least squares.

Johann Franz Encke (1832-34) gives an exposition, in three parts, of the method of least squares. The first part contains the "proof" by Gauss (1809), an attempt by Encke to demonstrate that the arithmetic mean is necessarily the best average of a number of observations, a discussion of weights and probable errors, and two tables of the probability integral $(2/\sqrt{\pi})\int e^{-t^2} dt$. In this part, the author also gives a proof, which he credits to his colleague Dirichlet, of the expression for the probable error of the median absolute deviation, which Gauss (1816) had given without proof. He gives the value of the numerical coefficient in this expression as 0.786716 [correct to six decimal places], as compared with the values 0.78671 [Hauber (1830a)] and 0.7520974 [Gauss(1816)], neither of which he mentions,as well as a numerical example of application of this method to actual data. The second part of the article contains Gauss' algorithm for the solution of normal equations and his method of determining weights, while the third deals with conditioned observations.

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) $y = au + bv + cw + \cdots$ where a,b,c, $\cdots$ are constant coefficients. Now the question is, 1st, 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; 2ndly, 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 $y_i (i=1, \cdots, n)$ and the corresponding values of x_i (and hence of $u_i, v_i, w_i, \cdots$) related by n equations (2) $y_i = a_i u_i + b_i v_i + c_i w_i + \cdots$. The author proposes successive approximations based on neglecting all but one, two, $\cdots$ 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 (1765a). 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 practiced by such earlier authors as Boscovich (1755, 1757), Lambert (1760,1765a), 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 meter, 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), Laplace (1812), Gauss (1823) 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 endeavors 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 interquatile 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 weights 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 (1923).

Jules Bienaymé (1852) reviews the development of the theory of least squares from the early work of Legendre (1805), Gauss (1809,1823) and Laplace (1811a,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 (1853a) 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 (1853a) argues, however, that the two methods are completely different and even that a contradiction exists. Cauchy (1853b) maintains that in many investigations his method of interpolation is preferable to the method of least squares. Cauchy (1853c) 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 (1853d,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 (1853f) shows that the most probable values may sometimes differ from those found by the method of least squares. Bienaymé (1853b) reviews some of Cauchy's articles [Cauchy (1853d,e,f)] and maintains that the mean of the sum of squares of the errors is under all circumstances a measure of precision of the observations. Cauchy (1853g) 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.

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 rewrites 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 Biddell 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 favor 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 systems 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 favorable 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 $2m^{\text{th}}$ powers of the residuals, where m is an integer which tends to infinity as a limit.

C. G. von Andrae (1869) 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 m^{th} ordered observation, where m=n/2. By analogy, if s observations are to be used, he chooses the m_i^{th} ordered observations

$(i=1,2,\cdots,s)$, where $n_i = n/(s+1)$.

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 ϵ that may be expected to occur in n observations is $2n\int_{\epsilon}^{\infty} \phi(t)dt = n\psi(\epsilon)$, where $\phi(t) = e^{-t^2/2}/\sqrt{2\pi}$, an observation deviating from the mean by an amount greater than ϵ should be rejected if the quantity $n\psi(\epsilon)$ 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 DeMorgan (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.

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, 1799,1810,1811a,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 $n^{\underline{th}}$ root of the mean of the $n^{\underline{th}}$ 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 (1818a,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 the method of 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 the errors 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 (1873a,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.

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 von Andrae (1860) and was proved by 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.

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 1444 measurements of an angle of approximately 15°, 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 favorable remarks on Peirce's criterion, based on twenty years of use in various investigations.

Francis Ysidro Edgeworth (1883a) 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 (1883b) 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 (1873b). In a later paper [Edgeworth (1887a), p. 373 (footnote)], he acknowledges Stone's priority, of which he was unaware at the time he wrote this paper.

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 unfavorable 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.

4. THE AWAKENING (1885-1945)

Edgeworth (1885) discusses the choice of measures of central tendency and of variability. He insists that, while the arithmetic mean and the root-mean-square deviation from it are most accurate for samples from a normal population, other measures (median and mean absolute deviation or quartile deviation) are more convenient and little less accurate, while for other populations they may be more accurate as well as more convenient. With regard to measures of variability he writes (pp. 188-189): "'''When the observations really conform to a [normal] probability-curve, there are several formulae for the modulus [$c=\sigma\sqrt{2}$, where σ is the standard deviation] which are little inferior to the above [root-mean-square error] in respect of accuracy, and two of them which are superior in respect of convenience. If we call the preferential method the method of mean square of errors, one of the rival methods might be called the method of mean first power; the other the method of mean zero powers. ''' [The last] method is that described by Mr. Galton [(1875)] '''; the same in principle as that which was employed by Quetelet [(1846)]. The essence of this method is to note the points between which

are comprised quarters (eights or other fractions) of the total number of observations, and then to equate the distance thus given by observations to the corresponding multiple of the modulus as assigned by theory. For example, if we take two points so that between them there occur half the total number of given observations, and outside each of them a quarter of the total number, the distance between these two points ought theoretically to be equal--is equatable--to twice the modulus x 0.476." On pp. 190-191, Edgeworth discusses the use of the median. He finds that the fluctuation [the square of the modulus c(or twice the variance σ^2)] of the distribution of medians of sets each consisting of m observations is equal to the reciprocal of $2my^2$, where y is the maximum ordinate of the probability curve divided by its area.

Edgeworth (1886) explores in detail the relative advantages of the arithmetic mean, the median, and the mode, with less attention given to other possible means. On the grounds of precision, he declares the arithmetic mean to be superior to the others for the normal law and others near it, but says the median is better "when the apex of the curve is very high and its extremities very much extended." (p. 167). "In respect of convenience, [the Mode] has a considerable advantage over the Arithmetical Mean and a less marked advantage over the Median." (p. 168). The author makes passing reference to the use of the quartile deviation by Quetelet (1846) and of quartiles and deciles by Galton (1875) in estimating the probable error, and to the method of situation and the most advantageous method [Laplace (1799, 1812)].

Simon Newcomb (1886) considers the problem of combining a number of

observations of the same quantity so as to obtain the best result. He raises two objections to the criterion for rejection of doubtful observations proposed by Peirce (1852): (1) It disregards any a priori knowledge of the probable error of the observations and seeks to determine it from the observations themselves; and (2) It does not take account of the fact that the a priori probability of an observation varies from one observer to another. Given n observations assumed to have come from a generalized law of error which is a mixture of m normal laws, with proportions p_i having precision $h_i (i=1,\ldots,m)$, Newcomb says the best result is a weighted mean (with the weights proportional to the probabilities of the hypotheses on which they depend) of m^n weighted means of the observations, each mean being obtained by making a hypothesis concerning the distribution of the m measures of precision among the n observations.

Edgeworth (1887a), in discussing the diversity of methods for the treatment of discordant observations, makes the following statement (p. 365): "Different methods are adapted to different hypotheses about the cause of a discordant observation; and different hypotheses are true, or appropriate, according as the subject-matter, or the degree of accuracy required, is different." He specifies three hypotheses and divides the different methods of treating discordant observations into four groups. He rates the first three types of method on their appropriateness under each of the hypotheses, deferring discussion of the fourth method (use of the median instead of the arithmetic mean) to a later paper [Edgeworth (1887d)].

Edgeworth (1887b) drops Boscovich's Condition (I) [that the sums of

positive and negative deviations be equal in magnitude] and uses only his Condition (II) [that the sum of the absolute values of the deviations be a minimum], which, as we have already seen, requires the choice of the median rather than the arithmetic mean. Edgeworth gives the following description (pp. 279-282) of his procedure: "It is proposed here to treat those difficulties in the reduction of observations which are peculiar to the case of plural quaesita. ··· Consider, first, the simple case in which there are only two quaesita. Let the given equations be of the form $a_1x+b_1y-w_1=0, a_2x+b_2y-w_2=0$, &c., where w_1, w_2, &c., are observations subject to equal error. According to the usual procedure we obtain for one locus (of the sought point xy) the 'normal equation' $a_1[a_1x+b_1y-w_1]+a_2[a_2x+b_2y-w_2]+$ &c.$=0$; which may be thus interpreted. Substitute any assigned value for y in the original equations. Of the n values for x thus presented, the (weighted) Arithmetical Mean is given by substituting the assigned value for y in the 'normal' equation. The analogous procedure is to find a locus such that if we substitute any assigned value of y in the original equations, the Median of the corresponding n values of x may be given by the locus. The series of points, which in the case of the Arithmetical Mean is obtained by a single stroke of analysis, must, in the case of the Median, be traced one by one. That is, we must substitute in the given equations successive values of y (e.g. $0, \delta, 2\delta$, &c.), find the Median value for x corresponding to each assigned y, and plot the series of points. A second Median Curve is afforded by the Medians of the y components; and the intersection of these Median Curves gives the Median Point. The method is perfectly general. As an illustration we may take the case of two quaesita, x and y, the equations for which involve only one of the variables. The Mean loci are

in this case lines parallel to the axes. And it follows from considerations which I have elsewhere [Edgeworth (1887d)] put together, that the Median, as compared with the Arithmetical Mean, affords a solution nearly as good when the typical [normal] probability-curve prevails, and better when the observations are 'discordant'". The author (p. 280) calls the Boscovich-Laplace method [Laplace (1799), Sec. 40] a "remarkable hybrid between the Method of Least Squares and the Method of Situation", because Boscovich's Condition (I) requires that deviations be taken from the arithmetic mean, as in the method of least squares, instead of from the median, as in Edgeworth's version of the method of situation, where that condition has been dropped.

Edgeworth (1887c) writes as follows (pp. 222-223) concerning the method developed in the preceding paper: "The method may be thus described in the case of two variables, x and y. Find an approximate solution by some rough process (such as simply adding together several of the equations so as to form two independent simultaneous equations). Take the point thus determined as a new origin, and substitute in the n (transformed) equations for one of the variables x a series of values $\pm \delta, \pm 2\delta$, &c. Corresponding to each of these substitutions we have n equations for y. For each of these systems determine the Median according to Laplace's Method of Situation. This series of Medians forms one locus for the sought point. A second locus is found by transposing x and y in the directions just given. The intersection of these loci is the required point. The method may be extended to any number of variables. ... The advantages claimed for the new method are that, while in the typical case of the laws of facility being all [normal] Probability Curves, the generalized Method of Situation is only slightly less accurate, and considerably less

laborious, than the Method of Least Squares; in the abnormal case of Discordant Observations the proposed method is not only more convenient, but better. It is much to be wished that some practical astronomer would give this method a trial by employing it in some laborious and important calculation."

Edgeworth (1887d) considers the question of the choice of means in the special case of discordant observations. On p. 270 he writes: "The criterion whether the Median or Arithmetic Mean is the better reduction is presumably the character of the correlated Probability-Curve. The reduction which corresponds to the smaller Modulus is presumably the better; since thus we obtain a smaller 'probable' error ... . Which of the reductions will have the smaller Modulus will depend on the character of our facility-curve. For [normal] Probability-Curves, and presumably functions in their neighborhood, it is shown by Laplace [(1812), Supplement 2] that the Arithmetic Mean has the advantage. But for curves whose head reaches high, while their extremities strech out far, the Median has the advantage. Now the grouping of Discordant Observations is apt to assume this form. Accordingly the Median is proposed as the Mean proper to this class of observations. If we have been deceived by the appearance of Discordance ... and the facility-curve was really a normal Probability-Curve, yet we shall have lost little by taking the Median instead of the Arithmetic Mean. For the error of the former is of the same order as (only 1.3 [times] greater than) the error of the latter. And, if the observations are really discordant, the derangement due to the larger deviations will not be serious, as it is for the Arithmetic Mean." The author gives three numerical examples.

Edgeworth (1887e) gives two tests of symmetry, one based on a comparison

of the means of the positive and negative deviations from the arithmetic mean and the other on a comparison of the arithmetic mean and the median of the observations. The critical value which he gives for the latter test is erroneous; since the arithmetic mean and the median are correlated, the variance of their difference is not the sum of their variances, but that sum decreased by twice their covariance.

H. H. Turner (1887) comments as follows (pp. 466-470) on the method of Edgeworth (1887b,c): "Edgeworth invites attention to a method of reducing observations relating to several quantities, which he has suggested as a substitute for the ordinary process of the 'Method of Least Squares'. I have applied this method to an example for a particular case of two variables, and venture to offer the following remarks and suggestions for consideration. [Here follows a quotation from Edgeworth (1887c).] Some of the labour of this process, and sometimes the preliminary search for an approximate solution, may be avoided by the use of a graphical method [which the author describes]. ··· In the method of least squares the normal equations have a unique solution; but the intersection of two broken lines may be a series of points, and the two median loci may also have a common portion. The solution then becomes to some extent indeterminate. ··· It is possible that the number and distribution of these points of intersection afford real information as to the value and accordance of the observations. But, in practice, a single solution, although its singularity may be somewhat fictitious, is preferable to a variety; and unless some additional criterion for extracting a single solution from the median loci can be obtained, it is to be feared that we have here a somewhat serious objection to this method on the score of convenience. ··· Mr. Edgeworth

claims as advantages for the new method that (1) It is considerably less laborious than the Method of Least Squares. (2)In the case of Discordant Observations it is theoretically better. So far as my slight experience entitles me to express an opinion on these points, I should say that (1) is very doubtful. In trying a new method much time is liable to be wasted; but there would, I imagine, never be quite the same straightforwardness about the new method which makes the method of least squares so easy, although somewhat long. (2) is somewhat counterbalanced by the failure to give a unique solution".

Edgeworth (1888) restates (p. 184) the method he proposed a year earlier [Edgeworth (1887b,c)]: "A substitute for the Method of Least Squares has been proposed by me, based upon the following principle. The data being of the form $a_1x+b_1y \cdots - v_1=0$; $a_2x+b_2y \cdots -v_2=0$, &c., (where v_1, v_2, &c. are observations of equal worth), a solution is obtainable by taking x,y ··· such that the sum of the residuals (the left-hand members of the above written equations), each residual taken positively, should be a minimum". In a footnote (pp. 184-185) he writes: "This rule is derivable from the hypothesis that the law of error, the facility-curve under which the observations range, is of the form $y=(h/2)e^{-hx}$, x taken positively in both directions [Laplace's first law of error; see Laplace (1774)]. But the use of the rule does not commit us to the assumption of the hypothesis. The Method of Least Sum is in this respect exactly on a par with the Method of Least Squares. The rule of the latter Method is--Determine x and y so that the sum of the squares of the residuals may be the least possible. This rule is derivable from, and specially correlated with, the hypothesis that the law of facility is the [normal] Probability-curve. But it is thought legitimate by Laplace and other eminent authorities to apply the

rule even where the hypothesis is not assumed. No doubt the use of either method divorced from the law of facility appropriate to it is open to logical objections. But the difficulties are not greater for one method than for the other." The author continues (pp. 185-186): "The point thus designated must be on each of two, or more, loci analogous to the Normal equations of the ordinary method. Accordingly the intersection of the 'Median loci' was at first proposed by me as the solution. But Mr. Turner [(1887)] has shown that these loci are apt to have in common, not only several points, but even lines and spaces. ··· In this event common sense teaches us that we should adopt the middle of the indeterminate tract as the best point; and this presumption is confirmed by a formal calculation of utility such as Laplace [(1312)], in the simplest case of a single unknown quantity, has employed to discover the 'most advantageous point'." Having thus disposed of one of Turner's criticisms, Edgeworth endeavors (somewhat less successfully, it seems to the present writer) to answer the other, namely that the method is not less laborious than the method of least squares, as Edgeworth [(1887b,c)] had asserted.

Joseph Bertrand (1888a) states that the Gaussian law of probability is the only one for which, among several observations made under the same conditions, the mean value is the most probable. Given any other law for the probability of errors, it is possible to specify the combination of a series of measurements which will give the most probable value. The converse is not true; given a combination of observations, in most cases there is no probability law for which that combination gives the most probable value. For example, he says, there is no probability law for which the geometric mean or the harmonic mean of a number of observations is the most probable value. The latter part of

the paper deals with the rejection of observations, for which Bertrand (1888b) proposes a new criterion. The author repeats and elaborates on these results in his book [Bertrand (1889)].

Faye (1888) points out that the greatest discrepancy from the mean of a set of observations is not likely to be counterbalanced by a discrepancy of nearly the same magnitude but of opposite sign, and hence unless the number of observations in the set is very large it is likely to have an undue influence on the arithmetic mean, so that the arithmetic mean of all the observations is not the most probable value. Given forty observations, he computes the arithmetic means of the largest and smallest observations, the second largest and second smallest, and so on (the midrange and the quasi-midranges). The former has the value 4.315 and the latter range from 3.815 to 3.995. He attributes this discrepancy to the fact that the largest deviation from the arithmetic mean, 6.35-3.93 = 2.42, is not matched by a comparable deviation in the opposite direction. He therefore recommends rejecting the largest observation, which changes the arithmetic mean from 3.93 to 3.87, a value which he considers to be more probable. In general,however he holds that observations should be rejected only if they are considered doubtful at the time they are made, or at least before any computations have been made. Otherwise the calculator can too easily make the results agree with his preconceived and sometimes erroneous opinion.

Galton (1888,1889) advocates the use of the median as a measure of central tendency and the quartile deviation as a measure of dispersion. In the latter publication, he writes: "The median, M, has three properties. The

first follows immediately from its construction, namely, that the chance is an equal one, of any previously unknown measure in the group exceeding or falling short of M. The second is, that the most probable value of any previously unknown measure in the group is M. ··· The third property is that whenever the curve of the Scheme [of distribution] is symetrically disposed on either side of M ···, then M is identical with the ordinary Arithmetic Mean or Average." (p. 41). "As the M[median] measures the Average Height of the curved boundary of a Scheme, so the Q[quartile deviation] measures its general slope. ··· Our Q has the further merit of being practically the same as the value which mathematicians call the 'Probable Error' ···." (p.53).

Emanuel Czuber (1890) advocates the method of maximum likelihood to find the most probable system of values of m unknown elements p,q,r, ··· in the law of error $\phi(x)$ of specified form, given that $x_1, x_2, \cdots, x_n$ are the errors of n observations. He attributes this method to Gauss (1809), but we have already seen that it was used even earlier by Lambert (1760) and by D. Bernoulli (1778). The author enumerates three conditions under which the usual method of finding the maximum likelihood estimates, based on solving likelihood equations formed by equating to zero the partial derivatives of the likelihood function $\Omega=\phi(x_1)\phi(x_2)\cdots\phi(x_n)$, fails.

J. E. Estienne (1890) endeavors to prove that "the best value to adopt, as a measure of a quantity of which experience has furnished values tainted by accidental errors, is, in every case, the median value···." In Chapter I, he proves the following theorem (p. 241): "The most probable value, determined by the rule of the median value, is that for which the arithmetic sum of the deviations is a minimum." Chapter II is devoted to the proposition that the

rule of the median value is independent of the law of errors and should be applied to the exclusion of every other rule, whatever be this law. Chapter III deals with the consequences of the law of the median value, of which the author gives several, including the use of the method of least first powers in solving n inconsistent equations in m unknowns (n>m). An example dealing with artillery fire is given in the appendix. The author is, of course, incorrect in declaring the universal superiority of the median, but no more so than earlier authors who insisted on that of the arithmetic mean.

Czuber (1891a) gives a detailed study of the theory of linear observational errors, the method of least squares, and the theory of errors in the plane and in space. The first six sections of Part 1 deal with laws of error [with particular emphasis on the normal (Gaussian) law, though Laplace's first law and others are mentioned]; the early work of Simpson (1756,1757) and Lagrange (1774) on the advantages of taking averages, of Laplace (1774, 1781) on his "most advantageous method", and of Daniel Bernoulli (1778) on the method of maximum likelihood; the work of Legendre (1805), Adrain (1808), Gauss (1809), Laplace (1812), and later writers on the method of least squares; and the problem of the choice of means and its relation to the choice between the method of least squares and rival methods. Sections 7 and 8 deal with estimation of the precision of a series of observations from the true errors and the apparent errors (residuals), respectively. Section 9 compares the (normal) error law with experience. Section 10 deals with the smallest and largest errors in a set of observations, and Section 11 deals with the treatment of outlying observations. Part 2 is concerned with the details of the computational procedure for the method of least squares, and Part 3 deals with

the theory of errors in the plane and in space.

Czuber (1891b) examines the rule proposed by Estienne (1890) and its consequences, and shows that the median and the method of least absolute first powers, far from being valid whatever the law of error, as asserted by Estienne, are, as pointed out by Glaisher (1872), tied just as firmly to the first law of error of Laplace (1774) as are the arithmetic mean and the method of least squares to Laplace's second (Gauss') law.

P. Pizzetti (1892) gives a useful summary of work to date on the theory of errors, with a bibliography of 503 items, but presents few if any new results.

Edgeworth (1893), in connection with a study of averages of correlated observations, proposes two principles of wider application: "(1)When observations are combined according to a system of weights different from that which is known to be best, it is in general advantageous to reject a certain class of the given observations. (2) When, as usual, the observations range under a [normal] probability curve, the median m corrected by the quartiles q_1 and q_2 affords a formula for the Mean, viz. $(1.2m + q_1 + q_2) \div 3.2$, which is more accurate than that method of combining such observations which has hitherto been supposed to be the most accurate, viz. the Arithmetic Mean. The principle may be applied with great ease and advantage to Discordant Observations." The author's statement that his new formula gives a result more accurate than the arithmetic mean for observations from a normal error law is incorrect. The source of his error lies in the assumption that the quartiles are independent of each other and of the median, whereas they are actually considerably correlated, as Karl Pearson (1920) has pointed out.

P. J. Éd. Goedseels and Paul Mansion (1893) discuss the theory of errors,

Goedseels observes that in reality no one has ever established the method of least squares in an absolutely conclusive manner, and that one should avoid assigning too great objective value to the results to which it leads. Mansion is of the same opinion. In reality, he says, the only definition which one can give of accidental errors is this: They are the errors which are eliminated by the method of least squares. But it is just to recall that Gauss took care to express himself with precision on what is arbitrary in the theory to which he gave such a perfect form. The great advantage of the method of least squares is that it allows the combination, in a simple and reasonable manner, of the results of observations of unequal value in a condensed form. This discussion is of interest because the budding dissatisfaction with the method of least squares expressed here later led to substantial contributions by the authors to the theory of rival methods.

Karl Pearson (1895), in connection with the first exposition of the Pearson system of frequency curves, discusses the relative position of mean, median, and mode of samples from a Pearson Type III distribution. He finds that the median lies about one third of the way from the mean to the mode.

Henri Poincaré (1896) discusses the theory of errors and the arithmetic mean, justification of the law of Gauss, errors in the position of a point, and the method of least squares. He expresses doubt as to the universal validity of the Gaussian law of error (and with it the arithmetic mean and the method of least squares),but offers no specific alternatives. Twice he raises the question as to whether one should reject outlying observations, but offers no definitive answer.

Fechner (1897) discusses various laws of error and various averages

(arithmetic mean, median, and mode), as well as the largest and smallest values, their difference (the range) and their sum (twice the midrange), and compares theoretical and observed values for several of these statistics.

Czuber (1899) applies the theory of probability to the results of measurements. He discusses the arithmetic mean and other averages, including the median and the power means [Fechner (1874)], and the estimation of the precision of a series of observations on the basis of the true errors. He mentions the method of situation of Laplace (1793,1799), based on the two conditions of Boscovich, as the first attempt at a systematic solution of the problem of solving an inconsistent system of observational equations. He devotes several sections to demonstrations of the method of least squares based on the work of Gauss (1809), Laplace (1812), and Gauss (1823) and related material. He also discusses measures of precision based on apparent errors and differences of observations, the comparison of theoretical and observed distributions, the largest and smallest errors in a series of observations, and the treatment of outlying observations.

Goedseels (1900) mentions three methods of solving a set of simultaneous equations greater in number than the number of unknowns--the method of Tobie Mayer (1750), the method of least squares, and the method of Cauchy (1837). He makes a detailed study of the method of Mayer and establishes the four following propositions: (1) The method of Mayer used in our days differs appreciably from the original method; (2) The modern method is susceptible to an important simplification; (3) There is room for returning, in certain cases, to the primitive procedure; (4) Both the primitive and modern procedures offer certain advantages not previously pointed out (especially in the case of only two unknowns).

A. A. Markoff (1900) devotes a chapter of his book on the calculus of probabilities to the method of least squares.

Goedseels (1901) proposes a simplification of the method of Cauchy (1837) for solving a system of m linear equations in n unknowns, m>n. Goedseels (1902) proposes an application of Cauchy's method to least squares.

Czuber (1903) discusses the theory of errors of observation, including laws of error, various measures of precision (including the mean absolute error, the square root of the mean square error, the probable error, the $m^{\underline{th}}$ root of the mean of the $m^{\underline{th}}$ powers of the absolute value of the error, and the mean difference of all pairs of observations), and the method of least squares.

J. C. Kapteyn (1903) develops the theory of skew frequency curves from a different point of view than that of Pearson (1895). He discusses the median and the method of calculating it for his theoretical curves, and compares its values in several examples with those of the mode and the arithmetric mean.

S. A. Saunder (1903) advocates the use of Peirce's criterion for the rejection of doubtful observations, for which he gives a new table of critical values. He also discusses the alternative to rejection of observations proposed by DeMorgan (1847) and Glaisher (1873).

Mansion (1906) summarizes important contributions to the history and the critique of the method of least squares contained in extracts of letters and papers of Gauss in the eighth volume of Gauss' collected works (published in 1900). He makes brief mention of the theory of combination of observations, introduced by Laplace (1786) [see also Laplace (1793,1799)], in which the largest error (in absolute value) is smaller than for any other system. He points out that Gauss (1809) criticized this method on the grounds that it uses for the final calculation of the unknowns only a number of equations

equal to the number of unknowns; the other equations are used only to decide the choice which one should make.

Goedseels (1909) proposes two methods, which he calls the most approximative method and the method of minimum approximation, for solving a system of n equations in p unknowns (n>p). In the most approximative method one assumes that the intervals (A_i,B_i) containing the respective residual errors r_i are known, and seeks to determine for each unknown (say x) the smallest interval (I,S) containing that unknown, i.e. an interval such that for every value of x less than I or greater than S, one or more of the residues r_i lie outside the given interval. If all observations are equally trustworthy and if positive and negative errors are equally likely, then $A_1 = A_2 = \cdots = A_n = A$, $B_1 = B_2 = \cdots = B_n = B$, and $-A=B=M$ (say). Consider a series of equations in a single unknown of the form x=m, having the same approximation M, and suppose the equations are arranged in order of increasing values of m: $x=m_1+r_1$, $x = m_2+ r_2$, $\cdots$, $x = m_n+ r_n$. Then the most approximative value is the midrange, $(m_1+ m_n)/2$, and the approximation of this value is the difference, $M-(m_n-m_1)/2$, between the given approximation, M, and the semirange, $(m_n-m_1)/2$. When $M-(m_n-m_1)/2=0$, the midrange $(m_1+ m_n)/2$ is the <u>exact</u> value of the unknown. When $M-(m_n-m_1)/2<0$, the data are absurd. When the limits of the errors are not known, Goedseels advocates use of the method of minimum approximation, which is the same as the method of Laplace (1786,1793,1799) in which the maximum absolute deviation (residual) is minimized; in the case of a single unknown, the result (average) obtained is the midrange. Goedseels discusses various other methods, including the method of least squares and the empirical methods of Mayer (1750) and Cauchy (1837). In summary, he states that

he prefers the most approximative method when the limits of error are known and the minimum approximation otherwise, especially in very important questions, even though the calculations become quite laborious for p>2; in questions of lesser importance, he says the method of least squares or even one of the empirical methods may be used to save labor.

Charles J. de la Vallée Poussin (1909,1911) states and proves the following theorems concerning the minimum approximation, which assume that any n of the first members of the system of equations (1) $a_i x + b_i y + \cdots + \ell_i u - m_i = 0 (i=1,2,\cdots,m;\ m>n)$ form a system of linearly independent expressions: I. The values of the unknowns which provide the minimum approximation M of the system (1) give at least n+1 residues attaining this limit M in absolute value. II. The minimum approximation and the corresponding values of the unknowns, for a system of n+1 equations in n unknowns, are obtained by general formulas. III. If m>n+1, the minimum approximation of a system of m equations in n unknowns is that of a certain system of n+1 equations which are part of the proposed system. One can deduce from these theorems an iterative procedure for determining the minimum approximation.

Goedseels (1910) summarizes the results given in his book [Goedseels (1909)] on the most approximate method, the method of minimum approximation, and the method of least squares, and applies all three methods to a numerical example involving the compensation of the coordinates of the vertices in a topographic survey. He insists that the method of least squares is not as good as the other two methods, even in the case of normally distributed errors, for which it gives the most probable values, since these values are inadmissible if they lie outside the interval (I,S) of the most approximative

method. In the numerical example, he first ascertains the the maximum error ε specified by the observer is admissible, i.e. that it is not less than the minimum approximation m. Then he proceeds to compensate the z-coordinates of the data points by the most approximative method and by the method of least squares. He rejects the results of the latter method as inadmissible, since the z-coordinate it gives for one point lies outside the interval given by the most approximative method. Finally, he ignores the maximum error stated by the observer and compensates the z-coordinates by the method of minimum approximation. The resulting values lie within the intervals given by the most approximative method, and the author shows that this must always be true if any admissible value $e \geq M$ (where M is the minimum approximation) is designated as ε by the observer.

Goedseels (1911) calls attention to an interesting study of the method of minimum approximation, including a simplification of that method, proposed by de la Vallée Poussin (1909,1911). He proposes, in turn, two other simplifications applicable both to that method and to the most approximative method. These simplifications apply only when there are only one or two unknowns, of which at least one is positive. The latter is really no restriction at all, since the data can be transformed so that at least one unknown is positive. The restriction to one or two unknowns is less serious than one might think, since one always proceeds by successive elimination of the unknowns, and the number of them is always eventually less than three; moreover it is in the final stages, where this condition is satisfied, that the methods in question become most complicated. In an abstract reprinted as a footnote on pp. 351-352, de la Vallée Poussin states that, with the simplifications proposed by

himself and by Goedseels, the calculations for the method of minimum approximation are even simpler than those for the method of least squares.

G. Udny Yule (1911), in a textbook that has gone through many editions, discusses averages [including the median (Mi) and its relation to the arithmetic mean (M) and the mode (Mo)], measures of dispersion [the range, the mean deviation about the median, and the quartile deviation $Q=(Q_3-Q_1)/2$, where Q_1 and Q_3 are the quartiles], and measures of skewness [$(M-Mo)/S \doteq 3(M-Me)/S$ (Pearson's measure, where S is the standard deviation) and $(Q_1+ Q_3- 2Mi)/2Q$]. He also discusses the standard errors of quantiles (median, quartiles, deciles, etc.) and of the semi-interquartile range (quartile deviation Q), as well as the correlation between the errors in two quantiles.

L. Tits (1912) states and proves three theorems which embody further simplifications, beyond those of de la Vallée Poussin (1911) and Goedseels (1911),of the most approximative method and the method of minimum approximation. He uses these theorems to simplify the solution of a numerical example given by Goedseels (1911).

Edward Lewis Dodd (1913) points out that Czuber (1891) recounted many of the attempts that have been made to relate the principle of the arithmetic mean as the most probable value with the Gaussian probability law, but quoted from Bertrand (1889), p. 180, an example to show that this law and principle are not strictly compatible. The author endeavors to show this incompatibility by other means. Specifically, he shows that, under certain conditions, the quadratic mean (root-mean-square) of two measurements from a Gaussian distribution has a greater probability than the arithmetic mean; also, there exist positive values of b (less than unity) for which the probability of bm is

greater than that of m (the arithmetic mean). The probability of the median of three or more measurements from a Gaussian distribution is, however, always less than that of the arithmetic mean, the ratio of probabilities approaching $\sqrt{2/\pi} \doteq .7979$ asymptotically.

Edgeworth (1913) gives empirical confirmation of Pearson's rule as to the relation between the arithmetic mean, the median, and the mode. He computes median, quartiles, deciles, and mean deviations for sums of 25 and of 16 random digits, and compares them with the theoretical values.

H. M. Goodwin (1913) discusses the rejection of observations, for which he gives a new criterion, which involves computing the arithmetic mean and the average deviation, omitting the doubtful observation, and rejecting that observation if its deviation from the mean is greater than or equal to four times the average deviation.

Mansion (1913) summarizes the work of various authors on three methods applied to the theory of errors, which involve minimizing respectively the largest error, the sum of the absolute values of the errors, and the sum of squares of the errors. He points out that these methods result in choice of the midrange, the median, and the arithmetic mean as the respective averages; also that an order-preserving (or order-reversing) transformation does not affect the method of minimum sum of absolute values, since the median of the transformed function is the transforming function of the median of the original data, but does affect the other two methods.

David Brunt (1917) discusses the law of error, making reference to the work of Gauss, Todhunter, and Glaisher. He deduces the law of error [Laplace's first] which results from the assumption that the median is the most probable

value. He also deals with the rejection of observations. He appears to favor use of the criterion of Wright & Hayford (1906), but he also states that Bessel opposed rejection of any observation unless the observer is satisfied that the external conditions produced some unusual source of error.

Warren M. Persons (1919) gives reasons for preferring the median to any other average of link relatives in computing indices of seasonal variation. P. J. Daniell (1920) discusses various measures of central tendency and of dispersion. Besides the usual arithmetic mean, median, standard deviation, mean numerical deviation, and quartile deviation, he also discusses discard averages and discard deviations. Since the middle observations (in order of magnitude) contain most of the information about central tendency and the extreme observations contain most of the information about dispersion, he proposes discarding some proportion (say 50%) of the outer observations in computing averages and of the inner observations in computing measures of dispersion.

Karl Pearson (1920) points out that Edgeworth (1893) was in error when he stated (p. 99, footnote) that the displacements of the two quartiles and the median are independent; they are, in fact, considerably correlated. The author determines the standard errors of quantiles of a sample of size N (assumed large) from any known population and the correlations of pairs of such quantiles.

R. M. Stewart (1920a) raises two objections to Peirce's criterion for the rejection of doubtful observations. First of all, the principle as stated by Peirce is erroneous when n (the number of observations to be rejected)is greater than unity. Even for the case n=1, Peirce's argument is based on an

unwarranted assumption, so that the most one can say is that if all residuals are less than the value obtained from Peirce's criterion, no observation should be discarded. The author remarks that Chauvenet's criterion for rejection of a single observation also contains an obvious fallacy. Stewart (1920b) proposes a new method for the treatment of discordant observations. Like Stone (1873b), Edgeworth (1883b), and Newcomb (1886), he assumes that the precision is not the same for all the observations, but he simplifies matters as much as possible by restricting the number of values of the precision constant h to two. He proposes a weighted mean which is a function of the residuals. He gives a method of finding the weights; once that has been done, the weighted mean can easily be found.

Dunham Jackson (1921) shows that, for each value of $p>1$, there is a definite number $x=x_p$ which minimizes the sum $S_p(x) = \sum_{i=1}^{n}|x-a_i|^p$, where $a_1, a_2, \cdots a_n$ are a set of real numbers. For $p=2$, x_p is the arithmetic mean of the a's. The limit of x_p, as $p\to 1$, is the median, while the limit of x_p, as $p\to\infty$, is the midrange.

Dodd (1922) studies the arithmetic mean, the median, the midrange, and other functions (averages) of the measurements with reference to their approximation to the so-called true value, by determining which has the greatest probability density at the true value. He summarizes his results as follows (p.158): "The present examination of functions of measurements is not exhaustive. The general conclusion, however, would probably be that in most cases where the law of error is symmetrical (the error function even) the arithmetic mean is better than other functions. The superiority of the arithmetic mean becomes somewhat doubtful: (1) When the number of measurements is small. ···

(2) When the probable error is not small compared with the arithmetic mean. ... (3) When the error curve,--as evidenced by the actual distribution of measurements--falls away from its maximum with some rapidity at first, but nevertheless persists at some distance from the maximum. In this case, the median may be better than the arithmetic mean. ... (4) When the error curve is perpendicular to the axis of errors, meeting this axis at equal distances from the origin. In this case, the average of the least and greatest measurements [the midrange] may be better than the average of all the measurements [the arithmetic mean]. ... "

Ronald Aylmer Fisher (1922) advocates the use of the sample median in estimating the central value of the Cauchy distribution, pointing out that the distribution of the sample mean is the same as that of a single observation, so that the sample mean is an entirely useless statistic. He also touches on the treatment of outlying observations, and lays a firm mathematical foundation for the method of maximum likelihood, here first given that name.

W. L. Crum (1923), in determining the indexes of seasonal variation in an economic series, takes the median of a series of link-relatives for a particular month as the unadjusted index for that month, as recommended by Persons (1919). To the observed data he fits (1) a normal curve, (2) a Charlier Type A curve, and (3) a composition of two normal curves. He notes that Yule (1911) has shown that if a distribution may be dissected into two normal distributions each of half the original frequency, and if the ratio between the two standard deviations is greater than 2.24, the median has a smaller probable error than the mean. He endeavors to extend this result to the case under consideration, in which the standard deviation of the larger portion, containing about 3/4 of the total number of cases, is 4.8 times that

of the smaller portion. He concludes that for this series and for many (but not all) economic series, the median is better than the mean.

Edgeworth (1923) restates the rationale of the method which he proposed much earlier [Edgeworth (1887b,c,1888)] for solving a redundant system of equations and amplifies the directions for its application given by Turner (1887). He then considers several numerical examples, and closes (pp. 1085-1088) with the following conclusions: "The accuracy of the double Median depends on much the same considerations as those which relate to the single Median. ··· The comparison of the (single or double) Median method with that of Least Squares is prejudiced by two misapprehensions exaggerating (a) one the defects of the Median, (b) the other the merits of Least Squares. It is presumed that determination by way of Medians is less exact because it sometimes leaves the segment of a line, or even (in the case of the double Median) a space within which no unique value is distinguished. But the comparative definiteness of the Arithmetic Mean is illusory, considering that the determination is liable to a probable error. ··· The Method of Least Squares enjoys an undue preference in virtue of its connection with the Normal Law of Error. For probably that law is not in general fulfilled by observations so perfectly as to justify the preference given. The preferability varies with the character of the observations. ··· When the curve representing the observations is quite abnormal [non-Gaussian] it is very possible that the Median should have the preference in respect of accuracy. ··· When the extremities of the curve representing the crude observations are abnormally protruberant, the Median is apt to be preferable. ··· In short, the use of the Median (single or double) is often easier, and sometimes more accurate, than the Method of Least Squares.

··· Altogether, we may conclude with Laplace that, in certain cases, the Method of Situtation is preferable to the Method of least Squares".

Edwin Bidwell Wilson (1923) re-examines the data of Crum (1923) and reaches the following conclusions (pp. 850-851): "With the exception of the extreme positive deviations which have not been well fitted by any of Professor Crum's three suggestions, these data give internal evidence of following Laplace's first law of error instead of his second law and should be fitted to that law. By simple graphical means using arith-log [semi-log] paper an extremely good fit [to Laplace's first law] may be had in a very few minutes' work (all that is necessary is to plot the grouped data, draw a straight line, and read the graph). ··· Professor Crum's plea for the use of the median in certain types of statistics is much reinforced by the behavior of these data when discussed in relation to the first law of error."

J. Haag (1924) studies the precision, for samples from a normal distribution, of the arithmetic mean, the median, and the quasi-midrange $X=(1/2)(x_p + x_{n-p+1})$, which is equal to the midrange for p=1 and to the median for $p = [(n+1)/2]$. He finds that the mean is the most precise, the median has asymptotic relative precision $\sqrt{2/\pi} \doteq 4/5$, and the midrange is the least precise, with asymptotic relative precision 0.

Jackson (1924), given a set of p simultaneous equations in n unknown quantities (p>n), studies the question of determining values for the unknowns so that these equations shall be approximatley solved, in the sense that the sum of the $m^{\underline{th}}$ powers of the absolute values of the errors is a minimum. For m=2, this is the classical problem of least squares. The author shows that the problem has at least one solution for every m>0 and a unique solution for

$m>1$. The limiting case as $m\to\infty$ is equivalent to the problem of minimizing the maximum error.

Edmund T. Whittaker & G. Robinson (1924) give the probable errors of the arithmetic mean and the median, as well as those of various measures of precision, based on the m^{th} powers of the absolute errors ($m=1,2,3,4,5,6$) and the median of the absolute deviations from the true value, all under the assumption that the samples are drawn from a normal distribution. They discuss the method of least squares at some length, giving an account of the contributions of Legendre (1805), Gauss (1809,1823,1828), Laplace (1812) and others. They mention three alternatives to the method of least squares: (1) the method of Tobias Mayer (1750);(2) the method of minimum approximation [Laplace (1799), Goedseels (1909), de la Vallée Poussin (1911)]; and (3) the method of Edgeworth (1887c,1888).

Julian Lowell Coolidge (1925) devotes one chapter of his book on probability to errors of observation. In his section on determination of the "best value" he states the following theorem, which he attributes to Fechner (1874): "The sum of the numerical values of the divergences of a number from a given series of numbers will be a minimum if the number in question be the median." He also includes sections on the Gaussian law of error, for which he gives Gauss' first deduction[Gauss (1809)], with passing mention of that of Hagen (1837) based on the composition of elementary errors, and on doubtful observations.

J. O. Irwin (1925a) finds a complicated expression for the exact distribution of the difference the p^{th} and $(p+1)^{\text{th}}$ individuals in order of magnitude (from largest to smallest) of a sample of size n from a normal distribution,

together with useful approximations for differences between first and second and between second and third individuals in order of magnitude, which he tabulates for selected values of n. Irwin (1925b) proposes a criterion for the rejection of observations based on these differences. Using the approximation developed in his earlier paper [Irwin (1925a)], he calculates the probabilities that the differences between the first and second and between the second and third individuals should be greater than λ times the standard deviation of the sampled population. If these probabilities become too small, he advocates rejecting the first (the last) or the first two (the last two) individuals as not belonging to the same homogeneous group as the remainder, so that a table of these probabilities for varying values of λ provides a criterion for the rejection of outlying observations.

Paul Lévy (1925) devotes a chapter in his book on probability to the theory of errors. One section deals with the determination of parameters of precision. He proposes using the interquartile distance to estimate twice the probable error (or 0.95a, where $a=\sigma\sqrt{2}$ and σ is the standard deviation). He gives two methods of determining the quartiles; later we shall present reasons for preferring a third method. Another section deals with the method of least squares; the author does not present any alternative methods.

Joseph Reilly, William Norman Rae & Thomas Sherlock Wheeler (1925) advocate the use of the criterion for rejection of observations given by Wright & Hayford (1906): "Reject each observation for which the residual exceeds 5 times the P. E. [probable error] for a single determination. Examine carefully each observation for which the residual exceeds 3.5 times the P. E. and reject it if any of the accompanying conditions are such as to produce lack of

confidence". The authors explain the determination of empirical constants by the method of least squares; they do not present any alternative methods.

L. H. C. Tippett (1925) tabulates the mean, standard deviation, β_1 and β_2 (measures of skewness and kurtosis), and values occurring with probabilities 5% and 1% for the distribution of the largest of n individuals in samples from a normal population for selected values of n up to 1000, as well as the mean, standard deviation, and β_1 and β_2 for the distribution of the range. He uses his results in deciding whether or not to reject outlying observations. Egon Sharpe Pearson (1926) extends Tippett's results.

Estienne (1926-27) proposes replacing the classical theory of errors of observation based on the arithmetic mean and the method of least squares by what he calls a rational theory based on the median and the method of least (absolute) first powers, which he derives from the notion that, for a conscientious observer, every measurement has the same subjective probability 1/2 of being too large as of being too small. Interestingly enough, he does not take the next logical step and propose replacing the Gaussian law of error by Laplace's first law of error. Instead, he insists that his so-called rational theory is consistent with any one of a whole family of error laws, including the Gaussian law, which may be used to approximate the unknown true law. He also proposes to treat the case of systematic errors, which Gauss (1823) specifically excluded from his treatment, by taking the median of the medians of repeated measurements under each of several conditions. He proposes a method of combining m linear equations in n unknowns ($2n>m>n$) based on his rational theory. Even when the restriction $m<2n$ is satisfied, his method does not appear to be practical. He insists that the importance of an error, at

least in many cases, is proportional to its absolute value and not to its square as postulated by Gauss (1823); that, however numerous be a set of measurements taken to the nearest unit, they cannot provide well founded reasons to adopt a value stated to a fraction of that unit; that the accuracy of a set of measurements cannot be judged by their consistency, since there may be systematic errors; and, finally, that time and money are better spent in perfecting measuring instruments so as to obtain more accurate observations than in increasing the number of observations.

"Student" [William Sealy Gosset] (1927) proposes a criterion for rejection of doubtful observations, based on the sample range. He also extends the work of Tippett (1925) and E. S. Pearson (1926) on the distribution of sample range.

Arthur L. Bowley (1928) gives a summary and an annotated bibliography (74 items) of Edgeworth's contributions, including important work on the law of error and the choice of means, with special emphasis on the median. In the latter connection, the author writes (pp.101-102): "The median has the disadvantage that its standard deviation of error (1 divided by $2\sqrt{n}$ times the greatest ordinate where the area of the frequency curve is unity) is greater by 25% than that of the arithmetic mean in the normal curve. This excess is not, however, serious in the rough measurements of credibility with which we are generally concerned in statistics, and in some non-normal curves the median is more accurate than the arithmetic mean. *** It has the well-known advantage that it can be computed when the measurements away from the centre are known only roughly, and generally in graded observations interpolation is only needed in the central parts ***. If no graduation is necessary the median is evidently

the easiest mean to compute, and if the maximum ordinate is known its probable error can at once be written down. As is well known, the median is that position which makes the sum of the deviations from it (all taken as positive) a minimum, the test of least detriment suggested by Laplace. ••• Finally, if a mean is required of discordant observations, where discordance signifies that the observations are taken from facility curves with different moduli, there is 'a peculiar propiety in the use of the Median.' •••. For these reasons Edgeworth attached great importance to the median in a considerable number of problems. His advocacy extends to its use for two or more unknowns •••." The last statement is exemplified by seven pages (pp. 103-109) on the method of situation of Laplace (1812) [1818] and its modification (by eliminating the restriction that the sums of positive and negative deviations be equal in magnitude) and extension (to two or more unknowns) by Edgeworth (1887b,c,1888,1923).

Jerzy Neyman and E. S. Pearson (1928) give an expression for the probability integral of the range of samples of size n and tabulate the mean, standard deviation, $\beta_1 = \alpha_3^2$ and $\beta_2 = \alpha_4$ for the standardized range W/σ for samples of size n = 3,4,6,10,20 from rectangular and normal distributions.

Edwin B. Wilson & Margaret M. Hilferty (1929) re-examine an extensive series of observations for which C. S. Peirce ["Theory of Errors of Observations," Report of the Superintendent of the U. S. Coast Survey (for the year ending November 1, 1870), Appendix No. 21, pp. 200-204 and Plate No. 27. U. S. Government Printing Office, Washington, 1873] concluded that the normal law was verified, and reach the opposite conclusion. Peirce's data consist of about 500 observations each day for 24 different days. The authors give the

standard deviations of the median and of the mean for each day, and proceed to compare them, reaching the following conclusions (p. 124): "The ordinary statement based on the normal law is that the determination of the median is 25% worse than that of the mean. A comparison of the standard deviations of the median and mean ··· shows that for these observations the median is better determined than the mean on 13 days, worse determined on 9 days, and equally well determined on 2 days. Roughly speaking, this means that mean and median are on the whole equally well determined." The results tend to show not only that the data have not come from a normal distribution but that for some distributions the median is more precise than the mean.

E. C. Rhodes (1930) gives (pp. 974-978) the following account of problems encountered in an application in a practical situation of the method of minimum deviations: "The writer recently was desirous of smoothing out the fluctuations in [a] series of figures [17 pairs of values of x=8(1)+8 and y]···. A parabola was fitted by the method of least Squares. ··· The result was not considered altogether satisfactory. ··· It was considered that the parabola was a bad fit. Two reasons suggested themselves for this. First, the original data from which the series was obtained did not involve absolutely random fluctuations; second, the parabola might not be the best curve for use in smoothing. ··· It was ··· decided to concentrate on the first consideration, which meant that although we had obtained the parabola of best fit by the method of least squares, yet it might not really be the best parabola which would smooth out the fluctuations in the series. This led us to the question of what other methods of fitting there were available, and Edgeworth's description of the use of medians in this connexion led to the attempt to fit by the method of

Minimum Deviations. This method may be briefly described. Suppose the equation to the parabola is $y=a_0+ a_1 x+a_2x^2$, and the given y's are y_{-8}, y_{-7}, $\cdots$, y_{-1}, y_0, y_1, $\cdots$, y_8, then instead of, as in the method of Least Squares, making $S_{x=-8}^{+8}(a_0+ a_1 x+a_2 x^2-y_x)^2$ a minimum, we make $S_{x=-8}^{+8}|a_0+ a_1 x+a_2 x^2-y_x|$ a minimum. His [Edgeworth's] description of the method and his arguments in its favour are briefly summarized [by] Bowley [(1928)], pp. 103 et seq., and are exposed by Edgeworth [(1888,1923)]. Unfortunately, Edgeworth confined himself in the working of the method to a reliance on a diagram (he used as illustrations the problem of two variables), which means in practice a rather laborious piece of work, and apparently did not notice that the method could be applied in a more simple manner. $\cdots$ The simpler method is as follows:-- Suppose we are dealing with a series of deviations, say, involving three unknowns, $A_1u+B_1v+C_1w+D_1$, $A_2u+B_2v+C_2w+D_2$, $\cdots$ $A_nu+B_nv+C_nw+D_n$, and we want to find values of u,v,w which make $S_{s=1}^{n}|A_su+B_sv+C_sw+D_s|$ a minimum. First, find for what values of v and w the expression is a minimum when u is given by $-(B_rv+C_rw+D_r)/A_r$, i.e. find a local minimum point in the plane $A_ru+B_rv+C_rw+D_r=0$, where r is any one of the values of s from 1 to n. This reduces the problem to one involving two variables only, i.e. what values of v and w will make $S_{t=1}^{n-1}|E_tv+F_tw+G_t|$ a minimum, where the E's, F's, G's are obtained from the A's, B's, C's, D's. To solve this, find for what value of w the expression is a minimum when v is given by $-(F_pw+G_p)/E_p$, i.e. find a local minimum point in the line $E_pv+F_pw+G_p=0$, where p is any one of the values of t from 1 to n-1. This reduces simply to the problem of finding a weighted median. $\cdots$ Then [the] point of intersection of $\cdots$ three planes [$A_lu+B_lv+C_lw+D_l=0$, $A_mu+B_mv+C_mw+D_m=0$, $A_nu+B_nv+C_nw+D_n=0$] is the true minimum point, and the values u,

v, w obtained from solving these equations make $S_{s=1}^{n}|A_s u+B_s v+C_s w+D_s|$ a minimum."

Tokishige Hojo (1931) studies the distribution of the median, quartiles, and interquartile distance in samples from a normal population. He compares the precision of the median with that of the mean and the midrange, and considers the problem of estimating the population standard deviation from the interquartile distance of a sample. In defining the sample median, M, and the sample quartiles, Q_1 and Q_3, of a sample of size n, the author distinguishes four cases according as (i)n=4m, (ii) n=4m+1, (iii) n=4m+2; (iv) n = 4m+3, where m is an integer. His definitions for these cases are as follows: (i) $Q_1=(x_m+x_{m+1})/2$, $M=(x_{2m}+x_{2m+1})/2$, $Q_3=(x_{3m}+x_{3m+1})/2$; (ii) $Q_1=(x_m+x_{m+1})/2$ $M=x_{2m+1}$, $Q_3=(x_{3m+1}+x_{3m+2})/2$; (iii) $Q_1=x_{m+1}$, $M=(x_{2m+1}+x_{2m+2})/2$, $Q_3=x_{3m+2}$; (iv) $Q_1=x_{m+1}$, $M=x_{2m+2}$, $Q_3=x_{3m+3}$, where x_i denotes the $i^{\underline{th}}$ smallest observation. The present writer prefers to define the median and the quartiles so that they divide the population into four intervals with equal probability 1/4, viz. $Q_1=x_{(n+1)/4}$, $M=x_{(n+1)/2}$, $Q_3=x_{3(n+1)/4}$, where a fractional subscript indicates interpolation between adjacent ordered observations. This definition agrees with Hojo's except for the quartiles in cases (i) and (iii) above, for which it yields, (i) $Q_1=(3x_m+x_{m+1})/4$, $Q_3=(x_{3m}+3x_{3m+1})/4$; (iii) $Q_1=(x_m+3x_{m+1})/4$, $Q_3=(3x_{3m+2}+x_{3m+3})/4$.

Richard von Mises (1931), in his volume on probability and its applications, includes a section on elementary descriptive statistics in which he discusses the median, quartiles, deciles and percentiles in addition to the arithmetic mean and the standard deviation. There is also a chapter on the theory

of errors and adjustment of observations in which the author follows the approach of Gauss (1809,1823).

Karl Pearson (1931) gives tables of criteria (Chauvenet's and Irwin's) for the rejection of outlying observations and of the distribution of range, median, and midrange in samples from a normal population. Walter A. Shewhart (1931) discusses the use of the median and the midrange as measures of central tendency instead of the arithmetic mean, and the use of the range as a measure of dispersion instead of the standard deviation.

Allen T. Craig (1932a,b) proves several useful theorems concerning the distributions of the sample median, mean, midrange, first quartile, and range. Harold Jeffreys (1932) offers an alternative to the rejection of observations. He takes the probability of an error to be given jointly by two normal [Gaussian] laws, one for the normal and the other for the abnormal errors, and provides a method of solution for the five unknowns (means and standard deviations for normal and abnormal errors and proportion of normal errors), together with an approximate solution by a method of weighting, the weight of an observation being a continuous function of its deviation.

Willem J. Luyten (1932) considers data on the differences of pairs of measures of the distance of double stars, and concludes that Laplace's first error curve fits the data much better than his second (the normal curve). He points out (p. 365) that, as a corollary of the use of the first Laplacean curve, it is no longer the arithmetic mean and the standard deviation but the median and the arithmetic mean error [mean deviation from the median] that are the significant constants of the distribution.

Egon S. Pearson (1932) gives a table summarizing available results on the

distribution of range in samples of 100 or less from a normal population. He discusses the method of computation of the table and experimental checks on the adequacy of the approximation employed, and gives illustrations of the use of the table.

P. R. Crowe (1933) proposes a graphical method, based on the median and the quartiles, of representing the distribution of monthly rainfalls. He compares the median with the mean and the mode, and the quartile deviation with the standard deviation and the mean deviation, giving advantages and disadvantages of each from the viewpoint of the climatologist.

A. T. McKay and E. S. Pearson (1933) develop theory which leads to certain new results regarding the form of the range curve at its terminals and provides the exact distribution of the range of samples of 3 from a normal population. They also give the exact distributions of the range in samples of n from rectangular and right triangular universes, the former having previously been given by Neyman and Pearson (1928).

Paul Reece Rider (1933) summarizes the history of criteria for rejection of observations from Peirce (1852) to Jeffreys (1932), and draws the following conclusion (pp. 21-22): "From the various methods cited above it is easy to see that devices for rejecting discordant observations could be invented without number. The choice of which to use, if any, is largely an individual matter. If one is willing to subscribe to the hypothesis laid down in a given criterion, he should be willing to abide by the result of applying the criterion to a set of data. ... In the final analysis it would seem that the question of the rejection or retention of a discordant observation reduces to a question of common sense. Certainly the judgment of an experienced

observer should be allowed considerable influence in reaching a decision. This judgment can undoubtedly be aided by the application of one or more tests based on the theory of probability, but any test which requires an inordinate amount of calculation seems hardly to be worthwhile, and the testimony of any criterion which is based upon a complicated hypothesis should be accepted with extreme caution".

Hans Münzner (1934) studies the precision of the $\alpha^{\pm h}$ absolute moment ($\alpha>0$) for the generalized Gaussian distribution function $\phi_\chi(\varepsilon)= [h_\chi/2\Gamma(1/\chi)]e^{-h^\chi|\varepsilon|^\chi}$ $\chi\geq 1$. He shows that the maximum precision is attained when $\alpha=\chi$. He points out that in the special case $\alpha=\chi=2$, this reduces to the statement that the standard deviation is the most precise measure of dispersion for the Gaussian distribution, as shown by Gauss himself. Another special case, not emphasized by the author, is $\alpha=\chi=1$, in which the result reduces to the statement that the mean deviation is the most precise measure of dispersion for Laplace's first distribution. We have already seen that the mean deviation is a minimum when taken about the median rather than about the arithmetic mean.

Harry S. Pollard (1934) gives the following results concerning medians: (1) an exact expression $\sigma_M = 1/(2\sqrt{2n+3})$ for the standard deviation of the median of samples of (2n+1) items (n an integer) from a rectangular population with probability density function $f(x)=1$ over a unit interval, which compares with the classical (large-sample) approximation $\sigma_M=1/[2f(0)\sqrt{s}]$ where s is the sample size and f(0) is the value of the p.d.f. at the population median; (2) upper and lower limits for the standard deviation of the median for samples from any population, and (3) a method of determining the probable error (or any percentile) of the distribution of the median for samples of size (2n+1), for which Dodd

(1922) has given the p.d.f.

Maurice Fréchet (1935) compares the precision of the mean and the median. He admits that the mean is to be preferred to the median in the majority of cases, but insists that it is not always so--median life or median income provides a more representative value than the corresponding mean. Certain statisticians object to the use of the median on the ground that its precision is less than that of the mean, which is true for the normal law of error (Laplace's second law), but not, as the author shows, for certain other probability laws. He studies two laws for which the median is more precise than the mean, at least for samples of size three. Let μ' be the standard error of the mean and μ'' be that of the median. For samples of size three from Laplace's first law of error with cumulative distribution function $F(x)= e^{x}/2$ for $x\leq 0$, $F(x)=1-e^{-x}/2$ for $x\geq 0$, the median is slightly more precise [$\mu'^{2}= \mu''^{2}+ 1/36$], while for samples of size three from the probability law $F(x)= 0$ if $x<1$, $F(x)= 1-x^{-\alpha}$ if $x\geq 1$, with $1<\alpha<2$, μ'' is finite but μ' is infinite. The author presents supporting evidence from a paper by Wilson & Hilferty (1929).

R. C. Geary (1935) considers the ratio w_n of the mean deviation to the standard deviation [$\sqrt{2/\pi}$ for infinite random samples from a normal population] as a test of normality. He takes the mean deviation from the mean rather than from the median. He notes that $\sqrt{\beta_1}$ is a test of symmetry rather than of normality, while the frequency distribution of β_2 (a measure of kurtosis) is unknown. E. S. Pearson (1935) compares β_2 and Geary's criterion. He concludes that there are strong practical grounds for choosing the latter as a test of whether the population sampled is platykurtic or leptokurtic unless the sample is very large, when β_2 may be used, but that $\sqrt{\beta_1}$ is the best criterion of

skewness, and that two tests ($\sqrt{\beta_1}$ and either β_2 or Geary's criterion) are required for departure from normality.

A. T. McKay (1935) studies the distribution, for samples from a normal distribution, of $u=X-\bar{x}$, where X is the highest observation and $\bar{x}$ is the mean. He uses the statistic u as the basis of a criterion for rejection of outliers, and compares this criterion with the criterion of Irwin (1925) and one based on the distribution of range tabulated by E. S. Pearson (1932).

William R. Thompson (1935) derives the distribution of $\tau=\delta/s$, where s is the sample standard deviation and δ is the deviation of an arbitrary observation from the sample mean, and uses this statistic, for which he tabulates critical values τ_0, as a criterion for the rejection of observations deviating from the mean by more than $s\tau_0$.

Georges Darmois (1936) discusses the relative precision of the median and the arithmetic mean of samples from various populations, and the related question of whether to use the method of least squares or the method of least absolute first powers.

E. S. Pearson and C. Chandra Sekar (1936) study the criterion for rejection of outlying observations proposed by Thompson (1935). They point out that it provides complete control over the probability of Type I error (rejecting the hypothesis that all the observations have been drawn from a single normal population, with unspecified mean and standard deviation, when that hypothesis is true). Nevertheless, they show that the criterion is quite inefficient in the presence of two or more outliers unless the sample is quite large.

Emil J. Gumbel (1937) studies the precision of the arithmetic mean and of the median, and verifies that the former is more precise for uniform,

Gaussian, and extreme-value distributions, and the latter for the symmetric double exponential (Laplace's first) distribution.

Curtis Bruen (1938) considers various methods of combining observations based on the concept of power-means, as defined by Fechner (1874). The $p^{\underline{th}}$ order power mean of a set of observations, $x_i (i=1,2,3,\cdots,n)$ is that value, x, which makes the sum, $\sum|x_i-x|^p$, a minimum. It is well known that the median is the first-order power-mean, the arithmetic mean is the second-order power-mean, and the midrange is the limiting value of the $p^{\underline{th}}$ order power-mean as $p\to\infty$. Not so well known is the fact,which the author attributes to R. M. Foster (1922), that the mode is the limiting value of the $p^{\underline{th}}$- order power-mean as $p\to 0$. The author generalizes the concept of the power-mean from the case of direct observations to that of indirect observations or of implicit functional observations, for which it leads to the method of least power-sums of the absolute values of the deviations. Corresponding to mode, median, mean, and midrange one has then the methods of least number (least sum of zero powers), least sum of first powers, least sum of squares, and least maximum (least sum of infinite powers) of the absolute deviations. Jackson (1924) has studied the existence and the uniqueness of solutions by the method of least $p^{\underline{th}}$ powers $(0<p\leq\infty)$ of sets of n simultaneous linear equations in m unknowns, when $m<n$. Bruen reviews the contributions to the theory of errors of Mayer, Boscovich, Laplace, Legendre, Gauss, Cauchy, Glaisher, Fechner, Edgeworth, Turner, Goedseels, de la Vallée Poussin, Rhodes and others. He closes with a discussion as to the choice of method, in which he points out that the choice depends on the presumed distribution of deviations, each method being best for a particular distribution--the mode in one variable or

the modal point in two or more variables for a spike distribution (single isolated value), the median or median loci for a symmetric exponential (first Laplacean) distribution, the mean or mean loci for a normal (Gaussian or second Laplacean) distribution, and the midrange or midpoint of least range for a uniform (rectangular) distribution.

E. L. Dodd (1938) reproduces some of the results of Jackson (1923) concerning the median, quartiles, and other positional means (quantiles) and enumerates some of the properties of these measures.

Jose Barral Souto (1938) shows that the mathematical expression $M_h = (\sum_{i=1}^{n} p_i a_i^h)^{1/h}$, where the a_i are non-negative real numbers and the p_i are positive weights whose sum is 1, leads for particular values of h to the following "means": $h=-\infty$ gives the smallest of the a_i, $h=-1$ gives the (weighted) harmonic mean, $h=0$ gives the (weighted) geometric mean, $h=1$ gives the (weighted) quadratic mean, and $h=\infty$ gives the largest of the a_i. If the a_i are replaced by their absolute deviations from a certain value of x, M_h is transformed into $M_h(x) = (\sum_{i=1}^{n} p_i |x-a_i|^h)^{1/h}$. Souto shows that the values of x which minimize $M_h(x)$ are the following for particular values of h: for $h \to 0$, the mode; for $h=1$, the (weighted) median; for $h=2$, the (weighted) arithmetic mean; and for $h=\infty$, the midrange. The corresponding minimum values of $M_h(x)$ are respectively the geometric mean deviation (zero), the mean (absolute) deviation, the root-mean-square deviation (standard deviation), and the semirange. These results are a generalization and extension of those given by Bruen (1938) and earlier authors.

Gumbel (1939) studies the determination of the median, quartiles, and other quantiles from small samples. He takes as the three quartiles the

$[(n+1)/4]^{\underline{th}}$, $[(n+1)/2]^{\underline{th}}$, and $[3(n+3)/4]^{\underline{th}}$ values from below, where n is the sample size. In an analogous manner, he takes as the quantile corresponding to cumulative probability λ the $[(n+1)\lambda]^{\underline{th}}$ value from below.

Jeffreys (1939) discusses the use of the median instead of the mean and the rejection of observations. He points out that the mean is the best average only when the underlying distribution is normal, in which case the standard deviation is the best measure of dispersion; in the same way, the median is associated with Laplace's first distribution and the mean (absolute) deviation. He offers the opinion that there is much to be said for the use of the median when the form of the law of distribution is unknown because it is less affected by a few abnormally large residuals than is the arithmetic mean. He criticizes Peirce's and Chauvenet's criteria for the rejection of observations, and offers a modified form of the alternative which he proposed earlier [Jeffreys (1932)], with a table of weights for its implementation.

Niels Arley (1940) generalizes the results of W. R. Thompson (1935). Let $x_1, x_2, \cdots, x_n$ be independent normal variates with mean $\varepsilon(x_i) = \sum_{j=1}^{m} a_{ij} p_j$, where the a's are known coefficients and the p's unknown parameters. Let the variance of x_i be $\sigma_i^2 = \sigma^2/P_i$, where σ^2 is unknown but the weights P_i are known. Finally let ξ_i denote the estimate of $\varepsilon(x_i)$ and S_i^2 the estimate of the variance of $x_i - \xi_i$, both obtained by the method of least squares. Arley shows that the probability density function of $r_i = (x_i - \xi_i)/S_i$ is given by $p(r) = \text{const.}\ (n-m-r^2)^{(n-m-3)/2}$ for $|r| \leq n-m$. He applies this result to obtain a criterion for the rejection of observations, which he compares with criteria proposed by various other authors.

Fréchet (1940a) compares certain measures of dispersion of the sample

median and the sample mean for large samples from unimodal distributions with finite variance. Let X be a random variable with unimodal distribution, σ_x its standard error, θ_x its mean absolute error, and E_x its quartile deviation, and let M be the median and V the arithmetic mean of a sample from this distribution. Fréchet shows that for sufficiently large n, (i)$\sigma_M \leq 1.74\sigma_x/\sqrt{n}$, (ii)$\theta_M \leq 1.60\theta_x/\sqrt{n}$, (iii)$E_M \leq 1.35\ E_x/\sqrt{n}$. For V, there is the equality $\sigma_V = \sigma_x/\sqrt{n}$ corresponding to (i), but there are no inequalities for V corresponding to (ii) and (iii). Fréchet concludes that the sample median should be more widely used except when the distribution is known to be such that the sample mean is better. Fréchet (1940b) obtains the following inequalities for the measures of dispersion of the random variable X itself, the notation being the same as in the preceding paper: $0 \leq \theta_x/\sigma_x \leq 1; 0 \leq E_x/\theta_x \leq 2,\ 0 \leq E_x/\sigma_x \leq \sqrt{2}$. If Δ_x is the upper bound of the mean probability density of X between x_1 and x_2: $\text{Prob}\{x_1 < X < x_2\}/(x_2 - x_1)$ when x_1, x_2 vary, then he shows that the following inequalities hold: $0 \leq 1/\Delta\theta_x \leq 4$; $0 \leq 1/\Delta\sigma_x \leq 2\sqrt{3}$; $0 \leq 1/\Delta E_x \leq 4$. Moreover, he shows that none of the twelve inequalities among σ_x, θ_x, E_x and Δ_x can be replaced by a sharper one.

Edwan Paulson (1940) finds the distribution of the median of a random sample of size (2n+1), where n is an integer, from a symmetric population (whose mean and median are both zero) in terms of the incomplete Beta function, which has been tabulated by Karl Pearson. He suggests that his results are especially useful in sampling from populations such as the Cauchy distribution, for which the mean is not a consistent statistic.

Robert R. Singleton (1940) points out that the method given by Rhodes (1930) for estimation of the parameters in a regression equation by minimizing the sum of absolute values of deviations is iterative and recursive, and is

presented without proof. He uses geometric methods and terminology to develop proofs for various methods and to obtain a new method which reduces the labor by eliminating the recursive feature.

Abraham Wald (1940) considers two sets of random variables x_i, y_i ($i=1,\cdots, N$; N even). Neither the true values X_i, Y_i nor the coefficients α and β of the linear relation [$Y_i = \alpha X_i + \beta$] between them is known. As an estimate of α [cf. Lambert (1765a)] he uses $a=a_2/a_1$, where $a_1=[(x_1+\cdots+x_m)-(x_{m+1}+\cdots+x_N)]/N$, $a_2=[(y_1+\cdots+y_m)-(y_{m+1}+\cdots+y_N)]/N$ and $m=N/2$. He points out that the greater $|a_1|$ the more efficient is the estimate a of α, and that $|a_1|$ is a maximum when one orders the observations so that $x_1 \leq x_2 \leq \cdots \leq x_N$.

E. B. Wilson (1940) points out that the usual formula for the standard error of the median of random samples of size n, $\sigma_M = 1/2\phi_M\sqrt{n}$, where ϕ_M is the value of the probability density function at the median, unlike the corresponding formula $\sigma/\sqrt{n}$ for the mean, is not universally valid. He explores various pathological cases for which it gives incorrect results. He then sets out to find a true expression for σ_M, restricting himself to samples of odd size $n=2k+1$ (k an integer) from a population that is symmetric about the origin, with probability density function $\phi(x)$ and cumulative distribution function $\Phi(x)$ [not the author's notation]. The probability density function of the median is then $\psi(x)=[(2k+1)!/(k!)^2][\Phi(x)]^k[1-\Phi(x)]^k\phi(x)$, its mean is zero, and its variance is given by $\sigma_M^2 = \int x^2\psi(x)dx$, where the integration extends over the whole range of the function $\phi(x)$. The author applies this result to show that the standard error of the median of samples of size n from the Cauchy distribution $\phi(x)= 1/\pi(1+x^2)$ is infinite for $n=3$, but finite for $n=5,7,\cdots$.

R. J. Brookner (1941) shows that, for a random sample of size N from a

rectangular population of length 1 around an unknown value θ [that is, $f(x)=1$ if $\theta-1/2\leq x\leq\theta+1/2, f(x)=0$ otherwise], the variance of the sample midrange t is $1/[2(N+1)(N+2)]$. This compares with a variance of $1/12N$ for the sample mean $\bar{X}$. Both the mean and the midrange are unbiased estimators of θ. The ratio of the variance of t to that of $\bar{X}$ is $6N/[(N+1)(N+2)]$, which is less than one for $N>2$ and approaches zero as $N\to\infty$, so the mean is a poor estimator of central tendency for a rectangular population.

Maurice Fréchet (1941) examines two methods of demonstrating the validity of the normal law of error--the method of Gauss (1809) based on the postulate that the arithmetic mean is the best average of a set of equally reliable observations and the method [due to Hagen (1837), whom the author does not mention] based on the composition of a large number of small elementary errors. He does not find either method convincing, and concludes that verfication is possible only experimentally, by comparison with actual data. This can be accomplished in various ways--by comparing theoretical and observed frequencies in the various classes of a grouped frequency distribution; by computing the Pearsonian measure of kurtosis $\beta_2=\alpha_4=\mu_4/\mu_2^2$, where μ_r is the r^{th} moment about the mean, for the data and comparing it with the theoretical value [3 for the normal (second Laplacean) law, 6 for the first Laplacean law]; or by computing the ratios $D=(q_2-q_1)/(d_2-d_1)$ and $C=(q_2-q_1)/(c_2-c_1)$, where q_1 and q_2 are the first and last quartiles, d_1 and d_2 are the first and last deciles, and c_1 and c_2 are the first and last centiles, and comparing them with the theoretical values [$D=0.5263$ and $C=0.2899$ for Laplace's second law, $D=0.4307$ and $C=0.1772$ for Laplace's first law]. The author applies the latter method to data on artillery fire. Returning to theory,

he shows the correspondence between laws of error and measures of central tendency and dispersion, the arithmetic mean and the standard deviation corresponding to Laplace's second law and the median and the mean absolute deviation(from the median) to Laplace's first law, etc. He extends this correspondence to the multivariate case (two or more dimensions).

H. O. Hartley (1942) writes the probability integral of the range W in random samples of size n from a population with probability density function f(x) in the form $P_n(W) = n\int_{-\infty}^{\infty} f(\xi)\left(\int_{\xi}^{\xi+W} f(x)dx\right)^{n-1} d\xi$. Hartley and E. S. Pearson (1942), using numerical integration, tabulate $P_n(W)$ to 4 decimal places for samples of size n=2(1)20 from a standard normal population at intervals of 0.05 in W.

K. Raghavan Nair and M. P. Shrivastava (1942) propose a simple method of curve fitting by grouping the residuals into as many groups as there are unknown coefficients to be estimated, and using the group averages to estimate the coefficients.

George A. Barnard (1943) studies the use of the median in place of the mean in quality control charts.

Samuel S. Wilks (1943) gives an excellent textbook treatment of order statistics and functions of order statistics, including the largest or smallest sample value, the sample median, and the sample range.

R. C. Geary (1944) compares the mean, the midrange, and the median as measures of central tendency for a rectangular population with known range 1.

E. J. Gumbel (1944) studies the moment characteristics of the distribution, in a sample of size n, of the $m^{\underline{th}}$ largest and $m^{\underline{th}}$ smallest values and of their sum [+2] and difference, the $m^{\underline{th}}$ midrange and $m^{\underline{th}}$ range. He assumes that n is

so large that the two m^{th} values may be regarded as independent. For m=1, the m^{th} midrange and m^{th} range [or quasi-midrange and quasi-range, as they are more commonly called] become simply the midrange and the range.

Herbert Robbins (1944) determines the expected values of the difference between the largest and smallest order statistics [the range R] and of the difference F between the values of the cumulative distribution function evaluated for the largest and smallest order statistics. The results are $E(F)=(n-1)/(n+1)$ and $E(R)=\int_{-\infty}^{\infty}\{1-f^n(t)-[1-f(t)]^n\}dt$, where f is the probability density function [not the author's notation]. From the latter it follows that the expected value of the range for n=3 is always 3/2 that for n=2, since $1-f^3-(1-f)^3 = (3/2)[1-f^2-(1-f)^2]$.

E. S. Pearson, H. J. Godwin and H. O. Hartley (1945) study and tabulate the probability integral of the mean deviation (from the arithmetic mean) of samples from a normal distribution.

5. THE MODERN ERA (1946-1972)

George A. Baker (1946) studies the distribution of the ratios of sample range to sample standard deviation in samples from normal distributions and from two different combinations of two normal distributions, one symmetrical but distinctly bimodal and the other weakly bimodal but strongly skewed. He tabulates various moment constants of the distribution for various sample sizes. He finds that the correlation between standard deviation and range of the same sample is negligible for samples of size $n \geq 100$ from the normal population, but not from the combinations.

George W. Brown and John W. Tukey (1946) study the distribution of sample

means for samples from various distributions, including "long tailed" ones, for which they find that the distance between any two percentage points of the mean of a sample of size n is ultimately larger than a positive power of n. They claim that these results show that (1) "the use of the mean of a sample as a measure of location ··· implies a belief that the tails of the underlying distribution are not too long; (2) it is probable that the relative efficiencies of mean and median are greatly affected by the length of the tail".

A. George Carlton (1946) shows that the range and midrange of a sample from a rectangular distribution are a pair of sufficient statistics, and maximum likelihood estimates, for the true range and true mean. He derives exact and limiting distributions of midrange, range, and their ratio, and calculates the 'efficiencies' of the sample mean and median as estimates of the true mean. The limiting distributions are non-normal, with standard error of order n^{-1} instead of the usual $n^{-1/2}$. For the one-parameter rectangular distribution $f(x)=1/\lambda$, $0<x<\lambda$, he finds that the largest observation v "is a sufficient statistic and is evidently the maximum likelihood estimate of λ".

Harold Cramér (1946) gives an excellent advanced treatment of the mathematical theory of statistics, including measures of central tendency (location) and of dispersion and the method of least squares and rival methods. Since all measures of location and dispersion are to a large extent arbitrary, each measure having its own advantages and disadvantages in various cases, and since the principle of least squares is associated with specific measures (mean and standard deviation), Cramér states that there is no logical necessity for adopting this principle. On the contrary, he says, it is largely a matter of convention whether we choose to do so or not, the main reason in favor of the

principle being the relative simplicity of the rules of operation to which it leads.

Joseph F. Daly (1946) proves that, for samples from a normal population, the mean and the range (or any other symmetric function of the sample variates which is invariant under a translation of the origin) are statistically independent.

E. J. Gumbel (1946) shows that in a sample of size n (large) the $m^{\underline{th}}$ observation from one extreme and the $k^{\underline{th}}$ from the other in order of magnitude may be regarded as independent provided that m and k are small with respect to n and that the population behaves in its tails in a certain exponential manner.

Maurice George Kendall (1946) gives a thorough treatment of the theory of linear and curvilinear regression. He points out that the most important use of least squares in statistical theory is in estimating the parameters (coefficients) in regression equations. He also mentions its use in estimating the parameters of statistical distributions, which will not be considered in detail in this report.

Frederick Mosteller (1946) suggests that certain "inefficient" statistics may be useful when data are inexpensive compared with the cost of computing "efficient" statistics. In particular, he proposes the use of linear combinations of order statistics, which he calls systematic statistics, to estimate the mean and standard deviation of a normal population. He compares the efficiencies of the estimates of standard deviation with those of other estimates which do not involve sums of squares or products, including the mean deviations about the mean and about the median.

Frank Ephraim Grubbs and Chalmers L. Weaver (1947) study the use of group ranges to estimate the population standard deviation from a sample from a normal population. They tabulate the moment constants (mean, standard deviation, α_3 and α_4) of the range for samples of size n=2(1)12 from a normal population.

E. Lord (1947) proves that the mean and the difference between the $p^{\underline{th}}$ and $q^{\underline{th}}$ order statistics of a sample of size n (which reduces to the range when p=1, q=n) from a normal population are independent.

K. R. Nair (1947) shows that the standard error of the mean deviation m' from the median is equal to or less than that of the mean deviation m from the mean for samples of 3 or 4 from a normal population. He suggests that, in view of greater simplicity in calculation, there would be strong practical grounds for using m' rather than m if expressions for the mean and variance of m' and tables of its probability integral were worked out and if the efficiency of m' relative to m for sample size n>4 were found to be not appreciably worse than for n=4.

R. L. Plackett (1947) determines an upper limit, independent of the form of the distribution, for the ratio d_n of the expected range in samples of size n to the population standard deviation. This limit is $n\{2[(2n-2)!-(n-1)!)^2]/(2n-1)!\}^{1/2}$, which is approximately $n^{1/2}$ for large n. Plackett finds distributions for which the limit is attained; for n=2,3 the distributions are rectangular.

Warren B. Purcell (1947) proposes saving time in life tests by using the median instead of the mean to indicate shifts in central tendency and the minimum value (first order statistic) instead of the range to indicate shifts in dispersion, thus making it possible to terminate the test as soon as [n/2]

+ 1 failures have occurred, where n is the number of items placed on test, and [n/2] is the largest integer less than or equal to n/2.

Seaman J. Tanenhaus (1947) proposes the use of the lot median or, better still, the average median of several sublots, as the most typical value of abrasion-resistance of yarns from distributions which are decidedly positively skewed, for which the mean tends to be atypical, being unduly affected by the extremes.

Churchill Eisenhart, Lola S. Deming and Celia S. Martin (1948a) show that the abscissa of the (one-tail) ϵ-probability point of the distribution of the median in random samples of size $n=2m+1$ from any continuous distribution is identical with that of the $P_{\epsilon,n}$- probability point of the parent distribution, where $\sum_{k=(n+1)/2}^{n} C_k^n P_{\epsilon,n}^k (1-P_{\epsilon,n})^{n-k} = \epsilon$ and $C_k^n = n!/k(n-k)!$ is the number of combinations of n things taken k at a time. Eisenhart, Deming and Martin (1948b) compare the ϵ-probability points, for various values of ϵ and n, of the median with those of the mean for samples from normal (Gaussian), Cauchy and double-exponential (Laplace's first) distributions and with those of the midrange for the rectangular (uniform) distribution. Their results give numerical verification of the fact that the mean is the best average for the normal distribution, the median for the double-exponential distribution, and the midrange for the rectangular distribution, while the median is the best of the three considered for the Cauchy distribution.

G. W. Housner and J. F. Brennan (1948) consider the problem of bivariate regression in which both variables are subject to error and have a finite number of means falling on a line and in which the number of sample observations taken about each mean is known. They estimate the slope b of the

regression line $Y=a+bX$ as the total of the differences of all pairs of observed values of the y's divided by the like total for the observed x's, and show that this estimate is consistent. For the case of ungrouped data, the proposed estimate reduces to $\hat{b} = \sum_{i=1}^{n} y_i(i-\bar{i})/\sum_{i=1}^{n} x_i(i-\bar{i})$, where the x's are ordered according to magnitude. In a particular numerical example, the authors show that this estimate compares favorably with others that have been proposed.

K. R. Nair (1948) studies the distribution of the extreme deviate from the sample mean, $w=x_k-\bar{x}$, where $x_1, x_2, \cdots, x_k$ are ordered values in a sample of size k from the unit normal distribution and $\bar{x}$ is their mean, as well as the distribution of its studentized form, w/s, where s^2 is an independent unbiased estimator of the population variance. He uses the latter distribution as the basis of a new criterion for rejection of outliers, which he compares with the criteria of Irwin, Tippett, Student, McKay and Thompson.

K. C. Sreedharan Pillai (1948) determines the information (as defined by Fisher) furnished by each order statistic x_i $(i=1,2,\cdots,n)$ in a sample of size n from a normal distribution concerning the mean μ and the variance σ^2, and tabulates results for $n=2,3,\cdots, 12$; $i=1,2,\cdots, [n/2]+1$. Not surprisingly, these tables show that the central values give the most information concerning μ and the extreme values concerning σ^2. The author determines a function of n which, when multiplied by the semirange $(x_n - x_i)/2$, yields an unbiased estimator of σ, and studies the distribution of the semirange.

K. R. Nair (1949), in a follow-up of his previous note [Nair (1947)] on the mean deviations from the median and from the mean, and their use in estimating the standard deviation σ of a normal population, shows that the coefficients of variation of the two mean deviations are almost the same for samples of size n when $2 \leq n \leq 10$.

W. R. Purcell (1949) elaborates on the use of the median life and the shortest life instead of the mean life and the range, as proposed in his earlier paper [Purcell (1947)], and gives an example of their successful use in saving time in life tests on incandescent lamps.

K. J. Shone (1949) studies the use of the sample range in estimating the standard deviation of nonnormal populations. Let σ, $\bar{r}$, σ_r and N represent the sample standard deviation, the mean range, the standard deviation of the range, and the sample size, respectively. For N=2, he finds that $2\sigma^2=\bar{r}^2+\sigma_r^2$ for all populations whose variance is finite; for N=3, $\bar{r}/\sigma \doteq 2.10-0.81\ \sigma_r/\bar{r}$ for eighteen discrete unimodal distributions; for N=4 and N=5, respectively, $\bar{r}/\sigma \doteq 2.29-0.69\sigma_r/\bar{r}$ and $\bar{r}/\sigma \doteq 2.41-0.46\ \sigma_r/\bar{r}$ for five selected populations of extreme form.

John Wilder Tukey (1949a,b,c,d) and Theodore E. Harris and Tukey (1949) report on a study of sampling from contaminated distributions. Tukey (1949a, c,d) studies the relative efficiencies and effectivenesses (in large samples) of various estimation procedures when the distribution differs from normality in the direction of long tails (resulting from a mixture of two normal distributions with the same mean and different standard deviations). Harris and Tukey (1949) and Tukey (1949b) consider the relative efficiencies of estimators obtained from the mean and standard deviation by removing the extreme γ % at each end of the sample, for varying degrees of contamination, both in large and in moderately large samples.

William John Youden (1949) exposes the fallacy in the common practice of making three measurements, averaging the two values closest together and discarding the other. Intuition suggests that if two of the three measurements

are in close agreement while the third is considerably removed from either of the others, then there may be grounds for suspecting and perhaps rejecting the third value. Analysis shows, however, that for samples of three from a normal distribution, one of the measurements will be at least 19 times farther away from its neighbor than the distance separating the two closest in one sample out of twelve; hence it appears that measurements that should be retained are often discarded.

Wilfrid J. Dixon (1950) proposes new criteria, based on the ratio of the differences of two pairs of order statistics, for rejection of outlying observations. He compares the performance of these criteria with those of Irwin, McKay, Thompson, Nair, and Grubbs for detecting contamination of samples from a normal population with mean μ and variance σ^2, $N(\mu,\sigma^2)$, by one or more observations from (a)$N(\mu + \lambda\sigma,\sigma^2)$ or (b)$N(\mu,\lambda^2\sigma^2)$.

Grubbs (1950) also proposes a new criterion for rejection of outliers, the criterion being the ratio of the sums of squares of deviations from the mean for the truncated sample (with the observation or observations in question omitted) and for the complete sample. He obtains and tabulates the distribution of this ratio for one extreme observation and for two extreme observations both at the same end; he does not examine the criterion for one extreme at each end.

Theodore E. Harris (1950) gives a simple explanation, with a numerical example, of a procedure, essentially that of Edgeworth (1923) and Rhodes (1930),

for fitting a regression line Y=a+bX by minimizing the sum of the absolute deviations rather than the sum of squares of the deviations. He points out the relation between this problem and linear programming.

F. M. Henry (1950) studies the loss of precision from discording discrepant data. In particular, he applies the rule of Goodwin (1913): "When the number of observations is small, reject any observation that deviates more than 4 A.D. from the sample mean, the mean and A.D. [average deviation] being computed with the omission of the doubtful observation" to series of five measurements of a time interval with a stop watch, and the "best two out of three" procedure to series of three such measurements. In both cases he finds that use of the procedure results in an increased rather than a decreased error; in the case of three measurements, not only the average of all three measurements, but also the average of the two extreme measurements (the midrange) gives better results.

E. S Pearson (1950) investigates the estimation of the standard deviation of a population from the range of a sample of size n or the mean range of N items divided into m groups of n items each. Even when (a) the population is not normal or (b) the sample includes one or more outliers, he concludes that use of the range, with adjustment appropriate for a normal population, is justified provided $n \leq 10$.

K.C.S. Pillai (1950) finds, in a form suitable for numerical calculations, the distributions of the midrange and the semirange and their joint distribution for samples of size n from a standard normal population, N(0,1).

G. R. Seth (1950) finds the joint distribution of the two closest observations $x';x''(x'<x'')$ of the set $x_1,x_2,x_3(x_1\leq x_2\leq x_3)$, given the distribution of x_1,x_2,x_3; he also finds the joint distribution of $u=(x''-x')$ and $w=(x''-x')/(x_3-x_1)$ in general, and the joint density function of u and w and the marginal density functions of u and of w, all when the underlying distribution is normal with mean θ and variance unity, $N(\theta,1)$. He also obtains the joint density function of $u=x''-x'$ and $v=(x'+x'')/2$, as well as the marginal density function of v, which has mean θ and variance $1/2+\sqrt{3}/4\pi$.

R. K. Zeigler (1950) shows that, for a random sample of size 2k+1 from a distribution which has a finite second moment and which is continuous at $x=\theta$ with $f(\theta)\neq 0$, θ being the population median, the joint distribution of the sample median and the mean deviation from the sample median is asymptotically bivariate normal, and gives the asymptotic means, variances, and correlation coefficient.

D. H. Rhate (1951) shows that, for symmetrical probability functions which are members of the Pearson family, the mean of two symmetrically placed elements in an ordered sample (a quasi-median or quasi-midrange) is more efficient than the median as an estimate of the central value. He demonstrates by an example that this statement is not true for all symmetrical probability functions.

Brown and Mood (1951) propose a method based on medians for determining the coefficients in a multiple linear regression equation. Let the dependent variable y be distributed with median $\alpha_0+\sum_1^k \alpha_r z_r$ and suppose we have a sample of n sets of associated observations $y_i, z_{1i}, z_{2i}, \cdots, z_{ki}$ with $i=1,2,\cdots,n$. Then the coefficients α_r are estimated by the numbers $\hat{\alpha}_r$ such that median $z_{ri}<\hat{z}_r$

$(y_i-\tilde{\alpha}_0-\sum\tilde{\alpha}_r z_{ri})=$ median $_{z_{ri}>\tilde{z}_r}(y_i-\tilde{\alpha}_0-\sum\tilde{\alpha}_r z_{ri})=0, r=1,2,\cdots,k$, where $\tilde{z}_r$ is the median of the n observations z_{ri}.

Dixon (1951) finds the distribution of the ratio $r=(x_n-x_{n-j})/(x_n-x_i)$ for some small values of i and j, where $x_1,x_2,\cdots,x_n$ are the order statistics of a sample of size $n\leq 30$ from a population which is (1) rectangular or (2) normal. He tabulates 3-decimal-place percentage points, corresponding to cumulative probability $\alpha=.005,.01,.02,.05,.1(.1).9,.95$, for r when j=1,2 and i=1,2,3, for samples of size n=(i+j+1)(1)30 from a normal population. These tabular values are useful in applying the criteria for rejection of outliers proposed by the author in his earlier paper [Dixon (1950)].

H. O. Hartley and E. S. Pearson (1951) tabulate the moment constants of the distribution of the range in samples of size n=2(1)20 drawn from a normal population with unit variance. They note that there are some discrepancies between their table and some earlier results of Grubbs & Weaver (1947).

Ray Bradford Murphy (1951) treats the problem of outlying observations in samples from univariate normal populations as one in linear hypotheses. In particular, he introduces t-tests for outliers from a single universe and likelihood ratio tests for outliers from several universes. He discusses the problems of testing for all possible numbers of outliers, k, subject only to the restriction that $2k<n$, where n is the sample size.

J. H. Cadwell (1951) finds improved approximate formulas (polynomials in n^{-1}) for the ratio of the standard error of the median to the standard error of the mean for random samples of size n drawn from a normal population. Numerical examples show good agreement with the exact results of Hojo (1931). The author shows how to extend the result so as to obtain the corresponding ratio

for any quantile of a sample from a continuous population.

Anders Hald (1952a,b) gives theory and tables for the distribution of the range. He also discusses the distributions of the largest observation and of its deviations from the population mean and the sample mean, as well as criteria for the rejection of outlying observations.

Tsurochiyo Homma (1952) obtains possible limit laws for the range and the midrange of samples from a continuous population, and shows that the range and the midrange are not asymptotically independent.

Norman L. Johnson (1952) gives an approximation, valid for w small and n not too large, for the probability $P_n(w)$ that the range of n independent random variables does not exceed w. He also gives an approximation for the critical values w_α satisfying $P_n(w_\alpha)=\alpha$.

Julius Lieblein (1952) investigates the distributions of several statistics involving the closest pair of observations in a sample of size three from rectangular and normal populations, and calculates their means and standard deviations. He shows that the ratio of the difference of the closest pair of observations to the range is a poor criterion for rejecting outlying observations, and finds the distribution of the outlying observation for a rectangular population.

K. R. Nair (1952) extends his earlier table [Nair (1948)] of percentage points of the studentized extreme deviate from the sample mean to cover more sample sizes and more significance levels.

Calyampudi Radhakrishna Rao (1952) gives examples involving the use of the largest and/or smallest observations to estimate the parameter(s) of a rectangular population, the asymptotic distribution of quantiles, and the

efficiency of the sample median as an estimate of the mean of a normal population.

J. H. Cadwell (1953) presents a method for evaluating the probability density function of the $r^{\underline{th}}$ quasi-range, $w_r = x_{n-r} - x_{r+1}$ of a sample of size n from a normal population, and tabulates percentage points and moments of w_1 for n=10(1)30. He investigates the efficiency of quasi-ranges in estimating the population standard deviation, and finds w_0(the range) to be most efficient for $2 \leq n \leq 17$ and w_1 most efficient for $18 \leq n \leq 31$.

Dixon (1953) studies the problem of contamination of a sample supposed to be drawn from a normal population with mean μ and variance σ^2, $N(\mu,\sigma^2)$, by drawing a proportion γ of the observations from either $N(\mu+\lambda\sigma,\sigma^2)$ or $N(\mu,\lambda^2\sigma^2)$. He discusses the estimation of μ by use of the mean and the median, the estimation of σ^2(or σ) by the sample variance and the range, and gives recommended rules for processing data under various conditions of contamination.

Enoch B. Farrell (1953) proposes the construction of quality control charts using ranges and midranges within subgroups and medians of these statistics between subgroups. He contends that this method gives more useful estimates of the true population parameters than the conventional method when outlying observations due to the presence of assignable causes of variation are present, also that it is more effective in detecting and locating assignable causes, besides involving simpler computations.

Harman Leon Harter (1953) applies the principle of maximum likelihood to the problem of determining the regression equation of one variable on p others. He shows that for a normal distribution of residuals, the maximum likelihood solution is the least squares solution, found by minimizing the

sums of the squares of the residuals, while for a Laplace (first) distribution of residuals, the maximum likelihood solution is found by minimizing the sum of the absolute values of the residuals. For distributions of residuals with finite limits, only certain solutions are admissible, and either of the above methods may lead to an inadmissible solution. For a rectangular distribution of residuals, the likelihood function is a constant, and there is no unique maximum likelihood solution, one admissible solution being just as likely as another.

E. P. King (1953) shows that, when the criteria of Grubbs (1950) and Dixon (1951) are employed to detect the presence of a single outlier, the effect of using a test statistic based on the more deviant of the two extremes, thus testing a two-sided hypothesis, is approximately, but not exactly, to double the significance level of the standard test procedure.

Edwin Glenn Olds (1953) studies the problem of finding the coefficients a and b in the equation of the best-fitting straight line Y=a+bX when values of y are observed corresponding to a fixed set of x-values. He points out that when y has a normal distribution with constant variance for each x, the solution can be found either by the method of least squares or by the method of maximum likelihood. When y has a rectangular distribution, the method of maximum likelihood does not, in general, give a unique solution, and the method of least squares sometimes yields a solution which is inconsistent in that the residual (the difference between the predicted and observed y-values) for one or more x-values may lie outside the admissible interval (-c,+c), where 2c is the range of the rectangular distribution assumed for the residuals. The author adopts the least squares solution whenever it is consistent; when it is

not, he shows how to find a modified least squares solution which minimizes the sum of squares of the residuals subject to the restriction that the absolute value of each residual must be less than or equal to c.

Frank Proschan (1953) advocates, for the rejection of outlying observations, Dixon's criterion based on the extreme observation when no past data are available and Nair's criterion based on the studentized extreme deviate when past data are available for use in obtaining an independent estimate of the standard deviation of an individual measurement. He tabulates critical values at the 5% and 1% levels for both tests.

Youden (1953) summarizes available results on the situation in which three measurements are made and the two showing the best agreement are selected. The difference between the selected measurements averages about four-tenths $(3-3\sqrt{3}/2)$ that of the difference for honest duplicates [Lieblein (1952)]. The dispersion of the average of the selected pair is 12 per cent larger than that of the average of duplicates. Let d be the difference between the selected pair and D the difference between the discarded measurement and the nearer of the selected ones. The interval D is ten or more times as large as d in 15.7 percent of sets of three measurements. More than one-third of the time, D is at least four times as large as d. Values of the ratio D/d exceed 32.57 once in twenty times [Youden (1949)].

Cadwell (1954) gives an asymptotic expression for the probability integral of the range for samples from a symmetric unimodal distribution, and investigates its accuracy for the case of samples of size 20 to 100 from a normal population. For this range of sample sizes the errors are small, and they can be made less than 0.0001 by using a correction based on values given in the paper. The author

tabulates percentage points of the range for samples of size n=20,40,60,80 and 100 from a normal population.

David R. Cox (1954) studies the mean range and the coefficient of variation of the range in samples of size 2,3,4, and 5 from different types of populations covering a wide range of values of $\beta_1=\alpha_3^2$ and $\beta_2=\alpha_4$, the measures of skewness and kurtosis, including symmetric and asymmetric mixtures of normal distributions, the normal distribution,the rectangular distribution, exponential type distributions, the Pearson system, and Shone's numerical results for five discrete distributions. He tabulates the normalized mean range and the coefficient of variation of the range to 3 decimal places for samples of size 2, 3,4 and 5 for β_2=1.0(.2)2.0(.5)5.0(1.0)9.0; β_1 is not a determining factor.He compares the distributions of the range for samples from exponential and normal populations, and applies the results to estimation of dispersion by use of the range.

Herbert A. David (1954) finds the cumulative distribution function and the expected value of the range of samples from five non-normal populations, and makes numerical comparisons of the results with the corresponding ones for samples from a normal population.

H. A. David, H. O. Hartley and E. S. Pearson (1954) approximate the distribution of u=w/s [where $w=x_{max}-x_{min}$, $(n-1)s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$, and $x_1,\cdots,x_n$ is a random sample with mean $\bar{x}$ from a normal population] by selecting a curve from the Pearson system with the proper first four moments. For specific values of n they compare the result with that of an exact alternative derivation. After examining certain non-normal populations, they suggest that u may be useful in detecting departures from normality.

Gumbel (1954) derives the continuous cumulative distribution function with specified mean and variance for which the expected value of the largest of n independent observations is a maximum and the continuous c.d.f. with specified variance for which the mean range is a maximum. The latter result, obtained by a different method, was previously given by R. L. Plackett (1947). The former result, obtained independently, is given by Hartley and David (1954), who also obtain an upper bound for the expected value of the m^{th} order statistic and best upper and lower bounds of the sample range of x under the restrictions that the mean and variance of x are 0 and 1 respectively and values of x are restricted to the closed interval [a,b], where a and b are given constants. They derive the distributions for which the upper bounds are attained, and show that the lower bound is attained for a discrete distribution where x may assume only two values. These results are of interest in assessing the bias that may result from the unwarranted assumption of normality when using the sample range to estimate the population standard deviation.

E. S. Pearson and H. O. Hartley (1954) give tables of moment constants, probability integral and percentage points of the range; also tables of percentage points of the extreme standarized deviate from the population mean and from the sample mean, the extreme studentized deviate from the sample mean, and the ratio of range to standard deviation in the same sample, all for samples from a normal population. Various applications of these tables, including the rejection of outlying observations, are discussed in the introduction.

George J. Resnikoff (1954) discusses various approximations to the distribution of the average range, and tabulates percentage points of the average range for subgroups of size five, commonly used in quality control work, for

samples of size N=5m, where m is an integer.

F. Zitek (1954) discusses various measures of sample dispersion, including standard deviation, mean deviation, and range, which may be used in estimating the standard deviation of a normal population. For n=2(1)15, he tabulates normalizing factors which make these estimators unbiased and variances of the resulting unbiased estimators whenever these are available in the literature, and makes some observations on the efficiency of the estimators.

John T. Chu (1955a) obtains upper and lower bounds for the cumulative distribution function of the median $\tilde{x}$ of a sample of (2n+1) observations on a random variable X from a population with probability density function f(x) and unique median ξ. He shows that the approach to normality of the distribution of $\tilde{x}$ is rapid when X is normally distributed, but much slower when X has a rectangular or a Laplace (first) distribution. Chu (1955b) shows that, under very general conditions, $\text{var}\,\tilde{x} \geq \{4[f(\xi)]^2(2n+3)\}^{-1}$, as compared with the asymptotic variance $\{4[f(\xi)]^2(2n+1)\}^{-1}$, with the equality holding for the rectangular distribution. He shows that the sample mean $\bar{x}$ is more efficient (has smaller variance) than $\tilde{x}$ for many symmetric distributions, notable exceptions being the Laplace and Cauchy distributions.

Chu and Harold Hotelling (1955) show that, under certain regularity conditions, the central moments of the sample median are asymptotically equal to the corresponding moments of the asymptotic distribution, which is normal. They give a general approximation procedure for the moments of the median which involves expanding the inverse of the cumulative distribution function in a Taylor series; the approximation error can be made arbitrarily small by using a sufficiently large number of terms in the expansion. They apply the method

to the normal, Laplace, and Cauchy distributions; for the first two of these they obtain upper and lower bounds for the variance of the median by a much simpler procedure. They obtain detailed results concerning the medians of samples drawn from a normal population.

J. Arthur Greenwood (1955) expresses the differential of the probability of the n-th range [quasi-range] in terms of Bessel functions of the third kind, and integrates by parts to obtain the distribution of the n-th range.

Max Halperin, Samuel W. Greenhouse, Jerome Cornfield and Julia Zalokar (1955) tabulate, to three significant figures, the upper and lower 5% and 1% points of the studentized maximum absolute deviate $d=\max_i |x_i - \bar{x}|/s$, where the $x_i (i=1,\cdots,k)$ are independent and each $N(\mu,\sigma^2)$, and where ms^2/σ^2 is distributed as χ^2 with m degrees of freedom and independent of x_i, for k=3(1)10(5)20(10)40, 60 and m=3(1)10(5)20(10)40,60,120. They give examples to illustrate the use of the tables for various purposes, including an outlier test which is the two-sided version of Nair's test.

George William Thompson (1955) shows that bounds exist for w/s, the ratio of range to standard deviation in the same sample of size n, for all populations with non-zero variance. He tabulates upper and lower bounds for w/s to three decimal places for n=3(1)20(10)60(20)100(50)200,500,1000; also lower and upper 0.1%, 0.5%, 1.0%, 2.5%, 5.0% and 10.0% points and the median (50% point) of w/s to five decimal places for samples of size n=3 from a normal population.

Tukey (1955) shows that various characteristics (e.g. percentage points, expectation, reciprocal standard deviation) of the range of samples from a normal population behave asymptotically like the square root of a log (bn+c)

where n is the sample size and a,b,c are appropriate constants. He uses this fact as an aid in interpolating between tabular values of these characteristics.

W. U. Behrens (1956) proposes certain factors for use in determining the standard deviation approximately either from the mean deviation (taken from the mean) or from the range. He compares these factors with those developed by other authors [including Tippett (1925), Pearson (1932) and Pearson, Godwin & Hartley (1945)], and refers to the different bases of the two kinds of factors. He contends that his factors are useful for the objectives generally pursued by experiment stations.

Juan Bejar (1956) defines the median regression curve of a bivariate distribution $f(x,y)$ as the locus $y=g(x)$ of the median of the conditional distribution $f(y|x)$, and gives its general properties. Since $g(x)$ is not easy to obtain, the author introduces the linear regression, and then the polynomial regression, which minimize the mean deviation instead of the mean square deviation as in the mean regression curve. He points out that the calculation of the median regression involves minimizing a linear expression with the variables constrained by inequalities, and is therefore closely related to linear programming.

Chester I. Bliss, William G. Cochran and John W. Tukey (1956) propose a new criterion, based on the range, for the rejection of outliers. The test statistic is the largest range in k sets of n measurements divided by the sum of all the ranges. If the observed ratio exceeds the 5% critical value, which they tabulate for various values of k and n, they conclude that the set having the largest range contains an outlier, which is identified by inspection and rejected.

H. A. David (1956) gives tables of the upper percentage points of the studentized extreme deviate from the sample mean like those of Nair (1948, 1952) [reprinted by Pearson & Hartley (1954)], but corrected by using a better approximation.

Akio Kudô (1956) proposes a new criterion for the rejection of outlying observations. Given three sets of independent observations {x}: (i) n_1 from $N(m_i,\sigma^2)$, $i=1,2,\cdots,n_1$; (ii) n_2 from $N(m^{(2)},\sigma^2)$; and (iii) n_3 from $N(m^{(3)},\sigma^2)$. One of n_1+1 possible decisions, D_i (accept H_i), is to be made, where H_0: $m_1=m_2=\cdots=m_{n_1}=m^{(2)}$, H_i ($i\neq 0$): $M^{(2)}=$ each m_j (except m_i) $=m_i-\Delta$. The author presents a decision procedure for which: $\Pr(\text{acc. } D_0|H_0)=1-p$, $\Pr(\text{acc. } D_i|H_i)$ is maximized for $i\neq 0$. The optimum decision procedure involves $x_M=\max\{x_1,x_2,\cdots,x_i,\cdots,x_{n_1}\}$; $\bar{x}$, the mean of samples (i) and (ii); S, the overall standard deviation using $\bar{x}$ for (i) and (ii) and $\bar{x}_3$ for (iii). The decision rule is: select D_0 if $(x_M-\bar{x})/S\leq\lambda_p$, and select D_M if $(x_M-\bar{x})/S>\lambda_p$. If σ is known, S is replaced by σ; in this case set (iii) is not needed, and different λ_p are needed. In each case, the author states that the critical values λ_p are to be published later [see Kudô (1958)].

Harold Ruben (1956) shows that the product moments of the extreme order statistics in samples of even sizes from normal populations can be expressed as linear functions of the products of the contents of certain hyperspherical simplices, and uses this fact to obtain simple explicit expressions for the variance of the sample range for samples of size 2 and 4.

Juan Bejar (1957) gives a method, similar to linear programming, to determine the regression line $y=a+bx$ such that $\sum|y_i-a-bx_i|$ is a minimum or the regression plane $z=a+bx+cy$ such that $\sum|z_i-a-bx_i-cy_i|$ is a minimum. He gives two

examples in which he arranges the data to make the shortest calculations.

Dixon (1957) discusses several simple estimates of the mean and standard deviation of a normal population. Estimates of the mean considered are the median, the midrange, the mean of the best two (in the sense of minimum variance), and the mean of all but the largest and smallest. Estimates of the standard deviation studied are various linear combinations of quasi-ranges. The efficiencies of these estimates are compared with those of the sample mean and sample standard deviation and the best linear unbiased estimates for samples of size n=2(1)20.

A. Ghosal (1957) derives formulas for the distribution of the $r^{\underline{th}}$ quasi-range $W_r = x_{n-r} - x_{r+1}$ of samples of size n from rectangular and exponential distributions; tabulates their first four moment constants for r=0,1,2 and n=5,10,15,20; and compares the efficiencies of $W_r (r>0)$ and W_0 as estimators of the population standard deviation. For the exponential distribution, he finds that W_1 is more efficient than W_0 for $n \geq 9$ and W_2 is more efficient than W_1 for $n \geq 17$.

B. T. Harley and E. S. Pearson tabulate the probability integral and percentage points of the range for samples of size n=200 from a normal population. They indicate that the results will be useful in connection with a suggestion by David, Hartley & Pearson (1954) that a comparison of the range and root-mean-square estimators of the population standard deviation may serve as a test of homogeneity or as a routine check of accuracy in computation and also in connection with methods of interpolation suggested by Tukey (1955).

Motosaburo Masuyama (1957) derives upper and lower bounds on the ratios of the population standard deviation σ to the expectation of the sample range

and of the population variance σ^2 to the expectation of the square of the sample range, and suggests that the harmonic mean of the appropriate pair of these bounds may be used for all distributions as a multiplier of the sample range or its square in estimating σ or σ^2.

Rider (1957) studies the distribution of the midranges of samples from five symmetric populations of limited range and the relative efficiencies of sample midrange and mean in estimating the population midrange (which is identical with the population mean and median). He finds that the midrange is more efficient than the mean for all of the populations considered (which have standardized fourth moment α_4 = 2.19,2.14,1.19,1.19,1), and that its efficiency increases with decreasing α_4.

Masaaki Sibuya and Hideo Toda (1957), using an expansion formula given by Cadwell (1953), tabulate (to four decimal places) the probability density function of the range w in normal samples of size n=3(1)20 for w=0(0.05)7.65.

S. Babcock, A. Beck, A. Davies, B. Goldsmith and E. Torkelson (1958) introduce the median, quasi-range method for control of lot average and lot standard deviation for measurable lot quality characteristics which are normally distributed. They tabulate factors for computing upper and lower acceptance limits for the median and an upper acceptance limit for the optimal quasi-range for samples of size n=5(5)50.

D. E. Barton and D. J. Casley (1958) propose a quick estimate of the linear regression coefficient of y on x in a bivariate sample (x_i,y_i), i=1,2, ···, n, which they obtain by dividing the difference of the means of the k largest and the k smallest of the x's into the difference of the means of the corresponding y's. For large samples from a bivariate normal population,

the maximum efficiency (81 per cent) is attained when $k \doteq 0.27n$. For small samples the efficiency lies between 70 per cent and 80 per cent when k is between one-third and one-quarter of n.

Philip G. Carlson (1958) obtains a recurrence formula for $E(w_{2n+1})$, the expected value of the range of a sample of size 2n+1, in terms of $E(w_{n+i+1})$ for i=1,2,$\cdots$, n-1.

Ferrell (1958) suggests a method for computing control limits for samples from a strongly skewed universe which can be approximated by a lognormal distribution. The geometric range ρ= Max/Min is used in place of the range and the geometric midrange $m=\sqrt{\text{Max} \times \text{Min}}$ in place of the mean. The author describes corresponding changes in the computation of limits. This method accepts the skewness of the universe and allows a search for other assignable causes of variation.

Harter (1958) discusses the use of sample quasi-ranges in estimating the standard deviation of normal, rectangular and exponential populations. For the normal population, he tabulates the expected value, variance and standard deviation of the $r^{\underline{th}}$ quasi-range for samples of size n for r=0(1)8 and n=(2r+2)(1)100. For each pair of values of r and n, he also tabulates the efficiency of the unbiased estimator of population standard deviation based on one sample quasi-range. He also considers estimators based on a linear combination of two quasi-ranges, and gives a method for determining the weighting factor which maximizes the efficiency. The most efficient unbiased estimators based on one quasi-range for n=2(1)100 and on linear combinations of two adjacent quasi-ranges and of any two quasi-ranges ($r<r \leq 8$) for n=4(1)100 are tabulated, along with their efficiencies. These estimators are compared with those of

Grubbs and Weaver (1947) based on group ranges, and their use is illustrated by an example. For rectangular and exponential populations, the most efficient unbiased estimators based on one quasi-range are tabulated, together with their efficiencies and the bias when estimators which assume normality are used.

Bernard Ostle and J. M. Miesan (1958) express the distribution of the range of a sample of size n from a right triangular population in terms of the range of the population, and apply the result to an acceptance sampling problem.

Plackett (1958) examines the methods used by the ancient Babylonian and Greek astronomers in estimating parameters of observational data, and finds no evidence that they made use of the arithmetic mean of a group of comparable observations. He does trace its use as far back as the late sixteenth century, when Tycho Brahe applied it to astronomical observations in order to eliminate systematic errors. The concept of the mean as a more precise value than a single measurement was already known to de Moivre, Flamsteed and Maupertius early in the eighteenth century, but remained controversial until the second half of that century, when it was demonstrated conclusively by Simpson (1756, 1757) and Lagrange (1774), both of whom made use of results due to de Moivre. The author closes with an account of the work of Simpson and Lagrange, which we have already examined.

Rider (1958) considers the family of density functions $f(x)=c(1+|x-\theta|^{k})^{-h}$, $-\infty<x<\infty$, where the case h=1, k=2 is the well-known Cauchy density function, and compares the efficiency of the sample mean and the sample median as estimators of θ for various other values of h and k.

Jean Geffroy (1959) makes notable contributions to the theory of extreme values, including the proof of various results concerning stability in probability

and almost complete stability of midrange, quasi-midranges, and range of samples.

Gumbel (1959) expresses the asymptotic distribution of the reduced m^{th} range R_m, which is a certain linear transform of the m^{th} range w_m, in terms of the previously derived asymptotic distribution of R_1, the reduced range, and calculates its moments. For m close to n/2, where n is the sample size, he shows that the m^{th} range is asymptotically normally distributed and gives its mean and standard deviation.

Hermann Hänsel (1959) summarizes the results of investigations by Tippett (1925), E. S. Pearson (1932), Behrens (1956) and others on the use of range for the estimation of measures of variability. He points out that use of the range makes possible short-cut methods of ascertaining standard deviation with only a slight loss of accuracy which are applicable in every branch of biology.

Harter (1959) gives a revised and condensed version of the material in his earlier report [Harter (1958)] on the use of sample quasi-ranges in estimating population standard deviation. He points out that the standard deviation of an exponential population whose lower limit (location parameter) is known can be estimated more efficiently from a single order statistic than from a quasi-range.

Harter and Donald S. Clemm (1959) give a description of the computation and use of tables of the probability integral, percentage points and moments of the range for samples from a normal distribution. They include the following tables: (1) an eight-decimal-place table of the probability integral of the (standardized) range, W=w/σ, at intervals of 0.01, for samples of size n=2(1)20 (2)40(10)100;(2) a six-decimal-place table of percentage points of the range for the same values of n and cumulative probability P=.0001, .0005, .001, .005,

.01, .025, .05, .1 (.1) .9, .95, .975, .99, .995, .999, .9995, and .9999; and (3) a table of moments of the range [mean to 10 decimal places (11 significant figures), variance to 10DP(10SF), α_3 to 8 DP(8SF) and α_4 to 7DP(8SF)] for samples of size n=2(1)100.

Bernard Ostle and George P. Steck (1959) prove that the symmetry of the parent population implies that the sample mean and the sample range are uncorrelated, and construct an example to show that the converse is not true. They present a necessary and sufficient condition that the correlation between the mean and the range be positive (negative). They also prove that the symmetry of the parent population implies that the sample range and midrange are uncorrelated.

K.C.S. Pillai and Benjamin P. Tienzo (1959) develop, in series form, for n=3,4,5 and $\nu \leq 10$, the distribution of the standardized extreme deviate from the sample mean, $u=\max [(x_n - \bar{x})/\sigma, (\bar{x}-x_1)/\sigma]$ and the corresponding studentized deviate, $t_n = \max [(x_n - \bar{x})/s_\nu, (\bar{x} - x_1)/s_\nu]$, where $x_1 \leq x_2 \leq \cdots \leq x_n$ is an ordered sample of size n from a normal population with variance σ^2, $\bar{x}$ is the sample mean, and s_ν is the square root of an independent mean square estimate of σ^2 based on ν degrees of freedom. Pillai (1959) tabulates the upper 5% and 1% points of t_n for n=2(1)10,12 and ν=1(1)10, and discusses the method of preparation of this table.

Rider (1959) derives the distribution of the r^{th} quasi-range, $W_r = X_{n-r} - X_{r+1}$, where $x_1 \leq x_2 \leq \cdots \leq x_n$ are drawn at random from an exponentially distributed population, and gives the moment generating function and the cumulants of W_r. From these he shows that the mean of W_r slowly diverges with increasing sample size while the variance approaches a finite value: for example, $\pi^2/6 = 1.6449$ for r=0.

John Edward Walsh (1959) proposes a large-sample nonparametric criterion for rejection of outlying observations. Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be independent observations from continuous populations. The null hypothesis, H_0, is that these observations all resulted from independent random drawings from the same well-behaved population with unspecified shape. The alternative hypothesis is H_1: the i smallest observations are too small (or H_1': the i largest are too large) to be consistent with H_0, where i is a small number which should be specified without knowledge of the observations. The alternative H_1 is accepted if a statistic of the form $x_i - (1+A)x_{i+1} + Ax_k$ is negative, where $A>0$, k is the largest integer contained in $i+\sqrt{2n}$, and n is sufficiently large. Similarly, the alternative H_1' is accepted if $x_{n+1-i} - (1+A)x_{n-i} + Ax_{n+1-k}$ is positive. Two-sided tests are obtained by combining these one-sided tests. Tchebycheff's inequality yields an approximate upper bound for the significance level of the test for A suitably chosen.

H. Weiler (1959) shows that if $a>0$ is the smallest and b is the largest of n values whose arithmetic and harmonic means are $\bar{x}$ and H, respectively, then $0 \leq (\bar{x}-H)/H \leq (b-a)^2/4ab$, the first equality holding only if all n values are equal and the second only if half of them have the value a and the other half the value b. Moreover, since $H \leq g \leq \bar{x}$, where g is the geometric mean, the same inequality holds for $(\bar{x}-g)/g$. Thus $\bar{x}$ differs little from H or g if the n values have a small range and all are far removed from zero.

Frank J. Anscombe (1960) examines numerous criteria for the rejection of outliers proposed during a period of more than a century. He suggests that rejection rules should not be regarded as significance tests, as has usually been the case, but as insurance policies. He makes a detailed study of the effect

of routine application of rejection criteria to replicate (especially triplicate and quadruplicate) determinations of a single value, focussing attention mainly on rules appropriate when the population standard deviation σ is known, but giving some attention to studentized rules. He examines the following rules: Rule 0. For given C, reject every observation y_i ($i=1,2,\cdots,n$, where n is the number of observations) such that $|z_i|>C\sigma$, where $z_i = y_i - \bar{y}$, $\bar{y}=\sum y_i/n$. Estimate the mean μ by the mean of the retained observations. Rule 1. For given C, reject y_M if $|z_M|>C\sigma$, where M is the value such that $|z_M|>|z_i|$ for all $i\neq M$; otherwise no rejections. Estimate μ by the mean of the retained observations, thus $\hat{\mu} = \bar{y}$ if $|z_M|<C\sigma$, $\hat{\mu}=\bar{y}-z_M/(n-1)$ if $|z_M|>C\sigma$. Rule 2. Apply Rule 1. If an observation is rejected, consider the remaining observations as a sample of size n-1 and apply Rule 1 again; and so on. Estimate μ by the mean of the retained observations. The author finds Rule 0 unsatisfactory, since a single outlier, if it outlies sufficiently, can cause the entire sample to be rejected. He finds Rule 1 satisfactory for small samples ($n=3$ or 4), but since Rule 1 can reject only one outlier, Rule 2 must be considered for larger samples which may contain more than one outlier.

Dixon (1960) considers various estimators of population mean and standard deviation from censored normal samples. Among the estimators of the mean considered are the Winsorized means, in which the magnitude of an extreme observation which is unknown or poorly known (or suspected of being spurious) is replaced by the next largest (or smallest) observation, as proposed by Charles P. Winsor, instead of rejecting it entirely. Dixon finds that the efficiency of Winsorized means, when balance is maintained by Winsorizing the same number of observations at each extreme, is remarkably high relative to that of the

best linear systematic statistic.

Harter (1960) gives a condensed version of the material on the range of samples from a normal population contained in the report by Harter & Clemm (1959). The table of the probability integral is omitted, but those of the percentage points (abridged) and moments of the range are included, along with a section on interpolation in the tables which is not found in the report.

Robert Vincent Hogg (1960a) defines odd location statistics T and even location-free statistics S by $T(x_1+h, x_2+h, \cdots, x_n+h) = T(x_1,x_2,\cdots,x_n) + h$, $T(-x_1,-x_2,\cdots,-x_n) = -T(x_1,x_2,\cdots,x_n)$, $S(x_1+h, x_2+h, \cdots, x_n+h) = S(x_1, x_2,\cdots,x_n)$, $S(-x_1,-x_2,\cdots,-x_n) = S(x_1, x_2,\cdots,x_n)$ for all real values of h. He proves that the symmetry of a probability density function implies that the correlation between an odd location statistic and an even location-free statistic is zero. This generalizes two special results of Ostle & Steck (1959). The sample mean, the sample median, and the sample midrange are odd location statistics, while the sample variance, the sample range, the sample quasi-ranges, the sample mean deviation from the sample median, and any ratio of two of these statistics are even location-free statistics. Hogg (1960b) proves that if the distribution is symmetric about θ and ET exists, then $E\{T|S = s\} = \theta$, together with a multivariate extension useful in obtaining unbiased estimators of θ: e. g., $(R_2M_1 + R_1M_2)/(R_1 + R_2)$, where M_1 and M_2 are the medians and R_1 and R_2 the ranges of random samples from two distributions both of which are symmetric about θ.

William H. Kruskal (1960) gives an exposition of the problem of handling wild observations, or outliers. He suggests that such observations should be reported even though they may be excluded from the analysis; moreover, they

should not be discussed simply in terms of the propriety of including them in the analysis, of which the author gives illustrations, but treated as opportunities to learn something new. He classifies outliers into three categories according as there is (a) a priori knowledge, (b) a posteriori knowledge, or (c) no knowledge of a variant causal pattern. Those in the third category are the ones which cause the trouble, and the author expresses dissatisfaction with existing approaches to handling them.

Rider (1960a) compares exact variances with the values obtained by using the formula for asymptotic variance for the medians of small samples (n = 1,3, 5,7) from exponential, normal, cosine, parabolic, rectangular and inverted parabolic populations, which have standard fourth moment α_4= 9,3, 2.19, 2.1$\frac{1}{2}$, 1.8 and 1.61 respectively. He finds that the adequacy of the asymptotic formula increases with α_4. Rider (1960b) makes a similar comparison for the variance of the median of samples of size 2k + 1, for k=0(1)15, from a Cauchy distribution.

Tukey (1960) surveys sampling from contaminated distributions and reaches a number of conclusions, of which the following are relevant to the present study: (1) "In large samples the sample mean is not nearly so safe an indicator of location as is the mean of the observations which remain after a small percentage of the highest, and an equal percentage of the lowest, have been set aside (use of a lightly truncated mean)." (2)"In slightly large samples, there is ground for doubt that the use of the variance (or the standard deviation) as a basis for estimates of scaling type is ever truly safe." (3)"In moderately or very large samples, ··· the variance or standard deviation is safely used only [for certain purposes which the author specifies]." (4) "Nearly

imperceptible non-normalities may make conventional relative efficiencies of estimates of scale and location entirely useless." (5) "If contamination is a real possibility (and when is it not?), neither mean nor variance is likely to be a wisely chosen basis for making estimates from a large sample." (6) "As an interim measure, the use of truncated variances is likely to be quite satisfactory." (7) "In smaller samples, the use of the mean deviation may be a frequently useful compromise".

Anscombe (1961)considers four statistics designed to reveal certain types of departure from the ideal statistical conditions (independent and normally distributed residuals with zero mean and constant variance) under which the least-squares method of estimating the parameters in a regression equation is unquestionably satisfactory. He gives information about the distributions of these statistics under the null hypothesis of ideal conditions, but states that a thorough investigation of the appropriateness of the least-squares method would have to go further, and would encounter grave difficulties. He states that for most fields of observation, outliers may be expected to occur, so that significance tests to determine whether extreme observations do in fact occur with frequency incompatible with the ideal conditions may be irrelevant. He writes: "The day-to-day problem with outliers ... is not: is the ordinary least-squares method appropriate? but: how should [it] be modified? not: do gross errors occur sometimes? but: how can we protect ourselves from the gross errors that no doubt occasionally occur? The type of insurance usually adopted (it is not the only kind conceivable) is to reject completely any observation whose residual exceeds a tolerance calculated according to some rule, and then apply the least-squares method to the remaining observations."

He refers to his own earlier paper [Anscombe (1960)] containing suggestions for choosing a routine rejection rule and to the Bayesian approach of de Finetti (1961).

Eisenhart (1961,1962) summarizes the work of Boscovich on the combination of observations. He points out that Boscovich was the first to devise a completely objective procedure for uniquely determining the coefficients of a two-parameter line $y = \alpha + \beta x$ from a set of three or more observational points. He also notes that Boscovich's procedure, like the median, is comparatively insensitive to the more extreme of a set of observations, and is especially well suited to summarizing the linear trend evidenced by a more or less heterogeneous set of data compiled from various sources, or obtained by a measurement procedure that has a tendency to yield occasional discordant values. Besides Boscovich's geometric algorithm, the author discusses the algebraic formulation of Boscovich's method by Laplace and the modification by Edgeworth, who advocated unrestricted minimization of the sum of absolute values of the residuals, dropping the restriction that their algebraic sum must be zero, thus in effect requiring that the line pass through the double median point $(\tilde{x},\tilde{y})$ instead of the center of gravity $(\bar{x},\bar{y})$ of the observations. He also mentions the more recent work of Rhodes, Singleton, Harris, and Bejar, as well as the classical work on rival methods, including least squares.

Thomas S. Ferguson (1961a) derives locally best tests, based on the sample skewness $a_3 = \sqrt{b_1}$ and the sample kurtosis a_4 respectively, of the null hypothesis H_0 that a number of observations were all drawn at random from the same normal population $N(\mu,\sigma^2)$ against the alternatives H_A that one or more outliers came from $N(\mu+\lambda\sigma,\sigma^2)$ and H_B that one or more outliers came from

$N(\mu,\lambda^2\sigma^2)$. He compares the power of these tests with those proposed by Grubbs (1950) and Dixon (1950). Ferguson (1961b) surveys the literature on the rejection of outliers from the time of Peirce (1852) to date, with special emphasis on the period after 1950. He devotes one section to the relation between rejection and estimation, in which he discusses trimming and Winsorization.

Bruno de Finetti (1961) proposes a Bayesian approach to the treatment of outlying observations in which observations are never rejected, though the influence of outlying observations on the final distribution may be weak or almost negligible. He distinguishes three cases, in which the errors are (a) independent, (b) exchangeable, or (c) partially exchangeable, where independence means "independence with known error distribution", exchangeability translates "independence with unknown error distribution", and partial exchangeability translates "independence with an unknown conditional error related to visible features of the individual observations."

Harter (1961) gives examples of the use of tables of percentage points of the range [Harter & Clemm (1959) and Harter (1960)], including an application to rejection of outliers based on use of the test statistic $W = w/\sigma$ (the standardized range) as proposed by Dixon (1950).

M. G. Kendall (1961) reports the results of a historical study of the work of Daniel Bernoulli on the method of maximum likelihood. He sets the stage by reviewing the earlier contributions of Cotes (1722), Euler (1749), Mayer (1750), [Maire &] Boscovich (1755), Simpson (1756,1757), Lagrange (1774), and Laplace (1774). Kendall's comments are followed by English translations of the paper by Bernoulli (1778) and the related one by Euler (1778), which Kendall regards as less valuable.

C. P. Quesenberry and H. A. David (1961) point out that one may approach

the problem of testing for outliers differently depending on the object in view. If the primary interest is in pruning the observations so as to secure a more accurate analysis of what is left (e.g. to obtain the most reliable estimate of a mean) the criterion may be the effect on the standard error of estimate, whereas if the interest lies in identifying the exceptional observations so as to create a new insight into the phenomena under study, the criterion may be the risk of wrongly deciding whether an observation is exceptional or not. The authors take the second point of view. They modify the test statistics proposed by Nair (1948) and by Halperin et al (1955) for one-sided and two-sided tests, respectively, by replacing the independent estimate s_ν of population variance in the denominator by the pooled estimate $s^* = \{[(n-1)s^2 + \nu s_\nu^2]/(n+\nu-1)\}^{1/2}$, which makes use also of the internal estimate s^2 from the sample of size n. They compute and tabulate percentage points of the modified statistics.

Pranab Kumar Sen (1961a) studies some properties of the asymptotic variances of the sample quantiles and quasi-midranges and discusses the role of the sample median. He shows that, among the class of sample quantiles, the sample median has asymptotically the smallest variance only under somewhat restrictive regularity conditions; while among the class of sample quasi-midranges, the sample median has asymptotically the smallest variance only for a class of non-regular parent density functions. He tabulates the relative efficiency of the sample median with respect to the optimum quasi-midrange for nine common parent distributions. Sen (1961b) studies the stochastic convergence of the sample extreme values for distributions having a finite end-point and the asymptotic convergence of their moments to the corresponding ones of their

limiting distributions. He applies the results to the estimation of the population midrange from the sample midrange.

K. S. Srikantan (1961) treats the general problem of testing a regression model against the alternative hypothesis of a single outlier. He develops test criteria which are generalizations of those of Grubbs (1950) and of Pearson & Chandra Sekar (1936), the latter based on the work of Thompson (1935), and tabulates their 5% and 1% critical values for regression on m variables (m=1,2,3).

J. Tiago de Oliveira (1961) gives a general proof of the asymptotic independence of the sample mean and extremes for an absolutely continuous distribution satisfying the conditions of Gumbel (1946) for asymptotic independence of the extremes.

Simeon M. Berman (1962) shows, under general conditions, that if the standardized largest observation has a limiting distribution, then the studentized largest observation has the same limiting distribution and the studentized largest absolute deviate has a limiting distribution of the same form.

Giovanni Cancelliere (1962) gives a new proof of the theorem that the sum of the absolute values of the deviations of a set of observations $x_1, x_2, \cdots, x_n$ from a number x is a minimum when x is the median of the x_i $(i=1,\cdots,n)$.

Odoardo Cucconi (1962) proposes a criterion for the rejection of outlying observations from a k-dimensional distribution (k=1,2,3,$\cdots$) which is assumed to be k-variate normal, but possibly contaminated by spurious observations. This criterion is a generalization of that of Thompson (1935), to which it reduces for k=1. The criterion is $[N/(N-1)]\sum_{i=1}^{k}\sum_{j=1}^{k}(\Delta_{ij}/\Delta)({}_sx_i-m_i)({}_sx_j-m_j)$, $s=1,2,\cdots$, N, where ${}_rx_i$ $(r=1,2,\cdots,N;\ i=1,2,\cdots,k)$ is the $i^{\underline{th}}$ coordinate of the $r^{\underline{th}}$ observation, Δ is the value of the determinant of the matrix whose element $(c_{ij}=c_{ji})$ is given

by the sum of the products of the deviations of the ${}_rx_i$ and ${}_rx_j$ from their respective means m_i and m_j, and A_{ij} is the cofactor of c_{ij}. In order to test the hypothesis H_0 that the $s^{\underline{th}}$ (s=1,2,$\cdots$,N) observation is homogeneous with the others, the value of the criterion is calculated and compared with the critical value $r^{(k)}_{\alpha/N}$. The values of $r^{(k)}_{\alpha/N}$, obtained from the incomplete Beta distribution, are tabulated to two decimal places for α=0.05,0.01; k=1,2,3,4; and various values of N ranging from 5 to 100.

Harter (1962), as part of a study of the ratio of two ranges not otherwise relevant to the present topic, tabulates (to 8DP) the probability density function of the standardized range W=w/σ for samples of size n=2(1)16 from $N(\mu,1^2)$, at intervals of 0.01 in W. This table represents a considerable improvement over the earlier 4DP table of Sibuya & Toda (1957) at intervals of 0.05 in W. Harter's values for W a multiple of 0.05, when rounded to 4DP, agree with those of Sibuya & Toda except for an occasional discrepancy of one unit in the last place.

Bruce Marvin Hill (1962) proposes a test of linearity versus convexity of a median regression curve. Specifically, he proposes to test H_0: $Y_i=\alpha+\beta X_i+\epsilon_i$ against H_1: $Y_i=\phi(X_i)+\epsilon_i$, i=0,1,$\cdots$,n, where α,β and ϕ are unspecified and $\phi(x)$ is a nonlinear convex function, the ϵ_i are independent indentically distributed random variables with median zero and a continuous density function $f(\epsilon)$ such that $f(0)>0$, and the X_i are fixed and known. The test involves estimating a line from a central subset of the observations by the procedure (using medians) of Brown & Mood (1951), making a weighted count of the number of remaining observations lying above the line, and rejecting H_0 if this number, R_n, is too large. The author gives the asymptotic distribution of R_n under H_0, from which

he obtains critical values of R_n, and the asymptotic distribution of R_n under H_1, from which he obtains the power of the test. The test can be adapted to two-sided alternatives.

Alex Rosengard (1962) seeks to unify the existing theory of limiting distributions of the mean and of the extremes of a sample by studying the limiting joint distribution of these three statistics.

Sibuya (1962) examines an asymptotic formula for the expected value of the median of the ranges of N independent normal samples each of size n. He compares the approximate values obtained from this formula with exact values obtained by numerical integration for n=2, N=3(2)17.

M. M. Siddiqui (1962) makes a numerical study of the method proposed by Chu & Hotelling (1955) for approximating the moments of the sample median by use of a Taylor series expansion of the inverse of the cumulative distribution function. He applies this method to various distributions and presents the results in tabular form. They show that the relative error decreases monotonically with sample size, and generally support the author's expectations that properties of the parent population which contribute to rapidity of convergence are finite range, a low value of kurtosis, and symmetry.

Tukey (1962), in a study of the future of data analysis, devotes considerable attention to "spotty data" resulting from long-tailed fluctuation-and-error distributions, occasional causes with large effects, or irregularly non-constant variability. He offers a number of possible cures or palliatives, including trimming and Winsorizing samples, which result in a small loss in efficiency when the samples come from a normal distribution but a large gain in efficiency when they come from a very long-tailed one (e.g. the Cauchy distribution). For

two-dimensional arrays he proposes graphical methods to be applied to the residuals, including (1) a conventional plot on normal probability paper, (2) a modified plot of $z_i = (y_i - \tilde{y})/a_{i|n}$ against i, where the $y_i (i=1,2,\cdots,n)$ are the residuals, $\tilde{y}$ is their median, and $a_{i|n}$ is the standard normal deviate corresponding to the cumulative probability of the $i^{\underline{th}}$ order statistic of a sample size n, and (3) an arithmetic analogue of the modified plot called FUNOP (from FULl NOrmal Plot). He proposes a specific procedure called FUNOR-FUNOM (FULL NOrmal Rejection-FULl NOrmal Modification) because it uses FUNOP and first rejects and then modifies deviations. This procedure is a sort of two-dimensional analogue of trimming and Winsorization, since it first rejects (trims) the most extreme deviations (those greater than $A_R \cdot \sigma$) and then reduces to $B_M \cdot a_{i|n} \cdot \sigma$ the remaining deviations exceeding the latter value, both A_R and B_M being prechosen.

Anscombe and Tukey (1963) emphasize the importance of examination and analysis of residuals, which may furnish information about the presence of outliers and/or about inappropriateness of the fitted curve or the scale of measurement. They suggest use of both graphical and analytic techniques, but suggest beginning with the former, preferably in the form of a scatter diagram in which residuals are plotted against fitted values. When the most prominent sort of misbehavior of the data has been diagnosed, it is important, they say, to deal with it before seeking out other sorts of misbehavior. If outliers are detected they may be rejected outright or modified by Winsorization or by assigning them smaller weights which decrease smoothly as the size of the residual increases. If the signs of the residuals show a definite pattern, this may indicate that the type of curve fitted is inappropriate. If the spread of

the residuals is correlated with the fitted values, this may indicate that the scale of measurement is inappropriate. These two phenomena are, of course, not independent; e.g. a straight line may adequately fit the data resulting from a transformation of scale, even though there was evidence of curvilinearity on the original scale of measurement.

Eisenhart, Deming & Martin (1963) give tables to accompany their earlier abstracts [Eisenhart, Deming & Martin (1948a,b)] concerning the distributions of the median and the mean of samples from various populations. The abstracts are reprinted (slightly edited).

Harter (1963) tabulates (to 8DP) the probability integral of the standardized $r^{\underline{th}}$ quasi-range $W_r = w_r/\sigma$, at intervals of 0.01 in W_r, for r=0(1)8 and samples of size n=(2r+2)(1)20(2)40(10)100 from a normal population with unit variance, $N(\mu,1^2)$, obtained by numerical integration. He also tabulates (to 6 DP) percentage points of W_r corresponding to cumulative probability P=0.0001, 0.0005, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1(0.1)0.9, 0.95, 0.975, 0.99, 0.995, 0.999, 0.9995, 0.9999 for the same values of r and n, obtained by inverse interpolation in the table of the probability integral.

Joseph L. Hodges, Jr. and Erich L. Lehmann (1963) propose various estimates of location based on rank tests. Of particular interest is the estimate $\hat{\theta}$, which is the median of the N(N+1)/2 averages $(z_i + z_j)/2$ of the $i^{\underline{th}}$ and $j^{\underline{th}}$ order statistics ($i \leq j$) of a sample of size N. The authors show that this estimate, which we shall call the Hodges-Lehmann estimate, has, under specified conditions, certain properties of regularity, invariance, symmetry, median unbiasedness and asymptotic normality. For samples from a normal population, it has asymptotic efficiency $3/\pi \doteq .955$ relative to the sample mean.

Andre G. Laurent (1963) presents an analogue to the distribution of $(X_i - \bar{X})/s$ [Thompson (1935)], where X_i is a random observation from a sample with mean $\bar{X}$ and s^2 is an independent root-mean-square estimate of the population variance, in case the distribution of the underlying population is exponential. He discusses its use in obtaining minimum variance unbiased estimates of functions of the parameters and the probability distributions of the reduced i^{th} order statistic and the reduced range and in tests for exponentiality or the presence of outliers.

Gerald J. Lieberman and Rupert G. Miller, Jr. (1963) in a study of simultaneous tolerance intervals in regression, include a section on detection and correction of outliers using simultaneous confidence principles.

B. S. Niven (1963), recalling that the sample mean and the sample range are uncorrelated if the parent population is symmetric [Ostle & Steck (1959) and Hogg (1960)] and independent if it is normal [Lord (1947)--see also Daly (1946)], gives a method suitable for the calculation of their joint distribution when the sample size is small. She gives specific results for samples of sizes three and four from rectangular and exponential populations, and recalls [McKay & Pearson (1933)] that for samples of size 3 from a normal population, the distribution of the range may be written in terms of the normal probability integral.

John W. Tukey and Donald H. McLaughlin (1963) discuss trimming and Winsorization. Given n ordered observations $y_1 \leq y_2 \leq \cdots \leq y_n$, their arithmetic mean is $\bar{y} = (y_1 + y_2 + \cdots + y_n)/n$, their g-times (symmetrically) trimmed mean is $y_{Tg} = (y_{g+1} + y_{g+2} + \cdots + y_{n-g})/(n-2g)$, and their g-times (symmetrically) Winsorized mean is $y_{Wg} = (g \cdot y_{g+1} + y_{g+1} + y_{g+2} + \cdots + y_{n-g} + g \cdot y_{n-g})/n$. Clearly both

y_{Tg} and y_{Wg} pay less attention to extreme values than does $\bar{y}$, but $y_{W,g}$ does not divert attention from the tails of the sample so completely as does y_{Tg}. For underlying distributions whose shapes are very close to Gaussian, the Winsorized means are less variable than the trimmed means. When the underlying distribution is Gaussian , the efficiency of the trimmed means is quite high, the fractional loss being crudely 2g/3n (corresponding to efficiency of about 2/3 for the median), but that of the corresponding Winsorized means is much higher. On the other hand, trimmed means are clearly much more efficient than Winsorized means for samples from very long-tailed distributions. The authors raise, but do not answer, the question as to where the transition takes place. They define g-times (symmetrically) trimmed and Winsorized sums of squared deviations in an analogous manner: $SSD_{Tg} = (y_{g+1} - y_{Tg})^2 + (y_{g+2} - y_{Tg})^2 + \cdots + (y_{n-g} - y_{Tg})^2$; $SSD_{Wg} = g(y_{g+1} - y_{Wg})^2 + (y_{g+1} - y_{Wg})^2 + (y_{g+2} - y_{Wg})^2 + \cdots + (y_{n-g} - y_{Wg})^2 + g(y_{n-g} - y_{Wg})^2$.

Miles (1963) deals with the problem of identifying and testing a candidate set of a small number t of extreme sample elements as significant outliers in a sample of size n from a k-dimensional normal distribution with unknown parameters. He considers the problem in detail for t=1,2,3,4. He defines criteria r_1 and r_2, respectively, for testing a single observation as a significant outlier and a pair of observations as significant outliers, small values of r_1 and r_2 constituting the critical regions. Exact probabilities $P(r_1<r)$ and $P(r_2<r)$ are extremely complicated, but the author gives values of r_α for which the upper bounds of $P(r_1<r_\alpha)$ and $P(r_2<r_\alpha)$ have the value α for α=0.010, 0.025, 0.050, 0.100; k=1,2,3,4,5; and n=5(1)30(5)100(100)500.

Victor Chew (1964) discusses statistical criteria for the rejection of

observations suspected to contain gross errors, covering the cases of one population or several populations, univariate or multivariate data, and correlated or uncorrelated observations. He points out the weaknesses in some of the classical rejection procedures, develops new procedures, and recommends existing procedures when appropriate. He also tabulates critical values for many criteria. In the case of a sample of independent observations from a single normal distribution with unknown mean and variance, he recommends using Dixon's or Grubbs' criterion and avoiding Chauvenet's. If an independent estimate of population variance is available, he recommends the criterion of Nair or that of Quesenberry & David. For a random sample from a bivariate normal distribution, he proposes a procedure based on the maximum radial distance if the parameters are known; otherwise, he recommends Wilks' criterion. He points out that residuals from a regression analysis are not only correlated but [often] also have heterogeneous variances. Though he admits that not much work has been done in this area, he recommends trying the methods of Lieberman and Miller and of Srikantan, remarking that the latter may be more convenient. For g samples of n independent observations each, the method of Dixon or Grubbs can be applied to the deviations from the sample means, though the control chart approach may be more convenient for routine applications, especially if samples are obtained sequentially.

Cyrus Derman (1964) shows that, for the truncated Cauchy distribution with p.d.f. $g_z(x) = 1/2(1+x^2)\ \tan^{-1}z$ for $-z<x<z$ and 0 otherwise, the variance of the mean of a sample of size n is $(z-\tan^{-1}z)/n\ \tan^{-1}z$, while the asymptotic variance of the sample median is $(\tan^{-1}z)^2/n$. Hence the efficiency of the sample mean relative to the sample median is $(\tan^{-1}z)^3/(z-\tan^{-1}z)$, which exceeds unity

if and only if $z<3.41$, representing a truncation of more than 9%.

Eisenhart (1964), in a discussion of the meaning of "least" in least squares, points out that the method of least squares was developed originally from three distinct points of view which differ not only in their aims and in their initial assumptions, but also in the meanings that they attach to the numerical results common to all three. These viewpoints are: (1) Least Sum of Squared Residuals [Legendre (1805)]; (2) Maximum Probability of Zero Error of Estimation [Gauss (1809)]; and (3) Least Mean Squared Error of Estimation [Gauss (1823)]. He closes with the following remarks: "The robust survival of the Method of Least Squares as a valuable tool of applied science no doubt stems in part from the algebraic and arithmetical advantage of Least Sum of Squared Residuals and in part from the fact this procedure also yields estimates of Least Mean Squared Error in the important case when the end results are linear functions of the basic observations. This one-to-one correspondence between minimizing some function of the residuals and minimizing the same function of Errors of Estimation appears to be a unique property of Least Squares. And although the Method of Least Squares does not lead to the best available estimates of unknown parameters when the law of error is other than the Gaussian, if the number of independent observations available is much larger than the number of parameters to be determined the Method of Least Squares can be usually counted on to yield nearly-best estimates".

Friedrich Gebhardt (1964) points out that some of the many procedures that have been proposed for handling outlying observations are based on statistics with the optimum property of minimizing, for certain alternative hypotheses, the probability of the error of the second kind (accepting the null hypothesis

H_0 when it is false) given that of the error of the first kind (rejecting H_0 when it is true), the observations that are not rejected being used to estimate unknown parameters, e.g. the mean. He considers one such procedure which gives rise to a one-parameter family of estimators for the mean and compares their risks with those of the Bayes solutions with respect to a one-parameter family of prior distributions.

Peter J. Huber (1964) treats in detail the theory of robust estimation of a location parameter of a contaminated normal distribution with c.d.f. $F(t)=(1-\epsilon)\Phi(t)+\epsilon H(t)$, $0\leq\epsilon<1$, where ϵ is a known number, $\Phi(t)$ is the standard normal c.d.f., and $H(t)$ is an unknown c.d.f. He seeks an estimator, intermediate between the sample mean and the sample median, that is robust against deviations from normality. Let $x_1, x_2, \cdots, x_n$ be a random sample of size n, and let the estimator $T=T_n(x_1,x_2,\cdots,x_n)$ be chosen so as to minimize $\sum_{i=1}^{n}\rho(x_i-T)$, where ρ is a known function. If we take $\rho(t)=t^2$ we get the usual least-squares estimator, the sample mean, while $\rho(t)=|t|$ yields the sample median and $\rho(t)=-\log f(t)$, where f is the assumed density, yields the maximum likelihood estimator. Huber shows that the most robust estimator (the one with the lowest supremum of the asymptotic variance when H ranges over all symmetric distributions) corresponds to $\rho(t)=t^2/2$ for $|t|<k$, $\rho(t)=k|t|-k^2/2$ for $|t|\geq k$, where k and ϵ are related by $\sqrt{2\pi}(1-\epsilon)=\int_{-k}^{k} e^{-t^2/2}dt+(2/k)e^{-k^2/2}$. He also considers robust estimation of a scale parameter, which he finds more difficult and less satisfactory.

K. V. Mardia (1964) obtains the exact distributions of extremes, ranges and midranges in samples from any multivariate population. He lets $(x_{1j},\cdots,x_{kj})$, $j=1,\cdots,n$, be a random sample from a k-variate continuous population with

p.d.f.$f(x_1,\cdots,x_k)$, and denotes the minimum and maximum observations of the $i^{\underline{th}}$ variate by X_i and Y_i respectively, the $i^{\underline{th}}$ range by $R_i (= Y_i - X_i)$ and the $i^{\underline{th}}$ midrange by $V_i (= \{Y_i + X_i\}/2)$, where $i=1,\cdots,k$. He finds the distributions of $(X_1,\cdots,X_k, Y_1,\cdots, Y_k)$, $(R_1,\cdots, R_k)$ and $(V_1,\cdots, V_k)$, which reduce to the classical forms for $k=1$.

Mary G. Natrella (1964) compares the variances of the unique medians of samples of size n (for odd values of n) with those of the pseudo-medians (the averages of the two central observations) for n even. She investigates samples from normal, rectangular, and extreme-value distributions, and finds that no general conclusion is possible as to whether it is better to take n odd or even.

Albert Stanley Paulson (1964) gives a probability basis for the computation of certain measures of effectiveness of test statistics and derives analytical expressions for these measures. He computes these measures for several test statistics for the rejection of outliers and makes comparisons to show the degree to which some statistics are better than others. In particular, he finds that the standardized extreme deviate test is more efficient than the chi-square test in detecting location error when the population variance is known. When the population variance is unknown, he finds that the Quesenberry-David statistic using a pooled estimate of the population standard deviation in the denominator is more efficient than the studentized extreme deviate.

E. S. Pearson and M. A. Stephens (1964) extend the table of percentage points of the ratio $u=w/s$ of the range w of a sample of n observations from a normal population having standard deviation σ to the root-mean-square estimate s of σ derived from the same sample, which was computed by David, Hartley & Pearson (1954), and test the accuracy of the approximation used by those authors.

Alex Rosengard (1964a,c) establishes results on the limiting independence of means (quantiles) and extreme values not related to the existence of limiting distributions (reduced limiting distributions) for these statistics. Rosengard (1964b) shows that, when the variance exists, the joint distribution of the mean and a quantile has, as its limiting form, a specified bivariate normal distribution.

Thomas J. Rothenberg, Franklin M. Fisher and C. B. Tilanus (1964) propose a class of estimators of the center of the Cauchy distribution. Each estimator in the class is the arithmetic mean of a central subset of the sample order statistics. The sample median is a member of this class, but it is not the most efficient. The average of approximately the middle quarter of the ordered sample has the lowest asymptotic variance.

Asit Prokas Basu (1965) proposes some tests for outliers in the case of the exponential distribution with p.d.f. $f(x)=e^{-(x-\mu)/\theta}/\theta, x\geq\mu, \mu\geq 0, \theta>0$. When μ and θ are known, the following test statistics, whose distributions are easily obtained, may be used to test whether the largest value x_n of an ordered sample of size n is an outlier: $B_0=(x_n-\mu)/\theta$, $B_1=(x_n-x_1)/\theta$, $B_2=(x_n-x_{n-1})/\theta$. Similar tests can be devised for the smallest value x_1. As an overall test of the presence of outliers one may use the χ^2- test, since under the null hypothesis $2\sum_{i=1}^{n}(n-i+1)z_i/\theta$, where $z_i=x_i-x_{i-1}$ and $x_0=\mu$, follows the chi-square distribution with 2n degrees of freedom. When μ and θ are not known, one can use the standardized deviate $U_n=(x_n-x_1)/\sum(x_i-x_1)$, whose distribution has been derived by Laurent (1963).

G. P. Bhattacharjee (1965) investigates the effect of non-normality on the distribution of range by deriving the probability integral and the first two

moments of the range of a sample drawn from a population represented by the first four terms of an Edgeworth series. He then examines the use of range in place of root-mean-square deviation as an estimator of the standard deviation of a non-normal population. He concludes that the range estimator of the population standard deviation is better for a platykurtic parent population (and worse for a leptokurtic one) than the corresponding estimator for a normal population, thus contradicting an erroneous conclusion reached by Cox (1954).

Peter J. Bickel (1965) states the main results of the asymptotic theory of the Winsorized and trimmed means and outlines the proof. He discusses an alternative method of trimming and Winsorization (not equivalent to that of Tukey) which encompasses the efficient estimates proposed by Huber and generalizes to higher dimensions. He gives the minimum efficiency, with respect to the families of all symmetric and symmetric unimodal distributions, of Winsorized and trimmed means with respect to the mean. He compares the trimmed mean and the Winsorized mean with the Hodges-Lehmann estimate (the median of averages of pairs) and the principal estimate proposed by Huber with the mean and the Hodges-Lehmann estimate. He concludes that although all the proposed "nonparametric" estimates of location behave satisfactorily when compared with the mean, with the possible exception of the Winsorized mean, the Hodges-Lehmann estimate seems to be the "safest" among them.

H. A. David and A. S. Paulson (1965) summarize and extend the results given by Paulson (1964) on the performance of several tests for outliers. They note that such tests "generally have one of the following aims: (a) to screen data in routine fashion preparatory to analysis (the problem of 'rejection of outliers'); (b) to sound an alarm that outliers are present, thus indicating the

need for closer study of the data-generating process; (c) to pin-point observations which may be of special interest just because they are extreme." They do not treat case (a), which is the one of the primary interest in the present study but they do cite some references to it.

E. J. Gumbel, P. G. Carlson and C. K. Mustafi (1965) prove that if the initial distribution (parent population) is unlimited, differentiable, symmetrical and unimodal, the distribution of the midrange, for any sample size, is also unlimited, differentiable, symmetrical and unimodal. This extends a result of Gumbel (1944).

Shanti S. Gupta and Bhupendra K. Shah (1965) derive the exact expressions for the moments of the order statistics of samples of size n from a standard logistic distribution L(0,1), where $L(\mu,\sigma^2)$ has the c.d.f. $F(y;\mu,\sigma) = 1/[1+\exp\{-[(y-\mu)/\sigma]\cdot(\pi/3^{1/2})\}]$. They tabulate the first four exact moments of the $k^{\underline{th}}$ order statistic X(k) for n=1(1)10, k=1(1)n. They also tabulate percentage points of $X_{(k)}$ for n=1(1)10, k=1(1)n and for n=11(1)25, k=1, n and n/2, (n+2)/2 (n even) or (n+1)/2(n odd). They also derive expressions in closed form for the cumulative distribution function and the density function of the range, both of which they tabulate for n=2,3. Shah (1965) obtains the distributions of semirange and midrange of samples from the logistic population.

K. V. Mardia (1965) gives alternative proofs of the formulas of Tippett (1925) for the expected values E(R) and $E[R-E(R)]^m$, where m is a positive integer and R is the range of a sample of size n from a continuous population. These proofs are simpler than those given by Tippett and other authors, and hold for all n, whereas Tippett, in his proof of the latter formula, assumed n to be even. The author also finds the exact value of the variance of the range of a sample of

size n=3 from a normal population; similar results for n=2 and n=4 were given by Ruben (1956).

Michael E. Tarter and Virginia A. Clark (1965) show that the cumulative distribution function (c.d.f.) and the moment generating function (m.g.f.) of the logistic distribution can each be expressed as a Maclaurin series where the coefficients are simple functions of Bernoulli numbers. They give the m.g.f. of the median, and determine the variance of the median, also the efficiency of the median relative to the mean for various sample sizes, as well as its asymptotic efficiency. They also give the variance of any order statistic and the covariance of any two order statistics.

Tukey (1965) studies the informativeness of specific order statistics or blocks of consecutive order statistics in a sample.

F. J. Anscombe and Bruce A. Barron (1966) consider a particular procedure for rejecting outliers and also a particular procedure for modifying outliers, for samples of size three assumed to have been drawn from a common normal population, except that one of the three readings may have an added bias. They give numerical results illustrating the effects of the procedures on estimation of the location parameter. They conclude that estimation by least squares should usually be tempered by successive application of both a rejection rule and a modification rule.

R. P. Bland, R. D. Gilbert, C. H. Kapadia and D. B. Owen (1966) extend the results of McKay & Pearson (1933), Lord (1947), Resnikoff (1954), Harter & Clemm (1959) and other authors to obtain exact results for additional cases of the distributions of the range and mean range for samples from a normal population.

Dorian Feldman and Howard G. Tucker (1966) study consistent estimates of non-unique quantiles of a distribution function. As a special case they consider the problem of medians, especially the sample median of the set of averages of all $\binom{n+1}{2}$ pairs of observations $X_1, X_2, \cdots, X_n$ [the Hodges-Lehmann estimate of the location parameter]. They prove that this sample median converges almost surely to the center median of the original population, provided that the original distribution is symmetric about a median; otherwise, this sample median of averages of pairs need not converge, and even if it did converge, it might converge to a number which is not a median of the parent distribution.

Joseph L. Gastwirth (1966) discusses a procedure for finding robust estimators, based on robust rank tests, of the location parameter of the symmetric unimodal distributions. Not only can the Hodges-Lehmann estimator be constructed from the author's procedure, but this procedure can also be used to generate another estimator T, which is the best linear unbiased estimator of the location parameter. The best linear unbiased estimator corresponding to the least favorable distribution of Huber (1964) is the trimmed mean.

Friedrich Gebhardt (1966) relaxes the restriction in his earlier paper [Gebhardt (1964)] that the respective variances of stragglers and non-stragglers be known by requiring only that the ratio of these variances be known. The results for a variety of cases which he studies support the suggestion of Tukey (1962) to trim the sample by all observations that deviate substantially from the sample mean and to Winsorize those observations that deviate moderately, but trimming exactly two observations is almost always a better strategy than Winsorizing two.

J. Likeš (1966) finds the distributions of Dixon's statistics for rejection

of outliers in the case of a sample from an exponential population, tabulates their percentage points, and gives examples of their use.

Donald T. Searls (1966) proves that if a Winsorized mean is formed by replacing all sample values larger than a predetermined cutoff point t by the value t itself, there exists a region for t for certain common distributions such that the mean square error of the Winsorized mean is smaller than the variance of the ordinary mean. He presents an example which shows that a wide range of cutoff points can be chosen which still result in a gain.

O. B. Sheynin (1966) contends that J. H. Lambert should be given precedence over Gauss as the originator of the theory of errors. He gives a concise summary of Lambert's works on the subject, which was the principal source of information used by the present writer concerning these works [Lambert (1760, 1765a,b)].

Thomas A. Willke (1966) reports the results of a sampling study of the estimation of the mean and standard deviation from the closest two of three observations in a sample from a normal population contaminated by slippage of the mean. The results of Lieblein (1952), which indicated that use of the closest two out of three is not advisable for noncontaminated samples, are borne out by this study for contaminated samples as well.

J. N. Adichie (1967) defines point estimates $\hat{\alpha}$ and $\hat{\beta}$ of the parameters α and β in the linear regression equation $Y=\alpha+\beta X$ in terms of certain statistics used to test hypotheses concerning α and β. He shows that the least squares estimates are a special case of these estimates. He proves that "rank score" estimates exist and shows how to compute them. He discusses both the small-sample and asymptotic properties of these estimates. He shows that

$\hat{\alpha}$ and $\hat{\beta}$ are unbiased if the underlying distribution of observations is symmetric. He also proves that $\hat{\alpha}$ and $\hat{\beta}$ are jointly asymptotically normal, and that the asymptotic efficiency of $(\hat{\alpha},\hat{\beta})$ is the same as the Pitman efficiency of the rank tests on which they are based, relative to the classical tests. Finally he compares the efficiencies of these estimates and the Brown-Mood median estimates.

F. J. Anscombe (1967) reviews various topics relevant to the present study, including the effect of modern computers on statistical calculation, "stepwise regression", testing goodness of fit by examining residuals, and possible alternatives to the method of least squares appropriate when the distribution of errors has long tails. He points out that Laplace and Gauss justified using the method of least squares and restricting attention to linear combinations of the observations largely on the basis of computational simplicity and feasibility. In the age of computers this justification is no longer valid. Suppose one has n sets of observations on a dependent variable and p independent variables, denoted respectively by y_i and x_{ir} $(i=1,2,\cdots, n;\ r=1,2,\cdots,p)$. If it is assumed that the true y is a linear function of the x's, one can set $y_i = \sum_{r=1}^{p} x_{ir}\beta_r + \epsilon_i (i=1,2,\cdots,n)$, where the ϵ's are independent with zero mean, so that $\mu_i = \sum_{r=1}^{p} x_{ir}\beta_r$, where μ_i is the expected value of y_i. The problem is then to estimate the parameters β_r as precisely as possible. In the classical theory it is assumed that the ϵ's are normally distributed, so that the method of least squares is optimal. Anscombe points out that it is possible to fit the β's by stages (e.g., one at a time) and advocates testing goodness of fit by examining the residuals. He then raises the question as to what should be done if the distribution of the ϵ's is not normal. If the distribution is

skewed, he suggests that it may be symmetrized by transforming the y-scale, usually by raising each y to some fixed power (or by taking logarithms); he points out, however, that such a transformation has other consequences which may or may not be desirable. If the distribution of the ε's is symmetric but platykurtic (shorter-tailed than the normal) or leptokurtic (longer tailed than the normal), a different remedy is required, involving departure from the method of least squares. Anscombe studies in detail the case of a long-tailed distribution of errors. If ones wishes only to find estimates of the θ's without any indication of their precision, he suggests minimizing, instead of $\sum_i (y_i-\mu_i)^2$, $\sum_i \psi(y_i-\mu_2)$, where $\psi(\cdot)$ is the square function for small values of the argument but increases less rapidly for larger values and is constant for very large values. Specifically, he suggests choosing as the estimates $(\hat{\beta}_r)$ the values of (β_r) that minimize $\sum_{(1)}(y_i-\mu_i)^2 + \sum_{(2)}K_1(2|y_i-\mu_i|-K_1) + \sum_{(3)}K_1(2K_2-K_1)$, where K_1 and K_2 are chosen numbers $(K_2>K_1>0)$, $\sum_{(1)}$ denotes summation over the values of i such that $|y_i-\mu_i|\leq K_1$, $\sum_{(2)}$ denotes summation over the values of i such that $K_1<|y_i-\mu_i|\leq K_2$, and $\sum_{(3)}$ denotes summation over the remaining values such that $|y_i-\mu_i|>K_2$. If, on the other hand, one desires evidence from the data concerning the precision of the estimates $(\hat{\beta}_r)$, it is necessary to make some assumption about the true distribution of the ε's. Let the error density function be represented by $f(\varepsilon|\alpha,\sigma)$, where typically α is a shape parameter and σ is a scale parameter. He notes that apparently the only kind of density $f(\varepsilon|\alpha,\sigma)$ that permits easy integration with respect to σ in closed form is $f(\varepsilon|\alpha,\sigma)=(a_\alpha/\sigma)\exp(-c_\alpha|\varepsilon/\sigma|^\alpha)$, where a_α and c_α are functions of α. When $\alpha=2$ we have normality, and when $1<\alpha<2$ we have a smooth function of ε with longer tails than the normal density. When

$m=1$ we have the double exponential (Laplace's first) distribution, concerning which Jeffreys (1939) [2nd ed. (1948), Sec. 4.4] has commented: "The interest of the law is reduced somewhat by the fact that there do not appear to be any cases where it is true." Anscombe expresses the opinion that the same remark can be made about the distribution when $m=1.5$ (say). He advocates instead Jeffreys' form of the Pearson Type VII error distribution, with density function $f(\varepsilon|m,\sigma)=(a_m/\sqrt{2\pi}\sigma)(1+c_m\varepsilon^2/\sigma^2)^{-m}$ where $c_m = m^2/2(m-1/2)^3$, $a_m = \sqrt{2c_m}\,\Gamma(m)/\Gamma(m-1/2)$, which approaches normality as $m\to\infty$, for some appropriate value of m. Anscombe suggests $m=4$, for which the Type VII distribution is equivalent to the Student t distribution with 7 degrees of freedom. He proceeds to investigate the likelihood function of the Type VII distribution. He points out that maximizing this likelihood function is rather like minimizing the expression $\sum_{(1)} + \sum_{(2)} + \sum_{(3)}$ mentioned above. For the Type VII distribution, $m>1/2$; however, for negative values of m, the same density function gives a Type II distribution over a finite interval (a distribution with shorter tails than the normal).

Allan Birnbaum and Eugene M. Laska (1967a,b) present a general method of determining efficiency-robust estimation methods, which they use to derive admissible linear unbiased estimators (whose respective variances, over a specified family of symmetric shapes of the error distribution, cannot be jointly improved) and maximin-efficient linear unbiased estimators (which maximize the minimum asymptotic efficiency, within a class of estimators, for a family of densities).

Edwin L. Crow and M. M. Siddiqui (1967) derive robust estimators of the location parameter which are efficient over a class of two or more forms

(pencils) of continuous symmetric unimodal distributions. The pencils considered are the normal, double exponential, Cauchy, parabolic, triangular, and rectangular. The estimators considered are trimmed means, Winsorized means, "linearly weighted" means, and a combination of the median and two other order statistics. Asymptotically these are compared with the Hodges-Lehmann estimator. The best trimmed mean or linearly weighted mean has an asymptotic efficiency of at least 0.82, relative to the best estimator for any single pencil, over a range of pencils of distributions from the normal to the Cauchy, while the combination of the median and two other order statistics is at least 0.80 efficient over the same range.

Hodges (1967) makes a further study of the Hodges-Lehmann estimate T, which he recognizes as a member of a class of estimates. He explores this class for other members which are easier to compute, and finds that one of the simplest of these, D, which is defined as the median of the means of pairs of symmetric order statistics, corresponds to the one-sample analog of Galton's rank-order test. He applies D to the same samples used with T and obtains very similar results. Finally, he compares a number of estimates, including $\bar{X}$ (the mean), T,D, and the trimmed and Winsorized means with regard to normal efficiency, ease of computation, and extreme value tolerance. Bickel and Hodges (1967) derive the asymptotic theory of Galton's test and the related estimate D, which gives an explicit form to the limiting distribution of D only for rectangular and Laplace parent populations. Although the limit is not normal, they conclude that the scatter of D is quite close to that of T. Finally, they give the small-sample distribution of D for a rectangular parent. Although their evidence is incomplete, they conclude that D is robust as well as easy to compute.

Hodges and Lehmann (1967) compute, for samples of size n from a symmetric distribution satisfying suitable regularity conditions, an approximation to the variance of the median $\tilde{X}$ up to terms of order $1/n^2$ and a corresponding approximation to the efficiency of $\tilde{X}$ relative to the mean $\bar{X}$ up to terms of order $1/n$. For normal and rectangular distributions, these give a much closer approximation to the exact efficiency than does the usual asymptotic efficiency. The authors point out that, to the accuracy of their approximation, one should not use the median based on an odd number of observations since the median based on the next smaller even number is equally accurate. They extend their results to other averages of two symmetrically placed order statistics (quasi-medians), thus making possible, in some cases, a further reduction in sample size without loss of accuracy.

Robert V. Hogg (1967) finds an estimator T, which is a weighted mean of $T_1, T_2, \cdots, T_m$, where T_j is a reasonable estimator (e.g., a minimum mean square error estimator of the parameter θ of the family D_j of distributions, $j=1,2, \cdots, m$), such that T has the same asymptotic distribution as that of T_j, when the sample comes from D_j. The weights are functions of the sample items. The author gives empirical evidence that T is satisfactory for small sample sizes. He proves that if T_j and the weight W_j are odd location and even location-free statistics, respectively, then $T=\sum W_i T_i$, where $\sum W_i = 1$, is an unbiased estimators of the center of every symmetric distribution, provided certain expectations exist. This fact is useful in constructing the weight function W_j. The particular T which the author investigates empirically is given by $T=\bar{X}^c_{1/4}$ for $k<2.0$, $T=\bar{X}$ for $2.0 \leq k \leq 4.0$, $T=\bar{X}_{1/4}$ for $4.0<k\leq 5.5$, $T=m$ for $5.5\ k$, where $\bar{X}^c_{1/4}$ is the mean of the [n/4] smallest and the [n/4] largest items in the sample (an interior-

trimmed mean), $\bar{X}_{1/4}$ is the mean of the remaining interior sample items (an exterior-trimmed mean), $\bar{X}$ is the sample mean, m is the sample median, and $k=\sum(x_i-\bar{x})^4/ns^4$ is the sample kurtosis. He compares the ratio of the variances of $\bar{X}$,m,M(the Hodges-Lehmann estimate) and T for 200 samples each of sizes 7 and 25 from four populations with kurtosis K=1.9,2.7,3.9,9.9. For both sample sizes, T has the smallest variance for K=1.9, $\bar{X}$ for K=2.7, and m for K=3.9,9.9. The mean performs very poorly for distributions with long tails (high K), the median performs rather poorly for those with short tails (low K), and the overall performance of both M and T is good.

Fred C. Leone, Toke Jayachandran and Stanley Eisenstat (1967) report on an empirical study of the performance of the sample mean and the Hodges-Lehmann and Huber estimators of the location parameter when applied to contaminated distributions. They verify Huber's statement that his estimator T and the Hodges-Lehmann estimator are close competitors and his conjecture that, for a sample of size n, the distribution of the ratio of $\sqrt{n}$ times his estimator T of the location parameter to the estimator S of the scale parameter can be approximated by a Student t-distribution. They compare the sample variance of $n^{1/2}$. T with its maximal asymptotic variance as given by Huber. They also verify, by means of a chi-square test of goodness of fit, that the distributions of the Hodges-Lehmann estimator and of Huber's T can be approximated fairly well by appropriate normal distributions when $n\geq 20$.

Max Ray Mickey, Olive Jean Dunn and Virginia A. Clark (1967) point out that the examination of residuals is not always sufficient to identify outliers in a regression model, and propose stepwise regression. Their procedure finds the single observation whose deletion causes the greatest reduction in the sum

of squared residuals, then repeats the process on the remaining data. The selection of stopping rules is left open.

M. M. Siddiqui and K. Raghunandanan (1967) supplement the paper of Crow and Siddiqui (1967) by a study of the robustness properties of four estimators of location (a weighted average of the median and two other symmetric order statistics, the trimmed mean, the Winsorized mean, and the Hodges-Lehmann estimator) with respect to eight distribution types (normal, .01 and .05 contaminated normal, logistic, Student's t with 3 and 5 degrees of freedom, double exponential, and Cauchy). For each of these types the probability density function is continuous and symmetric about the mean and the range is infinite. The estimator with the highest guaranteed efficiency for the entire class of distributions is the mean of the middle 50% of the sample. The authors state that the Hodges-Lehmann estimator was first suggested by Tukey, but the present writer finds no evidence of this in the source cited, where Tukey does use the averages of all pairs of observations, which he calls Walsh averages, but does not suggest using their median as an estimator of the location parameter.

Chatter Singh (1967) obtains expressions for the raw moments and the probability integral of the largest (smallest) value and for the first two moments of the range of samples from non-normal populations represented by the first four terms of the Edgeworth series. He gives some conclusions about the nature of the effects of parental skewness and kurtosis. The mean largest (smallest) value is sensitive to parental skewness but not to kurtosis in small samples. However, parental kurtosis tends to increase or decrease the variance depending upon whether $\lambda_4=\alpha_4 -3$ is positive or negative. The mean range is quite insensitive to population changes in small samples, but both skewness and kurtosis have

a greater effect on the variance of the range. The author compares his results with those of Pearson (1950), Cox (1954), David (1954) and others.

G. C. Tiao and Irwin Guttman (1967) point out that a major difficulty involved in statistical procedures, designed to guard against the occurrence of outliers or spurious observations, which are based upon examining the magnitude of the residuals, is caused by the fact that the residuals are correlated. They show how to avoid this difficulty by adjusting the residuals on the basis of information from an auxiliary experiment so that the adjusted residuals become uncorrelated. This leads to a set of estimation procedures, for the unknown mean of a normal population $N(\mu,\sigma^2)$ with known σ^2, in which one or more observations for which the magnitudes of the adjusted residuals are largest will be excluded. The authors discuss certain properties of these procedures, give exact numerical results for the cases of one and two spurious observations, and generalize to the case of unknown variance.

G. K. Bhattacharyya (1968) obtains median and weighted median estimates for the linear trend parameters of a univariate time series by applying the Hodges-Lehmann method to some well-known nonparametric tests for trend. He extends the estimation procedure to the multivariate trend model, and studies its asymptotic efficiency properties relative to the classical estimates.

G. E. P. Box and G. C. Tiao (1968) consider the problem of outlying observations from a Bayesian viewpoint. They assume that each observation in an experiment may come from either a "good" run or a "bad" run. They specify the models corresponding to good and bad runs [$N(\mu,\sigma^2)$ and $N(\mu,k^2\sigma^2)$, respectively] and the prior probability α that a run is bad, and employ standard Bayesian inference procedures to derive the appropriate analysis. They give an example

of the application of their method to actual data, and examine the sensitivity of the results to changes in c and k.

Irving Wingate Burr and Peter J. Cislak (1968) give, in closed form, the density function of the median for odd sized samples from the Burr system of distributions [which has c.d.f. $F(x)= 1-(1+x^{c})^{-k}$, $x\geq 0$, c, k>0; $F(x)= 0$, $x<0$ and covers almost all of the regions of the main Pearson Types IV and VI and an important part of that of the main Type I (Beta distribution)]. All finite moments of the median $\tilde{x}$ are linear combinations of Beta functions. For samples of size n=3,5,7 and 11 from Burr populations with $\alpha_{3:x}$= 0,.50, 1.00, 1.50 and, corresponding to each $\alpha_{3:x}$, two well separated values of $\alpha_{4:x}$, the authors tabulate the following important characteristics of the median: Bias, $\alpha_{\tilde{x}}$, $\alpha_{3:\tilde{x}}$, $\alpha_{4:\tilde{x}}$ and efficiency relative to the sample mean. They point out that it appears that, for this system, the median begins to be more efficient than the mean at about the degree of non-normality of the exponential distribution. Burr (1968) tabulates, for samples of size n=2,3,4,5,8,10 from populations of the Burr system with 27 different combinations of $\alpha_{3:x}$ and $\alpha_{4:x}$, the following characteristics for the distribution of range R: standardized mean and standard deviation, $\alpha_{3:R}$, $\alpha_{4:R}$ and coefficient of variation. He confirms that the standardized mean range is highly stable for fixed n under varying non-normality,as has been pointed out in the literature, and finds that the same is true of the standardized standard deviation for the populations studied, which have $\alpha_{4:x}> 2.87$. He also finds evidence that the range is somewhat more robust and efficient than hitherto noted.

Theophilos Cacoullos (1968) presents a sequential scheme for detecting outliers in a sample from a p-variate normal population $N(\mu,\sigma^{2}I)$ in which

the components are independently normally distributed with the same variance σ^2. At the $k^{\underline{th}}$ stage of experimentation, when the first k observations are available, the most recent observation x_k is rejected as an outlier if and only if $|x_k - \bar{x}_k|^2 > c_\alpha \sigma^2$ when σ^2 is known or $|x_k - \bar{x}_k|^2 > c_\alpha s_k^2$ when both μ and σ^2 are unknown, where $\bar{x}_k = \sum_{i=1}^{k} x_i/k$, $s_k^2 = \sum_{i=1}^{k} |x_i - \bar{x}_k|^2/p(k-1)$, $|x|$ denotes the length of x and c_α denotes the upper α point of the χ^2- distribution with p degrees of freedom. The author considers the sequential stopping rule which stops taking observations as soon as the first outlier is rejected. He proves that this scheme terminates with probability one after a finite number of steps and that the number of observations N has moments of every order. The probability P_k that the $k^{\underline{th}}$ observation is rejected appears to be monotone increasing for reasonable critical values of c and approaches α as $k \to \infty$.

D. R. Cox and E. J. Snell (1968) enumerate the following types of departure, from the usual linear regression model of one independent variable on n independent variables with errors ε normally distributed with zero mean and constant variance, which can be detected by an appropriate analysis of residuals: (1) the presence of outliers; (2) the relevance of a factor omitted from the model, detected by plotting the residuals against the levels of that factor; (3) non-linear regression on a factor already included in the model, detected by plotting the residuals against the levels of that factor and obtaining a curved relationship; (4) correlation between different ε_i's, for example between ε_i's adjacent in time, detected from scatter diagrams of suitable pairs of residuals, or possibly from a periodogram analysis of residuals; (5) non-constancy of variance, detected by plotting residuals or squared residuals against factors thought to affect the variance, or against

fitted values; (6) non-normality of the distribution of the ϵ_i's, detected by plotting the ordered residuals against the expected values of the order statistics from a standard normal distribution. The authors give a more general definition of residuals and find some asymptotic properties. They also discuss some illustrative examples, including a regression problem involving exponentially distributed errors.

D. R. Cox (1968) makes miscellaneous comments on various aspects of regression analysis, including outliers and robust estimation. He points out that screening of data for suspect observations will often be required. Suspect values may be examined individually in order to decide whether or not to include them in any subsequent analysis; this is the usual procedure with limited data. Often it is necessary to perform analyses both with and without suspect values. When p observations are available for each individual, the best way of looking for outliers will depend on the type of effect expected, of which the author discusses three. With extensive data, he suggests the use of methods of robust estimation, such as that proposed by Huber (1964), which are insensitive to outliers.

D. R. Cox and D. V. Hinkley (1969) consider a linear regression model in which the errors are independent and identically distributed with zero mean. If the type of error distribution is specified, the asymptotic efficiency of least-squares estimates relative to maximum-likelihood estimates of the regression parameters can be found, and the authors calculate it explicitly for an Edgeworth series, for a Pearson Type VII distribution [suggested by Anscombe (1967)] and for a log Gamma distribution of errors.

A. S. C. Ehrenberg (1968) quotes various authors on the subject of the

justification for adopting the least-squares approach to regression analysis. He says it appears to be a case of the practical man accepting the theoreticians' judgement that it will give the "best" solution, and the latter assuming the "best fit" is what the former wants.

Frank Rudolf Hampel (1968) studies the problem of robust estimation. In the one-dimensional case, he finds the optimal solutions (with regard to asymptotic variance and sensitivity) for the class of (sufficiently regular) M-estimators as defined by Huber. The optimal estimators of the location parameter in the model of normality turn out to be the Huber estimators and the trimmed means.

J. A. Hartigan (1968) defines a Bayes measure of discordance of an observation x, given a set of observations $x_1, x_2, \cdots, x_n$, to be the distance between the posterior distributions of a parameter, in the presence or absence of x. He also proposes a measure of dissimilarity between two observations. When the number of observations is large, these two measures may be approximated by simple functions of the log likelihood, thereby avoiding dependence on prior distributions.

Arnljot Høyland (1968) studies the behavior of the Hodges-Lehmann estimator of location θ^* and the classical estimator $\hat{\theta}$(the arithmetic mean) in a situation where the data occur naturally grouped in n blocks, c observations per block, with the experimental conditions varying from block to block, thus invalidating the standard assumption that the observations are independent and identically distributed. In particular, he studies the asymptotic efficiency, as $n \to \infty$ with c fixed, of θ^* relative to $\hat{\theta}$ for normal and gross error models. The relative efficiency is less than 1 for the former and greater than 1 for the latter in all cases studied.

Peter J. Huber (1968), in a survey paper on robust estimation, begins by attacking the dogma of normality and the associated rule of the arithmetic mean. He points out that many mathematicians of the nineteenth and early twentieth centuries realized that they were not universally valid, but in most cases continued to behave as if they were, not realizing how bad the classical estimates could be in slightly non-normal situations. The turning point did not come until after World War II, when Tukey and his associates began to emphasize the shortcomings of the classical estimates and propose practicable alternatives to them. Huber defines what he means by robust estimators and enumerates four distinct goals to be achieved by them. He gives three methods of constructing robust estimates: (1) maximum likelihood; (2) linear combinations of order statistics; (3) estimates based on rank tests. He introduces the idea of asymptotic robustness, and attempts to answer criticisms that have been levelled at asymptotic theory, restriction to symmetric distributions and minimax theory, all of which play important roles in robust estimation. He also considers the question of ease of computation, pointing out that the Hodges-Lehmann estimate, for a sample of size n, requires $0(n^2)$ operations, as compared with approximately $0(n \log n)$ for Hodges' alternative, Huber's estimate, and the trimmed and Winsorized means. He points out, however, that Hodges' alternative estimate is not asymptotically normal. He closes by mentioning some other problems (largely unsolved) in robust estimation: (1) estimation of scale parameters; (2) estimation of location parameters in the multivariate case; (3) estimation in the absence of translation and scale invariance; and (4) regression and analysis of variance problems.

Masao Kogure and Hajime Makabe (1968) study the non-central distribution

of the standardized range and give an application to a process capability study through a control chart with trend line.

P. Prescott (1968) suggests a simple estimator of the standard deviation of a normal population as an alternative to the usual root-mean-square estimator. The proposed estimator, which is one-third of the difference between the means of the largest one-sixth and the smallest one-sixth of the observations, has an asymptotic efficiency of 0.956.

Pranab Kumar Sen (1968a) studies the robust-efficiency of the Hodges-Lehmann estimator when the n observations are drawn from distributions which are symmetric about their medians and have continuous c.d.f.'s $F_1(x), \cdots, F_n(x)$ which are not necessarily identical. Sen (1968b) studies a simple robust unbiased estimator of the regression coefficient β based on Kendall's rank correlation coefficient tau. The estimator is the median of the set of slopes $(Y_j - Y_i)/(X_j - X_i)$ joining pairs of points with $X_i \neq X_j$. Sen compares its properties with those of the least squares estimator and some other nonparametric estimators.

L. de Haan and J. Th. Runnenburg (1969) study the distribution of the quotient $h_n = x_{n+1}/(x_{2n+1} - x_1)$ of the sample median and the sample range for a sample of size 2n+1 from a standard normal distribution. They give the c.d.f. of h_1 and the p.d.f. of h_2, and show that h_n is asympotically normal. N. Bouma and A.Vehmeyer (1969) tabulate the percentiles and the second and fourth moments (when they exist) for n=1,2,3 and 4 (sample sizes 2n+1=3,5,7 and 9). Their tabulation is based partly on the theoretical results of de Haan and Runnenburg and partly on numerical results obtained by Monte Carlo methods. They approximate the distribution of h_n by a Student t distribution with $\nu_n = 2E^2(r_n^2)/\sigma^2(r_n^2)$

degrees of freedom.

M. Mahamunulu Desu and Robert H. Rodine (1969) point out that the sample median is a median unbiased estimator of the median of a continuous population for odd sample sizes, but not necessarily for even sample sizes. For symmetric populations, however, the median of a sample of even size, as an estimator of the population median ξ, is median unbiased and also unbiased in the usual sense, and there is a class of unbiased and median unbiased estimators of ξ which includes the sample median and the sample midrange. The authors give an estimator, using a random selection of pairs of symmetrically placed order statistics, Y_r and Y_{n-r+1}, which is a median unbiased estimator for any population and unbiased for symmetric populations.

James John Filliben (1969) examines the behavior of various linear estimators of location when the underlying distribution is known (simple estimation) and when it is not known, but is known to belong to a prespecified set S (robust estimation). The estimators considered include best linear unbiased estimators and various modifications thereof, as well as trimmed and Winsorized means. The set S consists of 34 symmetric unimodal distributions; optimal linear robust estimators are found for S and various subsets of S.

Joseph L. Gastwirth and Herman Rubin (1969) consider the problem of finding, for the location parameters of symmetric unimodal distributions, robust estimators which are linear functions of the ordered observations. In particular, they study the maximin efficient linear estimators and admissible linear estimators proposed by Birnbaum and Laska, and obtain asymptotic generalizations. They demonstrate that within a large class of linear estimators there is a unique maximin efficient linear estimator for general families of densities.

They discuss in detail the special case in which the family of densities contains the logistic and double exponential distributions, for which they find the maximin efficient linear estimator and compare it with the best convex combination of the individual optimum linear estimators and with a Hodges-Lehmann type estimator based on the corresponding maximin rank test. Because of computational difficulties, they look for a maximin efficient estimator in smaller classes of linear estimators which are easy to use, including the trimmed means and linear combinations of a few sample percentiles. They show that, under suitable regularity conditions, a maximin efficient estimator for each of these classes exists, and give some numerical examples.

F. E. Grubbs (1969) gives an expository treatment of procedures for determining statistically whether the highest observation, the lowest observation, the highest and lowest observations, the two highest observations, the two lowest observations, or more of the observations in the sample are statistical outliers. Included are statistical formulae and tables of critical values for tests of significance to be applied in detecting outliers in single samples, as well as examples of their application.

Irwin Guttman and D. E. Smith (1969) investigate the performance of three rules for dealing with outliers in small samples from the normal distribution $N(\mu,\sigma^2)$ when the primary objective of sampling is to obtain an accurate estimate of μ. They assume that at most one observation in the sample may have arisen from either $N(\mu+a\sigma,\sigma^2)$ or $N(\mu,(1+b)\sigma^2)$, and measure the performance of each rule in terms of "Protection", the fractional decrease in the mean square error obtained by using the rule when such an observation is actually present in the sample. Numerical results are given for $n \leq 10$ when σ^2 is known, but only for

n=3 when σ^2 is unknown.

Douglas M. Hawkins (1969) derives the distributions, under null and alternative hypotheses, of the statistic proposed by Quesenberry & David (1961) for detecting the presence of a single outlier.

Louis Alan Jaeckel (1969) considers various robust estimates of a location parameter. He definds three types of location estimators: maximum likelihood type estimators, linear combinations of order statistics, and estimators derived from rank tests. He gives some relationships among the three types, and shows that Huber's minimax result applies to all three. He considers two flexible estimation procedures in which the observations are used to choose an estimator from a family of possible estimators. The families considered include the trimmed means and a "weighted median" of pairwise means derived from an arbitrary rank test.

C. L. Narayana and M. Subrahmanyam (1969) suggest an alternative to the method of least squares in the theory of regression, the object being to reduce the computations to a minimum and still obtain fairly accurate estimates of the slope of the regression line. They first compute the slopes of the lines joining each pair of data points, and take (i) a simple average of these slopes, (ii) a weighted average with weights equal to the denominators of the respective slopes, or (iii) a weighted average with weights equal to the squares of the denominators of the respective slopes. They prove that (iii) is equivalent to the method of least squares, and show by examples that (i) and (ii) are simpler and nearly as efficient.

P. V. Rao and J. I. Thornby (1969) define a robust point estimator of the parameter β in the generalized regression model $y_j=\alpha+g_j(\beta)+z_j, j=1,2,\cdots,n$, where

α and β are unknown parameters, $g_1, g_2, \cdots, g_n$ are real-valued functions of a real variable satisfying suitable conditions and $z_1, z_2, \cdots, z_n$ are independent identically distributed random variables having a distribution function belonging to a specified class. An important special case of this model is the regression model obtained by setting $g_j(\beta) = \beta x_j$, $j=1,2,\cdots,n$, where the x's are known constants. For this case, Adichie has proposed a robust estimator of β of the Hodges-Lehmann type and Brown and Mood have proposed a median estimator. A third alternative is provided by the estimator proposed by the authors, which is also of the Hodges-Lehmann type.

Ram Swaroop, Kenneth A. West and Charles E. Lewis, Jr. (1969) present a statistical technique, and the related computer program, for identifying the outliers in univariate data.

Kei Takeuchi (1969) proposes an estimator of the location parameter of a continuous symmetric distribution which is a linear combination of the order statistics of a (fictitious) subsample of size k drawn randomly from the order statistics of a sample of size n. He proves that this estimator is asymptotically efficient for a wide class of distributions satisfying certain regularity conditions, and shows by a Monte Carlo study that a modified version with symmetric coefficients attains high relative efficiency for several varieties of distributions even for small sample size (n=10,15,20 or even n=5).

Jerry Thomas (1969) reports the results of a Monte Carlo investigation of the effect of non-normality on the distribution of Dixon's criteria for detecting outlying observations. He reaches the conclusion that Dixon's criteria are not robust and may yield incorrect decisions for skewed distributions.

John E. Walsh (1969) studies the sample size n required for approximate

independence between the sample median and the largest (or smallest) order statistic, which he measures by the maximum value ε of the difference between their true joint probability and the corresponding value assuming independence. He finds that the following inequality holds: $2n+1 \geq -1+e^{-2}/2\pi\varepsilon^2 \approx -1+.0215/\varepsilon^2$ ($\varepsilon \leq .02$).

Takashi Yanagawa (1969) proposes a new robust estimate of location defined by $L_{N,p} = \sum_{i_1<i_2<\cdots<i_p} \text{med}\,(X_{i_1}, X_{i_2}, \cdots, X_{i_p})/\binom{N}{p}$, which is the mean of the medians, $\binom{N}{p}$ in number, of p-tuples $(X_{i_1}, X_{i_2}, \cdots, X_{i_p})$ obtained from the original random sample of size N. He compares this estimate for p=3 with other estimates for small samples from normal and double exponential populations, and finds that it is the most robust in a class including the sample mean, the best linear unbiased estimate for the double exponential distribution, and the Hodges alternative to the Hodges-Lehmann estimate.

V. P. Zelenen'kiy (1969) considers methods, based on statistical decision theory, for the exclusion of anomalous measurements of random processes. He proposes various solutions with differing amounts of information about the a priori statistical characteristics of the measured processes and the measurement errors.

V. D. Barnett (1970) studies the problem, suggested by a medical example, of fitting a linear functional model with replicated observations and inhomogeneous error variances. For a particular error structure relevant to the example, he finds maximum-likelihood estimators of the parameters in the model (slope, intercept and error variances). He obtains simple closed-form expressions for the asymptotic standard errors of the estimators, even though the estimators have no simple explicit form and must be evaluated by iterative methods.

Allan Birnbaum and Valerie Mike (1970) develop approximate versions of optimally robust Pitman-type estimators of location, and show that they have full asymptotic efficiency for a prototype family of distributions. By means of a Monte Carlo study for moderate n (20 to 100), they show that these estimators have efficiencies of 88% or more for normal, logistic, double exponential, and contaminated normal distributions.

Wayne A. Fuller (1970) investigates simple estimators of the mean of skewed populations under the assumption that the tail of the distribution is well approximated by the tail of a Weibull distribution. He considers relatively simple estimators, in particular those that are linear in the order statistics or that may be expressed as a linear function of the order statistics with weights that depend on a preliminary test. The loss in efficiency when the proposed estimators are used for populations for which the sample mean performs well (such as the exponential) is very small relative to the gain for heavily skewed populations.

J. L. Gastwirth and M. L. Cohen (1970) discuss the small-sample behavior of various robust linear estimators. They find that for sample size 20 the variances of these estimators are well approximated by asymptotic theory. If the observations are assumed to come from a family of distributions consisting of the Cauchy, double exponential, normal, contaminated normal and logistic distributions, then the asymptotically maximin efficient estimator, the 27-1/2% trimmed mean, is the maximin efficient estimator for size 16 or greater, but its minimum relative efficiency for samples of size 16 is less than asymptotic theory suggests. If the Cauchy distribution is deleted from the above family, the 20% trimmed mean is the maximin efficient linear estimator for all sample sizes studied.

Harter (1970) collects in two volumes various results (theory and tables), from earlier publications authored or co-authored by him, on order statistics and their use in testing and estimation. Volume 1 includes tables of probability integral, percentage points and moments of the range of samples from a normal populations [first published by Harter & Clemm (1959)] and a table of the probability density function of the range of samples from a normal population [first published by Harter (1962)]. Volume 2 includes tables of expected values, variances, standard deviations, probability integral and percentage points of quasi-ranges of samples from a normal population [previously published by Harter (1958,1959,1962)] and for the probability integral and percentage points of the range of samples from a rectangular population [previously published by Harter (1961a)].

Huber (1970) considers the problem of studentizing robust estimates. Let T be a robust estimate of a (location) parameter θ. Huber notes that little has been written about the estimated standard deviation (e.s.d.) s(T) appropriate for T or about the studentized ratio $(T-\theta)/s(T)$, beyond the general philosophy outlined by Tukey & McLaughlin (1963): Choose the T in the numerator to achieve a high robustness of performance; then match it with a denominator s(T) to achieve a high robustness of validity over a broad range of distributions. As a result of his investigation, Huber draws the following conclusions: (1) The trimmed mean, scaled by the Winsorized e.s.d., has both excellent small sample and excellent large sample properties; since it is also easy to compute, it can be strongly recommended for practical use; (2) The trimmed mean and a maximum likelihood type estimate T developed by Huber behave well even if the underlying distribution fails to have a density, but the corresponding estimates s(T) do not.

D. G. Kabe (1970) expresses the distributions of Dixon's statistics for the rejection of outlying observations in the case of an exponential population in terms of finite series of Beta functions, from which the probabilities of rejection of suspected outliers can be easily calculated on a desk calculator, thus making tables such as those of Likeš (1966) unnecessary.

R. M. Loynes (1970) obtains bounds on the asymptotic relative efficiency, as an estimator of the central value of a symmetric distribution, of the power mean of order q with respect to the power mean of order p, where $1<p<q$, which are $(p-1)^2/(q-1)^2$ and ∞, both bounds being the best possible. If the underlying distribution is unimodal, the lower bound can be improved to $(2p-1)/(2q-1)$, again the best possible.

C. Singh (1970) obtains the probability integral of the range of samples from a population whose distribution can be represented by the first four terms of an Edgeworth series. He tabulates the numerical values of the corrective functions arising because of nonnormality. He compares the new theoretical results with the earlier results of various authors, including Pearson (1950), Cox (1954), David (1954) and Singh (1967). He concludes that, for Edgeworth type populations, the effect of parental skewness on the probability integral and percentage points of the range is comparatively small for samples of size $n<5$, but as n becomes larger this effect becomes as prominent as that of kurtosis except in the lower tail of the distribution. The effect of $\lambda_3^2 = \alpha_3^2 = \beta_1$ is almost always opposite to the effect of $\lambda_4 = \alpha_4 - 3 = \beta_2 - 3$, and they sometimes counterbalance each other so that the resultant is close to the normal theory value.

Paul Switzer (1970) discusses various methods of obtaining robust estimators

of a location parameter θ. He proposes choosing (say) three reasonable candidate estimators for which non-parametric estimates of the standard error are available, then actually computing each for the data at hand and using the one which has the smallest estimated standard error. One general procedure is to assume that the sample of size N can be divided into K blocks of equal size n=N/K, compute the three estimates for each block, $\hat{\theta}_i^k$, k=1,2,$\cdots$, K; i=1,2,3; then take the overall $\hat{\theta}_i$ to be the average of the block estimates, $\hat{\theta}_i = \sum_{k=1}^{K} \hat{\theta}_i^k/K$, and estimate its standard error by $S_i = [\sum_{k=1}^{K} (\hat{\theta}_i^K - \hat{\theta}_i)^2/K(K-1)]^{1/2}$; choose that $\hat{\theta}_i$ for which the estimated standard error S_i is smallest. As a special case he considers a sample size N divisible by 6; he divides the data into K=N/6 equal groups (at random). In each group k(of size 6), he computes, as the candidate estimates, $\hat{\theta}_1^k = (X_3 + X_4)/2$, $\hat{\theta}_2^k = (X_2 + X_5)/2$, and $\hat{\theta}_3^k = (X_1 + X_6)/2$, where the X's are the ordered values in the group. He reports the results of a Monte Carlo study of this procedure applied to samples of size N=30,60, and 120 drawn from a short-tailed (uniform) distribution, a normal distribution, and a long-tailed (contaminated normal) distribution. The results are as expected, with $\hat{\theta}_3$ strongly favored by the short-tailed distribution and strongly disfavored by the long-tailed distribution.

Allan Birnbaum, Eugene Laska and Morris Meisner (1971) determine maximin-efficient linear unbiased estimators (MLUEs), and their efficiencies, for ordered samples of sizes 5(5)20 from a family of nine distributions. Since these distributions admit simple orderings such that the MLUE over any subset is just the MLUE over the extreme pair in the ordered subset, they are able to summarize the results compactly. They also discuss relations to other estimators.

F. C. Duckworth (1971) demonstrates that the standard method of least squares

is unsuitable for analyzing certain types of data, either because it fails to appreciate the strong dependence of the prediction error on the independent variable or because it is unable to allow for any error introduced by the fitting of an inappropriate equation. He contends that a semi-subjective graphical technique may often give more useful results, though he admits that such a method also has disadvantages--besides being partially subjective, it is also rather cumbersome and not easily computerizable, and there is no rigorous method of determining the error of estimate.

Arthur L. Edwards (1971) presents a set of computer subroutines, FITTER, FUNGUS, and FUNNY, written for finding linear least-squares fits of weighted tabular data with any of several functional forms, and for evaluating the final function at specified values of the independent variable. Provision is made for several functional forms, including power series in the independent variable or its reciprocal, with or without a constant term, and other functional forms may be added as desired.

M. V. Johns, Jr. (1971) develops a sequence of estimators, indexed by an integer-valued parameter k, which are asymptotically efficiency-robust in the sense that, for any k, the corresponding estimator is consistent and asymptotically normally distributed (as the sample size n increases) for any F in a large subset $\mathcal{F}$ of the class of symmetric distributions and, for large k, the corresponding estimator is (nearly) best asymptotically normal for all $F \in \mathcal{F}$. The simplest non-trivial estimator in the proposed sequence (corresponding to k=2) exhibits quite high efficiencies for small to moderate sample sizes (n=10,20,40) for a collection of distributions comprising the normal, the Cauchy, the logistic, the double exponential, and the 10% contaminated normal.

Elmer E. Rasmenga and R. G. Burdick (1971) describe a stepwise computer procedure for identifying and setting aside extreme values from sets of data with minimum bias or subjectivity on the part of the analyst. Since truncation of a basically normal distribution causes a downward bias in the estimated variance, a graphical method is provided to compensate for the bias.

Ram Swaroop and William R. Winter (1971) present a statistical technique and the necessary computer program for editing multivariate data. The technique is especially useful when large quantities of data are collected and the editing must be performed automatically. One task in the editing process is the identification of outliers which deviate markedly from the rest of the sample. The technique presented, a multivariate analog of the univariate technique of Swaroop, West & Lewis (1969), considers the statistical linear relationship between the variables in identifying the outliers. It is assumed that the data are from a multivariate normal population and that the sample size exceeds the number of variables by at least two.

John Caso (1972) presents the results of an extensive literature search in the area of robust estimation techniques. He gives a descriptive analysis of several robust estimators of the location parameter of symmetric distributions. These estimators, chosen because they are computationally and theoretically tractable and can be easily understood by a practitioner, are the trimmed and Winsorized means, the Hodges-Lehmann estimator, Huber's estimator, Hogg's estimator and Switzer's estimator. Caso also reports the results of a Monte Carlo study, based on 4200 samples each of sizes 12 and 24 from five symmetric probability distributions (rectangular, triangular, normal, contaminated normal and double exponential), of the efficiency of the robust estimators relative to the

best estimator for the distribution under consideration. The results show that the robust estimators provide a higher guaranteed efficiency than the best estimator for any particular distribution in the family.

Eisenhart (1972) traces the development of the concept of the best mean of a set of measurements from antiquity to the present day. He reports instances of the use of the mode by the Greeks as far back as the fifth century B. C. and of the midrange by the Arabs around 1000 A. D. The median and the arithmetic mean apparently came into use much later, around 1600. By the early nineteenth century, the principle of the arithmetic mean had become widely (though not universally accepted), but it met its downfall at the hands of Poisson (1824), who pointed out that for the Cauchy distribution with p.d.f. $f(x)= 1/\pi(1+x^2)$, $-\infty<x<+\infty$, the arithmetic mean has the same distribution (for which the moments do not exist) as a single observation. The author also comments on the theory of statistical estimation from the time of Laplace, D. Bernoulli, and Gauss to the present, and closes with some comments on modern robust estimation, which he says arose out of World War II arguments as to whether mean deviation or standard deviation is a better measure of dispersion in gunnery and bombing situations.

6. CONCLUSIONS AND RECOMMENDATIONS

1. The best choices of measures of central tendency and dispersion and of methods for fitting linear (or nonlinear) regression equations depend upon the error law, i. e. the distribution of the errors or residuals. For the three common laws of error shown on the accompanying graph (Figure 1), the best choices are as follows:

L1 DOUBLE EXPONENTIAL
(LAPLACE'S FIRST)

L2 NORMAL-GAUSSIAN
(LAPLACE'S SECOND)

UN UNIFORM-RECTANGULAR

$$f(x) = \frac{1}{\sigma\sqrt{2}} e^{-\sqrt{2}|x-\mu|/\sigma} \quad (-\infty, +\infty)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (-\infty, +\infty)$$

$$f(x) = \frac{1}{2\sigma\sqrt{3}} \quad (-\sqrt{3}\sigma, +\sqrt{3}\sigma)$$

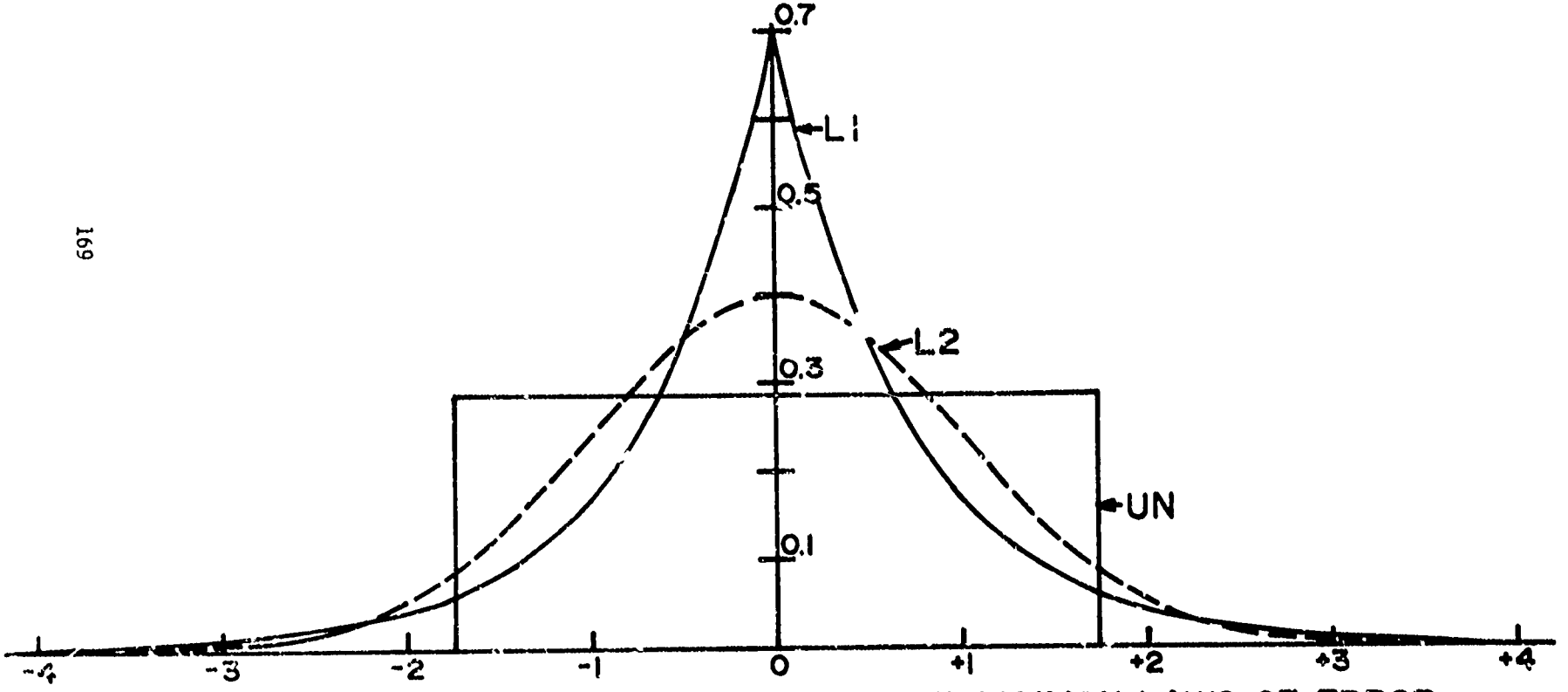


FIG 1 PROBABILITY DENSITY FUNCTIONS FOR THREE COMMON LAWS OF ERROR STANDARDIZED (MEAN=0, STANDARD DEVIATION=1, AREA UNDER CURVE=1)

Error Law*	α_4	AV*	DI*	LR(or NR)*
L3	6.0	MD	AD**	LF
L2	3.0	AM	SD	LS
UN	1.8	MR	SR	MM

*See glossary of code letters following list of references.

**Taken from the median, not from the arithmetic mean.

This conclusion follows from the theory of power means, as developed by Fechner (1874), Bruen (1938) and others. The results are corroborated by the following ratios of variances and asymptotic variances of median, mean, and midrange for the specified distributions:

$$\text{L1:}\quad \frac{\text{A Var (MD)}}{\text{A Var (AM)}} = \frac{1}{2}\text{;}\quad \text{L2:}\quad \frac{\text{A Var (MD)}}{\text{A Var (AM)}} = \frac{\pi}{2}\text{;}\quad \text{UN:}\quad \frac{\text{Var (MR)}}{\text{Var (AM)}} = \frac{6N}{(N+1)(N+2)}$$

2. For any other member of the exponential family of error laws, with probability density function of the form $f(x) = ce^{-a[|x-\mu|/\sigma]^p}$, the best choice of measure of central tendency is the power mean of order p, with which are associated the measure of dispersion which is the p^{th} root of the mean of the absolute p^{th} powers of deviations from the power mean of order p and the method of fitting the regression equation which minimizes that measure of dispersion of the residuals. When p is not an integer or is an integer greater than 2, this does not lead to simple procedures.

3. For error laws which are not members of the exponential family, it is not clear what the best choices are.

4. For the sake of simplicity, it is recommended that the choice be restricted to the three sets in paragraph (1) above. For symmetric distributions (whose standardized third moment about the mean, $\alpha_3 = \mu_3/\sigma^3$, is equal to zero), the value

of the standardized fourth moment about the mean, $\alpha_4=\mu_4/\sigma^4$, can be used as a criterion to decide which of the three sets of choices should be made. For α_4 near 3.0, the value for the normal distribution, one would obviously want to make the choices which are associated with that distribution. For those with substantially higher values of α_4, one would expect the choices associated with Laplace's first distribution (for which $\alpha_4=6.0$) to give better results. Similarly, for those with substantially lower values of α_4, one would expect the choices associated with the uniform distribution (for which $\alpha_4=1.8$) to give better results. On the basis of theoretical results of Rider (1957) and empirical results of Hogg (1967), it is recommended that the first set of choices in paragraph (1) above be made for $\alpha_4>3.8$, the second set for $2.2 \leq \alpha_4 \leq 3.8$, and third set for $\alpha_4 < 2.2$. If, as will usually be the case, the population value of α_4 is unknown, the corresponding sample value should be used.

5. All three of the above procedures are adversely affected by asymmetry in the distributions of errors or residuals, which occurs if positive and negative errors of the same magnitude are not equally likely. If the deviations from the chosen measure of central tendency in a one-dimensional array or the deviations from linear or nonlinear regression indicate asymmetry, as evidenced by a sample value of a_3 which differs significantly from zero (the standard error of a_3 is $\sqrt{6/n}$), consideration should be given to transforming the data so as to reduce the asymmetry as much as possible, and then analyzing the transformed data instead of the original data.

6. All three of the above procedures are also adversely affected by the presence of spurious observations which may have resulted from gross blunders or some undetected change in the quantity measured or in the conditions of

measurement. The adverse effect is most pronounced in the case of the procedure which is appropriate for the uniform distribution, which depends heavily on the extreme observations, and least so for the procedure which is appropriate for Laplace's first distribution, with the procedure appropriate for the normal distribution occupying an intermediate position. If there is any reason to suspect the presence of spurious observations, the procedure based on the uniform distribution should never be used, and that based on the normal distribution should be used only after applying one of the modern criteria for the rejection of outliers, most of which are based on normal theory. Outliers at one extreme only may produce false indications of asymmetry which vanish when they are rejected. Outliers at both extremes are likely to yield high values of α_4, which would lead to use of the procedures based on Laplace's first distribution; after they have been rejected, procedures based on the normal distribution may be appropriate. In such cases, it is often difficult to decide whether the extreme observations are spurious or whether they are genuine observations from a distribution (such as Laplace's first) with a high value of α_4.

7. Much can be learned by plotting the data. In the case of a one-dimensional array, one can form a preliminary impression as to skewness (as measured by α_3), kurtosis (as measured by α_4), and the presence of spurious observations. In the case of two variables, one can form a preliminary impression as to whether the relation is linear or nonlinear; if the latter, one can get some idea as to the type of curvilinear relation that should be fitted. After the regression equation has been fitted, the residuals should be plotted against the independent variable. The presence of any systematic pattern may indicate that the wrong type of relation has been fitted. Mention should be made of two systematic patterns

that may occur: (1) A correlation between the independent variable X and the magnitude of the residuals $|Y-\hat{Y}|$ may indicate the need to transform the data before analysis and find the regression on X, not of Y, but of log Y, Y^p, or Y/X; and (2) If the residuals tend to be of one sign for extreme values of the independent variable and of the opposite sign for intermediate values, this may indicate the need to fit a curvilinear instead of a linear relation. If no such pattern exists, the residuals may then be treated as a univariate array and analyzed accordingly (for skewness significantly different from zero, kurtosis for which the procedure used is inappropriate, or the presence of spurious observations which may not have been detected on the initial two-dimensional plot). If any of these conditions is discovered, appropriate steps to alleviate it can be taken and the resulting data reanalyzed.

8. If a function $y = f(x)$ has been computed (or measured) quite accurately and rounded values have been tabulated, the distribution of errors in the tabular values is uniform between -0.5 and + 0.5, in units of the last digit retained. This should be borne in mind in approximating the tabular values by a linear or nonlinear regression equation.

References

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 diticnem, et synopsis amplioris operis, ac habèntur 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 t 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 adnotationibus 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)

Lambert, J. H. (1765a). 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. (1765b). 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 Oeuvres 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 Oeuvres 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 177', 227-332(esp. 322-332). [Reprinted

in Oeuvres 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 Oeuvres 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 Oeuvres 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 Oeuvres 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,MI,TO)

Delambre, J. B. (editor)(1806-10). Base du Systeme 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 Oeuvres 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. (1811a). 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 Oeuvres Complètes de Laplace, Vol. 12, pp. 357-412. Gauthier-Villars, Paris, 1898]. (TE,AV,LR,AM,LS,EA)

Laplace, P. S. (1811b). Du milieu qu'il faut choisir entre les résultats d'un grand nombre des observations. Connaissance des Tems for 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 Oeuvres 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)

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 Anneé 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 Genauigkeit 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 (1818a). 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 (1818b). 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, 95-100; 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 à deux éléments détermines 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) for 1824, 92-99. (TE,AV,LR,MM)

Poisson, S. D. (1824,1829). Sur la probabilité des résultats moyens des observations. Connaissance des Tems for 1827, 273-302; for 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. (1830a). 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. (1830b). Verallgemeinerung der Poisson'schen Untersuchungen über die Wahrscheinlichkeit der mittleren [illegible]ltate der Beobachtungen in den "Additions à la Connaissance des Temps [illegible] 327." 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)

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

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 über 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 Dreiecksketten. Berlin. [Reprinted in part in Abhandlungen von Friedrich 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 der 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 on 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. 57-62]. (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,EC)

Bienaymé, Jules (1853a). 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,OM)

Bienaymé, Jules (1853b). 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,OM)

Cauchy, Augustin (1853a). 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,OM)

Cauchy, Augustin (1853b). 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,OM)

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

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

Cauchy, Augustin (1853e). 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,OM)

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

Cauchy, Augustin (1853g). 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,OM,MM)

Bertrand, J. (1855). Sur la méthode des moindres carrés. (Lettre à M. Élie 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 ubekjendt Störrelse ved givne umiddelbare 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-[illegible], tables, 593-599. [Appendix and related tables reprinted 1868 as A Treatise on the Method of Least Squares]. (TE,LR,LS,TO,PC,CC)

DeMorgan, 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 Genauigkeit 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)

Stone, E.J. (1873a). 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. (1873b). 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)

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. (1883a). The law of error. Philosophical Magazine (5) 16, 300-309. (TE,AV,LR,LS)

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

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)

Edgeworth, F. Y. (1885). Methods of statistics. Journal of the Royal Statistical Society Jubilee, 181-217. (TE,AV,DI,AM,MD,AD,QD)

Edgeworth, F. Y. (1886). Observations and statistics: an essay on the theory of errors of observation and the first principles of statistics. Transactions of the Cambridge Philosophical Society 14(2), 138-169; abstract, Proceedings of the Cambridge Philosophical Society 5(4)[1885], 310-312. (TE, AV,DI,LR,AM,MD,MO,QD,LF,EA)

Newcomb, Simon (1886). A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics 8, 343-366. (TE,TO,PC,NM)

Edgeworth, F. Y. (1887a). On discordant observations. Philosophical Magazine (5) 23, 364-375. (TE,TO,EM,SC)

Edgeworth, F. Y. (1887b). On observations relating to several quantities. Hermathena 6(13), 279-285. (TE,AV,LR,AM,MD,LS)

Edgeworth, F. Y. (1887c). A new method of reducing observations relating to to several quantities. Philosophical Magazine (5)24, 222-223. (TE,AV,LR,AM,MD, LF)

Edgeworth, F. Y. (1887d). The choice of means. Philosophical Magazine (5) 24, 268-271.(TE,AV,AM,MD)

Edgeworth, F. Y. (1887e). The empirical proof of the law of error. Philosophical Magazine (5) 24, 330-342. (TE,AV,AM,MD)

Turner, H. H. (1887). On Mr. Edgeworth's method of reducing observations relating to several quantities. Philosophical Magazine (5) 24, 466-470. (TE,AV,LR,AM, MD,LF,LS)

Bertrand, J. (1888a). Sur la loi de probabilité des erreurs d'observation. Comptes Rendus de l'Académie des Sciences (Paris) 106, 153-156. (TE,AV,AM, GM,HM,TO)

Bertrand, J. (1888b). Sur la combinaison des mesures d'une même grandeur. C. R. Acad. Sci. Paris 106, 701-704. (TE,AV,TO,AM,BC)

Edgeworth, F. Y. (1888). On a new method of reducing observations relating to several quantities. Philosophical Magazine (5) 25, 184-191. (TE,AV,LR,AM, MD,LF,LS)

Faye, H. E. (1888). Sur certain points de la théorie des erreurs accidentelles. C. R. Acad. Sci. Paris 106, 783-786. (TE,AV,TO,AM,MR,QM)

Galton, Francis (1888). Co-relations and their measurement, chiefly from anthropometric data. Proceedings of the Royal Society of London 45, 135-145. (TE,AV,DI,MD,QD)

Bertrand, J. (1889). Calcul des Probabilités. Gauthier-Villars, Paris. [esp. art. 166-169]. (TE,TO,BC)

Galton, Francis (1889). Natural Inheritance. Macmillan and Company, London. [esp. Chs. 4 and 5]. (TE,AV,DI,AM,MD,QD)

Czuber, E. (1890). Bemerkung über die wahrscheinlichsten Werte beobachteter Grössen. Archiv der Mathematik und Physik (2)9, 97-101. (TE,ML)

Estienne, J. E. (1890). Étude sur les erreurs d'observation. Revue d'Artillerie 36, 235-259. (TE,AV,LR,MD,LF)

Czuber, Emanuel (1891a). Theorie der Beobachtungsfehler. B. G. Teubner,Leipzig. (TE,TO,PC,CC,SC,GC)

Czuber, Emanuel (1891b). Über ein Ausgleichungsprincip. Technische Blätter 23, 1-9. (TE,AV,DI,LR,AM,MD,LS,LF)

Pizzetti, P. (1892). I fondamenti matematici per la critica dei risultati sperimentali. Atti della Regia Università de Genova 11, 113-333. (All)

Edgeworth, F. Y. (1893). Exercises in the calculation of errors. Philosophical Magazine 36, 98-111. (TE,AV,MD,AM,MQ)

Goedseels, ..; Mansion, P. (1893). Discussion sur la théorie des erreurs. Annales de la Société Scientifique de Bruxelles 17(1), 52-53. (TE,AV,LR,LS)

Pearson, Karl (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London (A) 186,343-414. [Reprinted in Karl Pearson's Early Statistical Papers (Cambridge Univ. Press, 1948), pp. 41-112]. (TE, AV,AM,MD,MO)

Poincaré, Henri (1896). Calcul des Probabilités. Gauthier-Villars, Paris. Second edition, 1912. (TE,AV,DI,AM,LR,LS,TO)

Fechner, G. Th. (1897). Kollectivmasslehre. Wilhelm Engelmann, Leipzig. (TE, AV,DI,AM,MD,MO,MR,EX,RA,LD)

Czuber, Emanuel (1899). Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. Jahresbericht der Deutschen Mathematiker Vereinigung 7 (2), 1-279. (All)

Goedseels, E. (1900). Étude sur la methode de Tobie Mayer. Annales de la Société Scientifique de Bruxelles 24(2), 37-58; rapport (par P. Mansion), 24 (1), 85-88. (TE,AV,DI,LR,MA,LS,CM)

Markoff, A. A. (1900). Calculus of Probabilites (Russian). St. Petersburg. Second edition, 1908; German translation, B. G. Teubner, Leipzig, 1912. (TE,AV,LR,LS)

Goedseels, E. (1901). Sur la méthode de Cauchy. Annales de la Société Scientifique de Bruxelles 25(1), 99-102,145-149. (TE,AV,LR,CM)

Goedseels, E. (1902). Sur l'application de la méthode de Cauchy aux moindres carrés. Annales de la Société Scientifique de Bruxelles 26(2), 148-156. (TE,AV,LR,CM,LS)

Czuber, Emanuel (1903). Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung, Volume I. B. G. Teubner, Leipzig-Berlin. Second edition, 1908. (TE,DI,AD,SD)

Kapteyn, J. C. (1903). Skew Frequency Curves in Biology and Statistics. Noordhoff, Groningen. Second edition (with M. J. van Uven), Hoitsema Bros., Groningen, 1916. (TE,AV,AM,MD,MO)

Saunder, S. A. (1903). Note on the use of Peirce's criterion for the rejection of doubtful observations. Monthly Notices of the Royal Astronomical Society 63, 432-436. (TE,TO,PC)

Mansion, Paul (1906). Sur la méthode des moindres carrés dans le Nachlass de Gauss. Annales de la Société Scientifique de Bruxelles 30(1), 169-174 (TE,AV, LR,LS,MM)

Wright, T. W; Hayford, J. F. (1906). Adjustment of Observations. D. van Nostrand, New York.[esp. pp. 87-92]. (TE,TO,WH,PC,CC,SC)

Goedseels, P.J.E.(1909). Théorie des Erreurs d'Observation. Charles Peeters, Louvain and Gauthier-Villars, Paris. [Third edition]. (TE,AV,LR,MR,AM,PA, MM,LS,MA,CM)

de la Vallée Poussin, Ch. J. (1909). L'approximation minimum d'un système d'équations (résume). Annales de la Société Scientifique de Bruxelles 32 (1), 173. (TE,AV,LR,MM)

Goedseels, P.J.Ed. (1910). Application de la théorie des erreurs de l'auteur à la compensation des coordonnées des sommets dans les levés topographiques. Annales de la Société Scientifique de Bruxelles 34(2), 257-287. (TE,AV,LR, PA,MM,LS)

Goedseels,P.J.E. (1911). Simplifications de la méthode la plus approximative et de l'approximation minima. Étalonnage des lunettes stadimétriques. Annales de la Société Scientifique de Bruxelles 35(1),351-368. (TE,AV,LR,PA, MM)

de la Vallée Poussin, Ch. J. (1911). Sur la méthode de l'approximation minimum. Annales de la Société Scientifique de Bruxelles 35(2), 1-16. (TE,AV,LR,MM)

Yule, G. Udny (1911). An Introduction to the Theory of Statistics. Charles Griffin and Co., Ltd., London. Fourteenth edition (with M. G. Kendall), 1950. (TE,AV,DI,AM,MD,MO,RA,AD,SD,QD)

Tits, L. (1912). Nouvelle simplification de la méthode la plus approximative et de la méthode de l'approximation minimum. Annales de la Société Scientifique de Bruxelles 36(2), 253-263. (TE,AV,LR,PA,MM)

Dodd, E. L. (1913). The probability of the arithmetic mean compared with that of certain other functions of the measurements. Annals of Mathematics 14, 186-198. (TE,AV,AM,MD,QA,GM)

Edgeworth, F. Y. (1913). On the use of the theory of probabilities in statistics relative to society. Journal of the Royal Statistical Society 76(2), 165-193. (TE,AV,DI,AM,MD,MO,AD)

Goodwin, H. M. (1913). Elements of the Precision of Measurements and Graphical Methods. McGraw-Hill Book Co., New York. (TE,AV,LR,AM,LS,TO,PC,CC,GR)

Mansion, Paul (1913). Sur les recherches de Laplace relatives à la théorie des erreurs. Annales de la Société Scientifique de Bruxelles 37, 107-117. (TE,AV, LR,MM,LF,LS)

Brunt, David (1917). The Combination of Observations. Cambridge University Press, Cambridge, England. Second edition, 1931. (TE,AV,LR,AM,MD,MO,LS,LF, TO,PC,CC,WH)

Persons, Warren M. (1919). Indices of business conditions. Review of Economic Statistics 1, 5-107. (TE,AV,AM,MD,MO,GM)

Daniell, P. J. (1920). Observations weighted according to order. American Journal of Mathematics 42, 222-236. (TE,AV,DI,AM,MD,MQ,DA,SD,AD,QD,DD)

Pearson, Karl (1920). On the probable errors of frequency constants, Part III. Biometrika 13, 113-132. (TE,AV,DI,AM,MD,QN)

Stewart, R. M. (1920a). Peirce's criterion. Popular Astronomy 28, 2-3. (TE,TO, PC,CC)

Stewart, R. M. (1920b). The treatment of discordant observations. Popular Astronomy 28, 4-6. (TE,TO,SM)

Jackson, Dunham(1921). Note on the median of a set of numbers. Bulletin of the American Mathematical Society 27, 160-164. (TE,AV,LR,AM,MD,MR,LS,LF,MM)

Dodd, E. L. (1922). Functions of measurement under general laws of error. Skandinavisk Aktuarietidskrift 5, 133-158. (TE,AV,AM,MD,MR)

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London (A) 222, 309-368. (TE,TO,AV,MD,ML)

Foster, R. M. (1922). Derivation of Mode as Special Case of Generalized Mean. Personal communication; result cited by E. V. Huntington, Mathematical Memoranda, Handbook of Mathematical Statistics (ed. H. L. Rietz), p. 7. Boston, 1924. (TE,PM,MO)

Crum, W. L. (1923). The use of the median in determining seasonal variation. Journal of the American Statistical Association 18, 607-614. (TE,AV,AM,MD)

Edgeworth, F. Y. (1923). On the use of medians for reducing observations relating to several quantities. Philosophical Magazine (6)46, 1074-1088. (TE,AV,LR,AM,MD,LS,LF)

Wilson, E. B. (1923). First and second laws of error. Journal of the American Statistical Association 18, 841-851. (TE,AV,AM,MD)

Haag, J. (1924). Sur la combinaison des résultats d'observations. Comptes Rendus de l'Academie des Sciences (Paris) 179, 1388-1390. (TE,AV,MD,AM,QM)

Jackson, Dunham (1924). On the method of least m^{th} powers for a set of simultaneous equations. Annals of Mathematics (2)25, 185-192; abstract, Bulletin of the American Mathematical Society 27(1921),304. (TE,AV,LR,LS,LF,MM)

Whittaker, E. T.; Robinson, G. (1924). The Calculus of Observations. Blackie & Son, Ltd., London-Glasgow. (TE,AV,LR,AM,MD,LS)

Coolidge, J. L. (1925). An Introduction to Mathematical Probability. Oxford University Press, Oxford, England. [Reprinted 1962 by Dover Publications, Inc., New York.] (TE,AV,LR,AM,MD,LS,TO,PC,CC,SC,GC)

Irvin, J. O. (1925a). The further theory of Francis Galton's individual difference problem. Biometrika 17, 100-128. (TE,AV,DI,OS)

Irvin, J. O. (1925b). On a criterion for the rejection of outlying observations. Biometrika 17, 238-250. (TE,TO,IC)

Lévy, Paul (1925). Calcul des Probabilités. Gauthier-Villars, Paris. (TE,LR,LS)

Reilly, J.; Rae, W. N.; Wheeler, T. S. (1925). Physico-Chemical Methods. D. van Nostrand, New York, (TE,LR,LS,TO,MS)

Tippett, L. H. C.(1925). On the extreme individuals and the range of samples taken from a normal population. Biometrika 17, 364-387. (TE,DI,RA,TO,TC)

Estienne, J. E. (1926-27). Introduction à une théorie rationnelle des erreurs d'observation. Revue d'Artillerie 97(1926), 421-441; 98(1926),542-562; 100 (1927), 471-487. (TE,AV,LR,LS,LF)

Pearson, Egon S. (1926). A further note on the distribution of range in samples taken from a normal population. Biometrika 18, 173-194. (TE,AV,DI,RA)

"Student" (1927). Errors of routine analysis. Biometrika 19, 151-164. (TE, DI,RA,TO,IC,ST)

Bowley, A. L. (1928). F. Y. Edgeworth's Contributions to Mathematical Statistics. Royal Statistical Society, London. (TE,AV,LR,AM,MD,LS, LF,TO,EM,S2)

*Wilson, E. B.; Hilferty, M. M. (1929). Note on C. S. Peirce's experimental discussion of the laws of error. Proceedings of the National Academy of Sciences 15(2), 120-125. (TE,AV,AM,MD)

*Neyman, J.; Pearson, E. S. (1928).
[See Additional References at end of list].

Rhodes, E. C. (1930). Reducing observations by the method of minimum deviations. Philosophical Magazine (7) 9, 974-992. (TE,AV,LR,MR,AM,MD,LS,LF)

Hojo, Tokishige (1931). Distribution of the median, quartiles, and interquartile distance in samples from a normal population. Biometrika 23, 315-360; appendix by Karl Pearson, 361-363. (TE,AV,DI,MD,QN,ID)

von Mises, Richard (1931). Vorlesungen aus dem Gebiete der Angewandten Mathematik. I. Band. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik. Franz Deuticke, Leipzig-Wien. (TE,AV, LR,MD,QN,LS)

Pearson, Karl (editor)(1931). Tables for Statisticians and Biometricians, Part II. Biometric Laboratory, University College, London. (TE,AV,DI,MR,MD,RA, TO,CC,IC)

Shewhart, W. A. (1931). The Economic Control of Quality of Manufactured Product. D. van Nostrand Co., Inc., New York. (TE,AV,DI,AM,MD,MR,SD,RA)

Craig, Allen T. (1932a). On the distribution of certain statistics. American Journal of Mathematics 54, 353-366. (TE,AV,DI,AM,HM,GM,MD,QN,RA)

Craig, Allen T. (1932b). Distribution of functions of middle items (abstract). Bulletin of the American Mathematical Society 38, 313. (TE,AV,DI,AM,MR,RA)

Jeffreys, Harold (1932). An alternative to the rejection of observations. Proceedings of the Royal Society of London (A) 137, 78-87. (TE,TO,JA)

Luyten, William J. (1932). Notes on stellar statistics. V: On the use of the first Laplacean error curve. Proceedings of the National Academy of Sciences 18, 360-365. (TE,AV,DI,AM,MD,SD,QD,AD)

Crowe, P. R. (1933). The analysis of rainfall probability. A graphical method and its application to European data. Scottish Geographical Magazine 49, 73-91. (TE,AV,AM,MD,MO,SD,AD,QD)

*Pearson, E. S. (1932).
[See Additional References at end of list].

and its application to European data. Scottish Geographical Magazine 49, 73-91. (TE,AV,AM,MD,MD,SD,AD,QD)

*Rider, Paul R. (1933). Criteria for Rejection of Observations. Washington University Studies-New Series, Science and Technology-No. 8. (TE,TO,PC,CC, SC,S2,EM,IC,SM,GC,TC,JA,AV,MD)

Hunsaker, Enos (1934). Uber die Bewertung der Potenzmomente. Monatshefte für Mathematik und Physik 41, 375-383. (TE,AV,DI,AM,MD,AD,AD,PM)

Pollard, Henry S. (1934). On the relstive stability of the median and arithmetric mean, with particular reference to certain frequency distributions which can be dissected into normal distributions. Annals of Mathematical Statistics 5, 227-262. (TE,AV,AM,MD)

Fréchet, Maurice (1935). Sur les précisions comparées de la moyenne et de la médiane. Aktuárské Vědy 5(3), 29-34. (TE,AV,AM,MD)

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. Biometrika 27, 310-332. (TE,DI,SD,AD)

**Pearson, E. S. (1935). A comparison of β_2 and Mr. Geary's w_n criteria. Biometrika 27, 335-352. (TE,DI,SD,AD)

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of the deviation to the sample standard deviation. Annals of Mathematical Statistics 6, 214-219. (TE,TO,TC,TM)

Darmois, Georges (1936). L'emploi des observations statistiques. Méthodes d'estimation. Actualités Scientifiquet et Industrielles, No. 356. Herman et Cie., Paris. (TE,AV,AM,MD,LR,LF,LS)

* McKay, A. T.; Pearson, E. S. (1933). [See Additional References at end of list].

** McKay, A. T. (1935). [See Additional References at end of list].

Pearson, E. S.; Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika 28, 308-319; appendix by J.M.C. Scott, 319-320. (TE,TO,TM)

Gumbel,E. J. (1937). La precision de la moyenne arithmétique et de la médiane. Aktuárské Vědy 6(4), 145-154. (TE,AV,AM,MD)

Bruen, Curtis (1938). Methods for the combination of observations: Modal point or most lesser-deviations, median loci or least deviation, mean loci or least squares, and midpoint of least range or least greatest-deviation. Metron 13 (2), 61-140. (TE,AV,MO,MD,AM,MR,LR,NR,LN,LF,LS,MM)

Dodd, E. L. (1938). Definitions and properties of the median, quartiles, and other positional means. American Mathematical Monthly 45, 302-306; note by E.J.M., 462. (TE,AV,MD,QN)

Souto, Jose Barral (1938). El modo y otras medias, casos particulares de una misma expression matematica. (English summary). Boletin Matematico (Buenos Aires) 11 (3), 29-41. (TE,AV,DI,AM,MD,MO,HM,GM,QA,AD,SD,SR)

Gumbel, Emil J. (1939). Les valeurs de position d'une variable aléatoire. Comptes Rendus de l'Académie des Sciences (Paris) 208, 149-151. (TE,AV,MD, QN)

Jeffreys, Harold (1939). Theory of Probability. Oxford University Press, London. (TE,AV,DI,AM,MD,AD,SD,TO,PC,CC,JA)

Arley, Niels (1940). On the distribution of relative errors from a normal population of errors. A discussion of some problems in the theory of errors.

Mathematisk-Fysiske Meddelelser udgivet af det Kgl. Danske Videnskabernes Selskab 18(3), 62 pp. (TE,LR,TO,AC,TM,CC,PC,SC,VC,HC,MZ,RC,TJ)

Fréchet, Maurice (1940a). Sur une limitation très générale de la dispersion de la médiane. Journal de la Société Statistique de Paris 81, 67-76; discussion, 76-78. (TE,AV,DI,AM,MD,AD,SD,QD)

Fréchet, Maurice (1940b). Comparaison des diverses mesures de la dispersion. Revue de l'Institut International de Statistique 8, 1-12. (TE,DI,AD,SD,QD)

Paulson, Edward (1940). A note on the distribution of the median. National Mathematics Magazine 14, 379-382. (TE,AV,AM,MD)

Singleton, Robert R. (1940). A method for minimizing the sum of absolute values of deviations. Annals of Mathematical Statistics 11, 301-310. (TE,AV,LR,NR,MD,LF)

Wald, Abraham (1940). The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics 11, 284-300. (TE,LR,OS)

Wilson, Edwin B. (1940). The sampling error of the median. Science 92, 58-59. (TE,AV,AM,MD,DI,SD)

Brookner, R. J. (1941). A note on the mean as a poor estimate of central tendency. Journal of the American Statistical Association 36, 410-412. (TE,AV,AM,MR,DI,SD)

Fréchet, Maurice (1941). Sur la correspondance entre certaines lois d'erreurs et certaines définitions de la distance. Revue Scientifique (Paris) 79, 3-14. (TE,AV,DI,AM,MD,SD,AD,QN)

Hartley, H. O. (1942). The range in random samples. Biometrika 32, 334-348. (TE,DI,RA)

*Pearson, E. S.; Hartley, H. O. (1942). The probability integral of the range in

* Nair, K. R.; Shrivastava, M. P. (1942). [See Additional References at end of list].

samples of n observations from a normal population. Biometrika 32, 301-310. (TE,DI,JA)

Barnard, G. A. (1943). The Use of the Median in Place of the Mean in Quality Control Charts. Technical Report No. Q.C./R/2. (British)Ministry of Supply Advisory Service on Quality Control. (TE,AV,AM,MD)

Wilks, S. S. (1943). Mathematical Statistics. Princeton University Press, Princeton, N. J.; revised and enlarged edition, John Wiley & Sons, Inc., New York, 1962. (TE,AV,DI,MD,RA,OS)

Geary, R. C. (1944). Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter. Biometrika 33, 123-128. (TE,AV,AM,MR, MD)

Gumbel, E. J. (1944). Ranges and midranges. Annals of Mathematical Statistics 15, 414-422. (TE,AV,DI,MR,QN,RA,QR,SR)

Robbins, Herbert (1944). On the expected value of two statistics. Annals of Mathematical Statistics 15, 321-323. (TE,DI,RA,OS)

Pearson, E. S.; Godwin, H. J.; Hartley, H. O. (1945). The probability integral of the mean deviation. Biometrika 33, 252-265. (TE,DI,AD)

Baker, G. A. (1946). Distribution of the ratio of sample range to sample standard deviation for normal and combinations of normal distributions. Annals of Mathematical Statistics 17, 368-369. (TE,DI,RA,SD)

Brown, George W.; Tukey, John W. (1946). Some distributions of sample means. Annals of Mathematical Statistics 17, 1-12 (TE,AV,AM,MD)

Carlton, A. George (1946). Estimating the parameters of a rectangular distribution. Annals of Mathematical Statistics 17, 355-358. (TE,AV,DI,MR,AM,MD,RA)

Cramér, Harald (1946). Mathematical Methods of Statistics. Princeton University

Press, Princeton, New Jersey (TE,AV,DI,AM,MD,MD,SD,RA,QN,LR,NR,LS,LF,ML)

*Gumbel, E. J. (1946). On the independence of the extremes in a sample. Annals of Mathematical Statistics 17, 78-81. (TE,OS,EX)

Kendall, Maurice G. (1946). The Advanced Theory of Statistics, Volume II. Charles Griffin & Co. Ltd., London (TE,LR,NR,ML,LS)

Mosteller, Frederick (1946). On some useful "inefficient" statistics. Annals of Mathematical Statistics 17, 377-408. (TE,AV,DI,AM,MD,AD,SD,RA,OS,QR)

Grubbs, Frank E.; Weaver, Chalmers L. (1947). The best unbiased estimate of population standard deviation based on group ranges. Journal of the American Statistical Association 42, 224-241. (TE,DI,RA)

**Nair, K. R. (1947). A note on the mean deviation from the median. Biometrika 34, 360-362. (TE,AV,DI,AM,MD,AD)

Plackett, R. L. (1947). Limits of the ratio of mean range to standard deviation. Biometrika 34, 120-122. (TE,DI,RA,SD)

Purcell, Warren B. (1947). Saving time in testing life. Industrial Quality Control 3(5), 15-18. (TE,AV,DI,AM,MD,RA,OS)

Tanenhaus, Seaman J. (1947). The median as a typical value for Walker yarn abrader data. Textile Research Journal 17, 281-286. (TE,AV,AM,MD,MD,OS)

Eisenhart, Churchill; Deming, Lola S.; Martin, Celia S. (1948a). The probability points of the distribution of the median in random samples from any continuous population (abstract). Annals of Mathematical Statistics 19, 598-599. (TE,AV,MD)

Eisenhart, Churchill; Deming, Lola S.; Martin, Celia S. (1948b). On the arithmetic mean and the median in small samples from the normal and certain non-normal populations (abstract). Annals of Mathematical Statistics 19, 599-600. (TE,AV,AM,MD)

*Daly, Joseph F. (1946). [See Additional References at end of list].
**Lord, E. (1947). [See Additional References at end of list].

Housner, G. W.; Brennan, J. F. (1948). The estimation of linear trends. Annals of Mathematical Statistics 19, 380-388. (TE,AV,AM,LR,LS,LF)

Nair, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form. Biometrika 35, 118-144. (TE,OS,ED,IC,TC,ST,MC, TM,NC)

Pillai, K.C.S. (1948). A note on ordered samples. Sankhyā 8, 375-380. (TE, AV,DI,AM,SD,SR)

Harris, Theodore E.; Tukey, John W. (1949). Measures of Location and Scale which are Relatively Insensitive to Contamination. Memorandum Report No. 31, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE,AV,DI,AM,RL,DA,MM,SD,DD)

Nair, K. R. (1949). A further note on the mean deviation from the median. Biometrika 36, 234-235. (TE,AV,DI,MD,AD)

Purcell, W. R. (1949). Saving time in testing life. Electrical Engineering (New York) 68,617-620. (TE,AV,DI,MD,OS,RA)

Shone, K. J. (1949). Relations between the standard deviation and the distribution of range in non-normal populations. Journal of the Royal Statistical Society (B) 11, 85-88. (TE,DI,SD,RA)

Tukey, John W. (1949a). Scaling by and for Power-Means--The Asymptotic Case. Memorandum Report No. 25, Statistical Techniques Group, Princeton University, Princeton, New Jersey. (TE,AV,PM)

Tukey, John W. (1949b). The Truncated Mean in Moderately Large Samples. Memorandum Report No. 32, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE,AV,AM,RL,DA,MM)

Tukey, John W. (1949c). Scaling by and for Percentiles and Exponential Averages.

Memorandum Report No. 33, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE,AV,ML)

Tukey, John W. (1949d). Skeleton Tables Related to Contaminated Distributions at Scale 3. Memorandum Report No. 34, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE,AV,ML)

Youden, W. J. (1949). The fallacy of the best two out of three. National Bureau of Standards Technical News Bulletin 33, 77-78. (TE,TO,ST)

Dixon, W. J. (1950). Analysis of extreme values. Annals of Mathematical Statistics 21, 488-506. (TE,TO,IC,DM,DC)

Grubbs, Frank E. (1950). Sample criteria for testing outlying observations. Annals of Mathematical Statistics 21, 27-58. (TE,TO,IC,TC,ST,DM,GS)

Harris, T. E. (1950). Regression using minimum absolute deviations (Answer to question 25). The American Statistician 4 (1), 14-15. (TE,LR,LF,LS)

Henry, F. M. (1950). The loss of precision from discarding discrepant data. Research Quarterly of the American Association for Health, Physical Education, and Recreation 21 (2), 145-152. (TE,TO,GR)

Pearson, E. S. (1950). Some notes on the use of range. Biometrika 37, 88-92. (TE,DI,RA,TO)

Pillai, K.C.S. (1950). On the distributions of midrange and semi-range in samples from a normal population. Annals of Mathematical Statistics 21, 100-105. (TE,AV,DI,MR,SR)

Seth, G. R. (1950). On the distribution of the two closest among a set of three observations. Annals of Mathematical Statistics 21, 298-301. (TE,TO, BT)

Zeigler, R. K. (1950). A note on the asymptotic simultaneous distribution of the sample median and the mean deviation from the sample median. Annals of

Mathematical Statistics 21, 452-455. (TE,AV,DI,MD,AD)

Bhate, D. H. (1952). A note on the estimates of centre of location of symmetrical population. Calcutta Statistical Association Bulletin 4(13), 33-35. (TE,AV,QM,MD)

Brown, G. W.; Mood, A. M. (1951). On median tests for linear hypotheses. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California-Berkeley, 1950)(ed. Jerzy Neyman), pp. 159-166. University of California Press, Berkeley-Los Angeles. (TE,AV,MD,LR, OS)

Dixon, W. J. (1951). Ratios involving extreme values. Annals of Mathematical Statistics 22, 68-78. (TE,OS,TO,DC)

Hartley, H. O.; Pearson,E. S. (1951). Moment constants for the distribution of range in normal samples. Biometrika 38, 463-464. (TE,DI,RA)

Murphy, Ray Bradford (1951). On Tests for Outlying Observations. Ph.D. thesis, Princeton University. University Microfilms, Ann Arbor, Mich. (TE,TO,IC,TC, MK,TM,NC,GS,DC)

Cadwell, J. H. (1952). The distribution of quantiles of small samples. Biometrika 39, 207-211. (TE,AV,AM,MD,QN)

Hald, A. (1952a). Statistical Theory with Engineering Applications. John Wiley & Sons, Inc., New York. (TE,DI,RA,OS,TO)

Hald, A. (1952b). Statistical Tables and Formulae. John Wiley & Sons, Inc., New York. (TE,DI,RA)

Homma, Tsuruchiyo (1952). On the limit distributions of some ranges. Reports of Statistical Application Research, Union of Japanese Scientists and Engineers 1 (4), 15-26. (TE,AV,DI,RA,MR)

Johnson, N. L. (1952). Approximations to the probability integral of the distribution of range. Biometrika 39, 417-419. (TE,DI,RA)

Lieblein, Julius (1952). Properties of certain statistics involving the closest pair in a sample of three observations. Journal of Research of the National Bureau of Standards 48, 255-268. (TE,DI,RA,TO,ET)

Nair, K. R. (1952). Tables of percentage points of the "Studentized" extreme deviate from the sample mean. Biometrika 39, 189-191. (TE,EX,TO,NC)

Rao, C. Radhakrishna (1952). Advanced Statistical Methods in Biometic Research. John Wiley & Sons, Inc. New York. (TE,AV,DI,AM,MD,RA)

Cadwell, J. H. (1953). The distribution of quasi-ranges in samples from a normal population. Annals of Mathematical Statistics 24, 603-613. (TE,DI,QR)

Dixon, W. J. (1953). Processing data for outliers. Biometrics 9, 74-89. (TE, AV,DI,AM,MD,SD,RA,TO)

Ferrell, Enoch B. (1953). Control charts using midranges and medians. Industrial Quality Control 9 (5), 30-34. (TE,AV,DI,MD,MR,RA)

Harter, H. Leon (1953). Maximum likelihood regression equations (abstract). Annals of Mathematical Statistics 24, 687. (TE,LR,NR,ML,LS,LF)

King, E. P. (1953). On some procedures for the rejection of suspected data. Journal of the Roval Statistical Association 48, 531-533. (TE,TO,GS,DC)

Olds, E. G. (1953). On Regression Analysis when the Dependent Variable is Rectangular (Preliminary Report). Technical Report No. 40, prepared under Contract No. AF 33(616)-294, Department of Mathematics, Carnegie Institute of Technology. (TE,LR,ML,LS)

Proschan, Frank (1953). Rejection of outlying observations. American Journal

of Physics 21, 520-525. (TE,TO,NC,DC)

Yorden, W. J. (1953). Sets of three measurements. The Scientific Monthly 77 (3), 143-147. (TE,TO,BT)

Cadwell, J. H. (1954). The probability integral of range for samples from a symmetrical unimodal population. Annals of Mathematical Statistics 25, 803-806. (TE,DI,RA)

Cox, D. R. (1954). The mean and coefficient of variation of range in small samples from non-normal populations. Biometrika 41, 469-481. (TE,DI,RA)

David, H. A. (1954). The distribution of range in certain non-normal populations. Biometrika 41, 463-468. (TE,DI,RA)

David, H. A.; Hartley, H. O.; Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. Biometrika 41, 482-493. (TE,DI,RA,SD)

Gumbel, E. J. (1954). The maxima of the mean largest value and of the range. Annals of Mathematical Statistics 25, 76-84. (TE,DI,RA,OS,EX)

Hartley, H. O.; David, H. A. (1954). Universal bounds for mean range and extreme observation. Annals of Mathematical Statistics 25, 85-99. (TE,DI,RA,OS,EX)

Pearson, E. S.; Hartley, H. O. (editors) (1954). Biometrika Tables for Statisticians, Volume I. Cambridge University Press for the Biometrika Trustees, Cambridge, England. (Second edition, 1958; third edition, 1966). (TE,DI,RA,TO,NC)

Resnikoff, G. J. (1954). The Distribution of the Average-Range for Subgroups of Five. Technical Report No. 15, Applied Mathematics and Statistics Laboratories, Stanford University. AD 40426. (TE,DI,RA)

Zítek, F. (1954). On certain estimators of standard deviation. (Polish. Russian and English summaries). Zastosowania Matematyki 1, 342-353. (TE,DI,SD,AD,RA)

Bonferroni, Carlo (1955). I valori mediani in una distribuzione continua. Statistica (Bologna) 15, 3-22. (TE,AV,AM,MD)

Chu, John T. (1955a). On the distribution of the sample median. Annals of Mathematical Statistics 26, 112-116. (TE,AV,MD)

Chu, John T. (1955b). The "inefficiency" of the sample median for many familiar symmetric distributions. Biometrika 42, 520-521. (TE,AV,AM,MD)

Chu, John T.; Hotelling, Harold (1955). The moments of the sample median. Annals of Mathematical Statistics 26, 593-606. (TE,AV,MD)

Greenwood, J. Arthur (1955). Distribution of the m-th range (abstract). Annals of Mathematical Statistics 26, 772. (TE,DI,QR)

Halperin, Max; Greenhouse, Samuel W.; Cornfield, Jerome; Zalokar, Julia (1955) Tables of percentage points for the studentized maximum alsolute deviate in normal samples. Journal of the American Statistical Association 50, 185-195. (TE,OS,EX,TO,NC)

Thomson, George W. (1955). Bounds for the ratio of range to standard deviation. Biometrika 42, 268-269. (TE,DI,RA,SD)

Tukey, John W. (1955). Interpolations and approximations related to the normal range. Biometrika 42, 480-485. (TE,DI,RA)

Behrens, W. U. (1956). Zür abgekürzten Berechnung der Standardabweichung. Zeitschrift für Acker-und Pflanzenbau 101, 459-464. (TE,DI,SD,RA)

Bejar, Juan (1956). Regresión en mediana y la programación lineal. (English summary). Trabajos de Estadística 7, 141-158. (TE,AV,LR,NR,AM,MD,LS,LF)

Bliss, C. I.; Cochran, W. G.; Tukey, J. W. (1956). A rejection criterion based upon the range. Biometrika 43, 418-422. (TE,DI,RA,TO,CT)

David, H. A. (1956). Revised upper percentage points of the extreme studentized

deviate from the sample mean. Biometrika 43, 449-451. (TE,OS,EX,TO,NC)

Kudô, A. (1956). On the testing of outlying observations. Sankhyā 17, 67-76. (TE,AV,DI,AM,SD,TO,TM,NC,GS,KC)

Ruben, H. (1956). On the moments of the range and product moments of extreme order statistics in normal samples. Biometrika 43, 458-460. (TE,DI,RA,OS,EX)

Bejar, Juan (1957). Cálculo práctico de la regresión en mediana. (English summary). Trabajos de Estadística 8, 157-173. (TE,AV,MD,LR,LF).

Dixon, W. J. (1957). Estimates of the mean and standard deviation of a normal population. Annals of Mathematical Statistics 28, 806-809. (TE,AV,DI,AM,MD, MR,QM,DA,RA,QR)

Chosal, A. (1957). The distribution of quasi-ranges in samples from rectangular and exponential distributions. Journal of the Institute of Actuaries Students' Society 14, 94-101. (TE,DI,QR)

Harley, B. I.; Pearson, E. S. (1957). The distribution of range in normal samples with n=200. Biometrika 44, 257-260. (TE,DI,RA)

Masuyama, Motosaburo (1957). The use of sample range in estimating the standard deviation or the variance of any population. Sankhyā 18,159-162. (TE,DI,RA,SD)

Rider, Paul R. (1957). The midrange of a sample as an estimator of population midrange. Journal of the American Statistical Association 52,537-542. (TE,AV, AM,MR)

Sibuya,Masaaki; Toda, Hideo (1957). Tables of the probability density function of range in normal samples. Annals of the Institute of Statistical Mathematics (Tokyo) 8, 155-165. (TE,DI,RA)

Babcock, S.; Beck, A.; Davies, A.; Goldsmith, B.; Torkelson, E. [Subcommittee of Joint Electron Tube Engineering Council Committee on Sampling Procedures

(JETEC-11)] (1958). Acceptance sampling of lots by the median, quasi-range method. Industrial Quality Control 15 (1), 8-11. (TE,AV,DI,MD,QR)

Barton, D. E.; Casley, D. J. (1958). A quick estimate of the regression coefficient. Biometrika 45, 431-435. (TE,AV,AM,LR,OS)

Carlson, Phillip G. (1958). A recurrence formula for the mean range of odd sample sizes. Skandinavisk Aktuarietidskrift 41, 55-56. (TE,DI,RA)

Ferrell, Enoch B. (1958). Control charts for Log-Normal universes. Industrial Quality Control 15 (2), 4-6. (TE,AV,DI,AM,GE,GG)

Harter, H. Leon (1958). The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation. WADC Technical Report 58-200, Wright-Patterson AFB. AD 151200. (TE,DI,QR,RA,SD) [See also Harter (1970), Volume 2: Chapter 1, Section 1 and Tables A1-A5; Chapter 2, Section 1 and Table B1].

Ostle, B.; Wiesen, J.M. (1958). An acceptance sampling plan based on the distribution of the range when sampling from a triangular population. Industrial Quality Control 15 (3), 8-9. (TE,DI,RA)

Plackett, R. L. (1958). Studies in the history of probability and statistics. VII. The principle of the arithmetic mean. Biometrika 45, 130-135. (TE,AV, AM,DI,SR)

Rider, Paul R. (1958). Generalized Cauchy distributions. Annals of the Institute of Statistical Mathematics (Tokyo) 9, 215-223. (TE,AV,AM,MD)

Geffroy, Jean (1959). Contribution à la théorie des valeurs extremes. II. Publications de l'Institut de Statistique de l'Université de Paris 8, 123-185. (TE,AV,DI,MR,QM,RA,QR)

Gumbel, E. J. (1959). The m^{th} range. Journal de Mathématiques Pures et Appliquées (9) 38, 253-265. (TE,DI,RA,QR)

Hänsel, Hermann (1958). Die Verwendung der Variationsbreite bei der Schätzung der Standardabweichung und bei der Varianzanalyse ungeordneter Blockanlagen, sowie eine weitere Vereinfachung des Hartleyschen Verfahrens zur direkten Bestimmung von Grenzdifferenzen. Bodenkultur (A) 10, 148-158. (TE, DI,RA)

Harter, H. Leon (1959). The use of sample quasi-ranges in estimating population standard deviation. Annals of Mathematical Statistics 30, 980-999; correction, 31(1960), 228. (TE,DI,QR,RA) [See also Harter (1970), Volume 2: Chapter 1, Section 3 and Tables A1-A5; Chapter 2, Section 1 and Table B1].

Harter, H. Leon; Clemm, Donald S. (1959). The Probability Integrals of the Range and of the Studentized Range: Probability Integrals, Percentage Points, and Moments of the Range. WADC Technical Report 58-484, Volume I. Wright-Patterson Air Force Base. AD 215024. (TE,DI,RA) [See also Harter (1970), Volume 1: Chapter 1, Section 2 and Tables A6-A8].

Ostle, Bernard; Steck, George P. (1959). Correlation between sample means and sample ranges. Journal of the American Statistical Association 54, 465-471. (TE,AV,DI,AM,MR,RA)

Pillai, K.C.S. (1959). Upper percentage points of the extreme studentized deviate from the sample mean. Biometrika 46, 473-474. (TE,AV,AM,OS,EX,TO,NC)

Pillai, K.C.S.; Tienzo, Benjamin P. (1959). On the distribution of the extreme studentized deviate from the sample mean. Biometrika 46, 467-472. (TE,AV,AM, OS,EX,TO,NC)

Rider, Paul R. (1959). Quasi-ranges of samples from an exponential population. Annals of Mathematical Statistics 30, 252-254; corrections, 1266-1267. (TE, DI,RA,QR)

Nalsh, J. E. (1958). Large sample nonparametric rejection of outlying observations. Annals of the Institute of Statistical Mathematics (Tokyo) 10, 223-232. (TE,OS,TO,MM)

Neiler, H. (1959). Note on harmonic and geometric means. Australian Journal of Statistics 1, 44-46. (TE,AV,AM,GM,HM)

Anscombe, F. J. (1960). Rejection of outliers. Technometrics 2, 123-147; discussion, 157-166. (TE,AV,AM,MD,LR,LS,TO,PC,CC,GC,MC,BM,BC,ST,TM,JA,GS,CT,AS)

Dixon, W. J. (1960). Simplified estimation from censored normal samples. Annals of Mathematical Statistics 31, 385-391. (TE,AV,DI,Mi,MM,RA,QR,TO)

Harter, H. Leon (1960). Tables of range and studentized range. Annals of Mathematical Statistics 31, 1122-1147. (TE,DI,RA) [See also Harter (1970), Volume 1: Chapter 1, Section 2 and Tables 26-A8].

Hogg, Robert V. (1960a). Certain uncorrelated statistics. Journal of the American Statistical Association 55, 265-267. (TE,AV,DI,AM,MD,MR,SD,PO,QR,AD)

Hogg, Robert V. (1960b). On conditional expectations of location statistics. Journal of the American Statistical Association 55, 714-717. (TE,AV,DI,MD,RA)

Kruskal, William H. (1960). Some remarks on wild observations. Technometrics 2, 1-3. (TE,TO,TM,DC,GS,BT,CT)

Rider, Paul R. (1960a). Variance of the median of small samples from several special populations. Journal of the American Statistical Association 55, 148-150. (TE,AV,MD)

Rider, Paul R. (1960b). Variance of the median of samples from a Cauchy distribution. Journal of the American Statistical Association 55, 322-323. (TE,AV, MD)

Tukey, John W. (1960). A survey of sampling from contaminated distributions.

Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (ed. Ingram Olkin et al), pp. 448-485. Stanford University Press, Stanford, Calif. (TE,AV,DI,AM,DA,SD,AD,DC,QR)

Anscombe, F. J. (1961). Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, California, 1960) (ed. Jerzy Neyman), Volume I, pp. 1-36. University of California Press, Berkeley-Los Angeles. (TE,LR,LS,TO,AR)

Eisenhart, Churchill (1961). Boscovich and the combination of observations. Roger Joseph Boscovich, S. J., F.R.S., 1711-1787: Studies of his Life and Work on the 250th Anniversary of his Birth (ed. Lancelot Law White), pp. 200-212. Allen and Unwin, Ltd., London. (TE,AV,AM,MD,LR,LS,LF,MA,EA)

Ferguson, Thomas S. (1961a). On the rejection of outliers. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, California,1960) (ed. Jerzy Neyman), Volume I, pp. 253-287. University of California Press, Berkeley-Los Angeles. (TE,TO,DC,GR,KC,FC)

Ferguson, Thomas S. (1961b). Rules for rejection of outliers. Revue de l'Institut Internationale de Statistique 29(3), 29-43. (TE,TO,PC,CC,TM,DC, GR,CT,KC,MR,FC)

de Finetti, Bruno (1961). The Bayesian approach to the rejction of outliers. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, California, 1960) (ed. Jerzy Neyman), Volume I, 199-210. University of California Press, Berkeley-Los Angeles. (TE,TO,JA)

Harter, H. Leon (1961a). The Use of Sample Ranges and Quasi-Ranges in Setting Exact Confidence Bounds for the Population Standard Deviation. I. ARL 31, Part I. Wright-Patterson AFB. AD 260325. (TE,DI,RA) [See also Harter (1970), Volume 2: Chapter 2, Section 2 and Tables B2, B3].

Harter, H. Leon (1961b). Use of tables of percentage points of range and studentized range. Technometrics 3, 407-411. (TE,DI,RA,TO,DC)

Kendall, M. G. (1961). Studies in the history of probability and statistics. XI. Daniel Bernoulli on maximum likelihood. Biometrika 48, 1-2; C. G. Allen's English translations of papers by D. Bernoulli (1778) and by Euler (1778), 3-18. (TE,AV,LR,MA,AM,MD,MA,ML,MM)

Quesenberry, C. P.; David, H. A. (1961). Some tests for outliers. Biometrika 48, 379-387. (TE,TO,BI,NC,GS,EC,AE)

Sen, Pranab Kumar (1961a). On some properties of the asymptotic variances of the sample quartiles and mid-ranges. Journal of the Royal Statistical Society (B) 23, 453-459. (TE,AV,MD,MR,QM,OS)

Sen, Pranab Kumar (1961b). A note on the large-sample behaviour of extreme sample values from distribution with finite end points. Calcutta Statistical Association Bulletin 10, 106-115. (TE,AV,MR,QM,OS,EX)

Srikantan, K. S. (1961). Testing for the single outlier in a regression model. Sankhyā (A) 23, 251-260. (TE,LR,LS,TO,IC,TC,ST,TM,NC,GS)

Tiago de Oliveira, J. (1961). The asymptotic independence of the sample mean and extremes. Universidade de Lisboa, Faculdade de Ciências (A) 8 (2), 299-309. (TE,AV,AM,EX)

Berman, Simeon (1962). Limiting distribution of the Studentised largest observation. Skandinavisk Aktuarietidskrift 45, 154-161. (TE,EX,TO,GS,NC)

Cancelliere, Giovanni (1962). Una dimostrazione del teorema sulla mediana. Rivista Italiana di Economia, Demografia e Statistica 16, 138-139. (TE,AV,MD)

Cucconi, Odoardo (1962). Un criterio per il rigetto delle osservazioni spurie. Scuola in Azione 21, 92-106. (TE,TO,IC,TM,NC,GS,CU)

Eisenhart, Churchill (1962). Roger Joseph Boscovich and the combination of observations. Actes du Symposium International R. J. Boškovic 1961, pp. 19-25. Beograd-Zagreb-Ljubljana. (TE,AV,AV,AM,MD,LR,LF,LS)

Harter, H. Leon (1962). Percentage Points of the Ratio of Two Ranges and Related Tables. AFL 62-378. Wright-Patterson AFB. AD 286848. (TE,EN,RA) [See also Harter (1970), Volume 1: Chapter 1, Section 3 and Table A1].

Hill, Bruce Marvin (1962). A test of linearity vs. convexity of a median regression curve. Annals of Mathematical Statistics 33, 1096-1123. (TE,AV, MD,LR,NR,OS,LS)

Rosengard, Alex (1962). Étude des lois-limites jointes et marginales de la moyenne et des valeurs extrêmes d'un échantillon. Publications de l'Institut de Statistique de l'Université de Paris 11, 3-55. (TE,AV,AM,OS,EX)

Sibuya, Masaaki (1962). A note on the use of median ranges. Annals of the Institute of Statistical Mathematics (Tokyo) 14, 37-39. (TE,AV,DI,AM,MD,RA)

Siddiqui, M. M. (1962). Approximations to moments of the sample median. Annals of Mathematical Statistics 33, 157-168. (TE,AV,MD)

Tukey, John W. (1962). The future of data analysis. Annals of Mathematical Statistics 3, 1-67. (TE,AV,DI,AM,MD,NR,RL,DA,MM,RA,TO,WI,TF)

Anscombe, F. J.; Tukey, John W. (1963). The examination and analysis of residuals. Technometrics 5, 141-160. (TE,LR,NR,LS,TO,JA,WI,TF)

Eisenhart, Churchill; Deming, Lola S.; Martin, Celia S. (1963). Tables Describing Small-Sample Properties of the Mean, Median, Standard Deviation, and Other Statistics in Sampling from Various Distributions. Technical Note 191, National Bureau of Standards. (TE,AV,DI,AM,MD,MR,AD,RA,SD)

Harter, H. Leon (1963). The Use of Sample Ranges and Quasi-Ranges in Setting

Exact Confidence Bounds for the Population Standard Deviation.II. ARL 31, Part II. Wright-Patterson AFB. AD 412352. (TE,DI,QR) [See also Harter (1970), Volume 2: Chapter 1,Section 2 and Tables A6, A7].

Hodges, J. L.,Jr.; Lehmann, E. L. (1963). Estimates of location based on rank tests. Annals of Mathematical Statistics 34, 598-611. (TE,AV,AM,MD,EL,HL,TO, MI,AE)

Laurent, Andre G. (1963). Conditional distribution of order statistics and distribution of the reduced i^{th} order statistic of the exponential model. Annals of Mathematical Statistics 34, 652-657. (TE,TO,TM,DC,GS,PC,LA)

Lieberman, Gerald J.; Miller, Rupert G.,Jr. (1963). Simultaneous tolerance intervals in regression. Biometrika 50, 155-168. (TE,LR,LS,TO)

Niven, B. S. (1963). The joint distribution of the sample mean and the sample range. Australian Journal of Statistics 5, 127-132. (TE,AV,DI,AM,RA)

Tukey, John W.; McLaughlin, Donald H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/ Winsorization I. Sankhyā (A)25, 331-352. (TE,AV,AM,RL,DA,MM)

Wilks, S. S. (1963). Multivariate statistical outliers. Sankhyā (A)25, 407-426. (TE,TO,IC,TM,NC,DC,GS,PC,MM)

Grubbs, Victor (1964). Tests for the Rejection of Outlying Observations. RCA Systems Analysis Technical Memorandum No. 64-7, Missile Test Project, Patrick Air Force Base, Florida. (TE,TO,CC,IC,TM,NC,DC,GS,CT,MR,PC)

Berman, Cyrus (1964). Some notes on the Cauchy distribution. Miscellaneous Studies in Probability and Statistics: Distribution Theory, Small-Sample Problems and Occasional Tables. NBS Technical Note 238, pp. 3-6. Statistical Engineering Laboratory, National Bureau of Standards. (TE,AV,AM,MD)

Dixon, W. J. (1964). Query 4: Rejection of outlying values. Technometrics 6, 238. (TE,TO)

Eisenhart, Churchill (1964). The meaning of "least" in least squares. Journal of the Washington Academy of Sciences 54, 24-33. (TE,AV,AM,MA,LR,LS,LF,MM,EA, ML)

Gebhardt, Friedrich (1964). On the risk of some strategies for outlying observations. Annals of Mathematical Statistics 35, 1524-1536. (TE,TO,GS,KC,AR,PC)

Huber, Peter J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73-101. (TE,AV,AM,MD,RL,MM,DA,HL,HV,DI,LR,LS,TO, AR)

Mardia, K.V. (1964). Exact distributions of extremes, ranges, and midranges in samples from any multivariate population. Journal of the Indian Statistical Association 2, 126-130. (TE,AV,DI,MR,RA,EX)

Natrella, Mary G. (1964). Variances of medians and pseudo-medians. Miscellaneous Studies in Probability and Statistics: Distribution Theory, Small-Sample Problems and Occasional Tables. Technical Note 238, pp. 9-12. Statistical Engineering Laboratory, National Bureau of Standards. (TE,AV,MD,QM)

Paulson, Albert Stanley (1964). On the Performance of Several Tests for Outliers. M. S. thesis, Virginia Polytechnic Institute, Blacksburg, Va. (TE,TO,TM,DC,GS, KC,AR,PC)

Pearson, E. S.; Stephens, M. A. (1964). The ratio of range to standard deviation in the same normal sample. Biometrika 51, 484-49?. (TE,DI,SD,RA)

Rosengard, Alex (1964a). Independance limite uniforme de la moyenne et des valeurs extrêmes d'un échantillon. Comptes Rendus de l'Académie des Sciences (Paris) 258, 5786-5788. (TE,AV,AM,EX)

Rosengard, Alex (1964b). Loi-limite jointe d'un quartile et de la moyenne d'un échantillon. Comptes Rendus de l'Académie des Sciences (Paris) 259, 2759-2761. (TE,AV,AM,QN)

Rosengard, Alex (1964c). Indépendance limite uniforme de la moyenne et des valeurs extrêmes d'un échantillon. Comptes Rendus de l'Académie des Sciences (Paris) 259, 2955-2956. (TE,AV,AM,OS,EX)

Rothenberg, Thomas J.; Fisher, Franklin M.; Tilanus, C. B. (1946). A note on estimation from a Cauchy sample. Journal of the American Statistical Association 59, 460-463. (TE,AV,AM,MO,DA)

Saxena, A. P. (1965). On some tests of hypotheses relating to the exponential distribution when some outliers are present. Journal of the American Statistical Association 60, 548-559. (TE,TO,TM,NC,GS,KC,DC,LA)

Bhattacharjee, G. P. (1965). Distribution of range in non-normal samples. Australian Journal of Statistics 7, 127-141. (TE,DI,RA)

Bickel, Peter J. (1965). On some robust estimates of location. Annals of Mathematical Statistics 36, 847-858. (TE,AV,AM,RL,DA,MM,HL,HV)

David, H. A.; Paulson, A. S. (1965). The performance of several tests for outliers. Biometrika 52, 429-436. (TE,TO,TM,DC,GS,KC,AR,PC)

Gumbel, E. J.; Carlson, P. G.; Mustafi, C. K. (1965). A note on midrange. Annals of Mathematical Statistics 36, 1052-1054. (TE,AV,AM,MR)

Gupta, Shanti S.; Shah, Bimpendra K. (1965). Exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. Annals of Mathematical Statistics 36, 907-920. (TE,DI,RA,OS)

Mardia, K. V. (1965). Tippett's formulas and other results on sample range and extremes. Annals of the Institute of Statistical Mathematics (Tokyo) 17, 85-91. (TE,DI,RA,EX)

Shah, B. K. (1965). Distribution of midrange and semirange from logistic population. Journal of the Indian Statistical Association 3, 185-188. (TE,AV,DI,MR,SK)

Tarter, Michael E.; Clark, Virginia A. (1965). Properties of the median and other order statistics of logistic variates. Annals of Mathematical Statistics 36, 1779-1786. (TE,AV,AM,MD,OS)

Tukey, John W. (1965). Which part of the sample contains the information? Proceedings of the National Academy of Sciences 53, 127-134. (TE,OS)

Anscombe, F. J.; Barron, Bruce A. (1966). Treatment of outliers in samples of size three. Journal of Research of the National Bureau of Standards (B) 70, 141-147. (TE,TO,BT)

Bland, R. P.; Gilbert, R. D.; Kapadia, C. H.; Owen, D. B. (1966). On the distributions of the range and mean range for samples from a normal distribution. Biometrika 53, 245-248. (TE,DI,RA)

Feldman, Dorian; Tucker, Howard G. (1966). Estimation of non-unique quantiles. Annals of Mathematical Statistics 37, 451-457. (TE,AV,MD,HL,QN)

Gastwirth, Joseph L. (1966). On robust procedures. Journal of the American Statistical Association 61, 929-948. (TE,AV,RL,DA,MM,HL,HU,GA,TO)

Gebhardt, Friedrich (1966). On the effect of stragglers on the risk of some mean estimators in small samples. Annals of Mathematical Statistics 37, 441-450. (TE,AV,AM,DA,MM)

Likeš, J. (1966). Distribution of Dixon's statistics in the case of an exponential population. Metrika 11, 46-54. (TE,TO,DC)

Searls, Donald T. (1966). An estimator for a population mean which reduces the effect of large true observations. Journal of the American Statistical Association 61, 1200-1204. (TE,AV,DA,MM)

Sheynin, O. B. (1966). Origin of the theory of errors. Nature 211, 1003-1004. (TE,AV,LR,MA,AM,ML,MM,TO)

Willke, T. A. (1966). A note on contaminated samples of size three. Journal of Research of the National Bureau of Standards (B)70, 149-151. (TE,TO,BT)

Adichie, J. N. (1967). Estimation of regression parameters based on rank tests. Annals of Mathematical Statistics 38, 894-904. (TE,AV,MD,RL,HL,LR,LS,BM)

Anscombe, F. J. (1967). Topics in the investigation of linear relations fitted by the method of least squares. Journal of the Royal Statistical Society (B) 29, 1-29;discussion, 29-52. (TE,AV,MD,RL,HU,LR,LS)

Bickel, P. J.;Hodges, J.L., Jr. (1967). The asymptotic theory of Galton's test and a related simple estimate of location. Annals of Mathematical Statistics 38,73-89. (TE,AV,RL,HL,HA)

Birnbaum, Allan; Laska, Eugene (1967a). Optimal robustness: A general method, with applications to linear estimators of location. Journal of the American Statistical Association 62, 1230-1240. (TE,AV,RL)

Birnbaum, Allan; Laska, Eugene M. (1967b). Optimally Robust Linear Estimators of Location. Technical Report, Courant Institute of Mathematical Sciences, New York University. Abstract, Annals of Mathematical Statistics 38,1932. (TE,AV,RL)

Crow, Edwin L.; Siddiqui, M. M. (1967). Robust estimation of location. Journal of the American Statistical Association 62, 353-389. (TE,AV,RL,HL,HU,DA,WM, LW,MT)

Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, California,

1965/66) (ed. Lucien M. LeCam and Jerzy Neyman), Volume I, pp. 163-186. University of California Press, Berkeley-Los Angeles. (TE,AV,RL,HL,HA)

Hodges, J. L., Jr.; Lehmann, E. L. (1967) On medians and quasi medians. Journal of the American Statistical Association 62, 926-931. (TE,AV,AM,MD,QM)

Hogg, Robert V. (1967). Some observations on robust estimation. Journal of the American Statistical Association 62, 1179-1186. (TE,AV,AM,MD,RL,HL,HU,DA,WM, HO)

Leone, Fred C.; Jayachandran, Toke; Eisenstat, Stanley (1967). A study of robust estimators. Technometrics 9, 652-660. (TE,AV,RL,HL,HU)

Siddiqui, M. M.; Raghunandanan, K. (1967). Asymptotically robust estimators of location. Journal of the American Statistical Association 62, 950-953. (TE,AV, AM,MD,RL,MT,DA,WM,HL)

Singh, Chatter (1967). On the extreme values and range of samples from non-normal populations. Biometrika 54, 541-550. (TE,DI,RA,EX)

Tiao, G. C.; Guttman, Irwin (1967). Analysis of outliers with adjusted residuals. Technometrics 9, 541-559. (TE,TO,DC,AR)

Bhattacharyya, G. K. (1968). Robust estimates of linear trend in multivariate time series. Annals of the Institute of Statistical Mathematics (Tokyo) 20, 299-310. (TE,LR)

Box, G. E. P.; Tiao, G. C. (1968). A Bayesian approach to some outlier problems. Biometrika 55, 119-129. (TE,TO,TM,GS,AR)

Burr, Irving W. (1968). On a general system of distributions. III. The sample range. Journal of the American Statistical Association 63, 636-643. (TE,DI, RA)

Burr, Irving W.; Cislak, Peter J. (1968). On a general system of distributions:

I. Its curve-shape characteristics; II. The sample median. Journal of the American Statistical Association 63, 627-635. (TE,AV,MD)

Cacoullos, T. (1968). A sequential scheme for detecting outliers. Bulletin de la Société Mathématique de Grèce (N.S.) 9 (1), 113-123. (TE,TO)

Cox, D. R. (1968). Notes on some aspects of regression analysis. Journal of the Royal Statistical Society (A) 131, 265-279; discussion, 315-329. (TE,LR)

Cox, D. R.; Hinkley, D. V. (1966). A note on the efficiency of least-squares estimates. Journal of the Royal Statistical Society (B) 30, 284-289. (TE, AV,LR,LS,ML)

Cox, D. R.; Snell, E. J. (1968). A general definition of residuals. Journal of the Royal Statistical Society (B) 30, 248-265; discussion, 265-275. (TE,LR,TO)

Ehrenfeld, A. S. C. (1968). The elements of lawlike relationships. Journal of the Royal Statistical Society (A) 131, 280-302; discussion, 315-329. (TE,LR,LS)

Hampel, Frank Rudolf (1968). Contributions to the Theory of Robust Estimation. Ph.D. dissertation, University of California-Berkeley. University Microfilms, Inc., Ann Arbor, Mich. (TE,AV,RL,DA,HU)

Hartigan, J. A. (1968). Note on discordant observations. Journal of the Royal Statistical Society (B) 30, 545-550. (TE,TO)

Høyland, Arnljot (1968). Robustness of the Wilcoxon estimate of location against a certain dependence. Annals of Mathematical Statistics 39, 1196-1201. (TE, AV,RL,HL)

Huber, P. J. (1968). Robust estimation. Mathematical Centre Tracts, Selected Statistical Papers 27, 3-25. (TE,AV,RL)

Kogure, Masao; Makabe, Hajime (1968). Non-central distributions of both standardized and studentized ranges and their applications-1. Range in a sample

from two populations. Reports of Statistical Application Research, Japanese Union of Scientists and Engineers 15, 71-77 (TE,DI,RA,TO)

Prescott, P. (1968). A simple method of estimating dispersion from normal samples. Applied Statistics 17, 70-74. (TE,DI,SD,DD)

Sen, Pranab Kumar (1968a). On a further robustness property of the test and estimator based on Wilcoxon's signed rank statistic. Annals of Mathematical Statistics 39, 282-285. (TE,AV,RL,HL)

Sen, Pranab Kumar (1968b). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association 63, 1379-1389. (TE,LR)

Bouma, N.; Vehmeyer, A. (1969). Percentiles and approximations of the sample median over the sample range for samples of size 3,5,7 and 9 from a standard normal distribution. Statistica Neerlandica 23, 235-239. (TE,AV,MD,DI,RA)

Desu, M. Mahamunulu; Rodine, Robert H. (1969). Estimation of the population median. Skandinavisk Aktuarietidskrift 52, 67-70. (TE,AV,MD,MR)

Filliben, James John (1969). Simple and Robust Linear Estimation of the Location Parameter of a Symmetric Distribution. Unpublished Ph.D. dissertation, Princeton University. (TE,AV,RL,DA,MM)

Gastwirth, Joseph L.; Rubin, Herman (1969). On robust linear estimators. Annals of Mathematical Statistics 40, 24-39. (TE,AV,RL,HL,HU,GA)

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics 11, 1-21. (TE,TO,TM,DC,CS,KC,AR,FC,MI)

Guttman, I.; Smith, D. E. (1969). Investigation of rules for dealing with outliers in small samples from the normal distribution: I. Estimation of the mean. Technometrics 11, 527-550. (TE,AV,TO)

de Haan, L.; Rummenberg, J. Th. (1969). Some remarks concerning the quotient of sample median and sample range for a sample of size 2n+1 from a normal distribution. Statistica Neerlandica 23, 227-234. (TE,AV,MD,DI,RA)

Hawkins, D. M. (1969). On the distribution and power of a test for a single outlier. South African Statistical Journal 3, 9-15. (TE,TO)

Jaeckel, Louis Alan (1969). Robust Estimates of Location. Ph.D. dissertation, University of California-Berkeley. University Microfilms, Inc., Ann Arbor, Michigan. (TE,AV,RL,ML,OS,DA,HL,HU)

Narayana, C. L.; Subrahmanyam, M. (1969). On an alternative method to the principle of least squares in the theory of regression. Sankhyā (B) 31, 427-434. (TE,LR,LS,AS)

Rao, P. V.; Thornby, J. I. (1969). A robust point estimator in a generalized regression model. Annals of Mathematical Statistics 40, 1784-1790. (TE,AV, MD,RL,HL,LR,NR,BM)

Swaroop,Ram; West, Kenneth A.; Lewis, Charles E., Jr. (1969). A simple Technique for Automatic Computer Editing of Biodata. NASA TN D-5275. National Aeronautics and Space Administration, Washington, D. C. (TE,TO)

Takeuchi, Kei (1969). A Uniformly Asymptotically Efficient Robust Estimator of a Location Parameter. Technical Report IMM 375, Courant Institute, New York University. (TE,AV,RL,DA,WM,HL,HU,HO)

Thomas, Jerry (1969). Monte Carlo investigation of the robustness of Dixon's criteria for testing outlying observations. Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing, pp. 437-483. (ARO-D Report 69-2, U. S. Army Research Office, Durham, N.C.). (TE,TO,DC)

Walsh, J. E. (1969). Sample sizes for approximate independence between sample median and largest (or smallest) order statistic. Australian Journal of Statistics 11, 120-122. (TE,AV,MD,EX)

Yanagawa, Takashi (1969). A small-sample robust competitor of Hodges-Lehmann estimate. Bulletin of Mathematical Statistics (Fukuoka) 13(3-4), 1-14. (TE, AV,RL,YE,HA)

Zelenen'kiy, V. P. [Zelenen'kiĭ, V. P.] (1969). Application of statistical decision theory to the exclusion of anomalous measurements. Izvestia Akademiia Nauk SSSR, Tekhnicheskaya Kibernetika 1969(2), 139-142 (Russian); translated as Engineering Cybernetics 1969(2), 122-126. (TE,TO)

Barnett, V. D. (1970). Fitting straight lines--the linear functional relationship with replicated observations. Applied Statistics 19, 135-144. (TE,LR)

Birnbaum, Allan; Mike, Valerie (1970). Asymptotically robust estimators of location. Journal of the American Statistical Association 65, 1265-1282. (TE,AV,RL,HU)

Fuller, Wayne A. (1970). Simple Estimators for the Mean of Skewed Populations. Prepared under Contract No. Cco-9165 with U. S. Bureau of the Census. Iowa State University, Ames, Iowa. (TE,AV,RL,MM,TO,DC,AR)

Gastwirth, J. L.; Cohen, M. L. (1970). Small sample behavior of some robust linear estimators of location. Journal of the American Statistical Association 65, 946-973. (TE,AV,RL,HL,HU,GA)

Harter, H. Leon (1970). Order Statistics and their Use in Testing and Estimation, Volumes 1 and 2. U. S. Government Printing Office, Washington. (TE,DI, RA,QR,OS)

Huber, P. J. (1970). Studentizing robust estimates. Nonparametric Techniques

in Statistical Inference (ed. M. L. Puri), pp. 453-463. Cambridge University Press, Cambridge, England. (TE,AV,ML,HU)

Kabe, D. G. (1970). Testing outliers from an exponential population. Metrika 15, 15-18. (TE,TO)

Loynes, R. M. (1970). On the asymptotic relative efficiencies of certain location parameter estimates. Journal of the Royal Statistical Society (B) 32, 134-136. (TE,AV,AM,MD,PM,RL,HU)

Singh, C. (1970). On the distribution of range of samples from non-normal populations. Biometrika 57, 451-456. (TE,DI,RA)

Switzer, Paul (1970). Comments and Suggestions on Efficiency Robustness. Technical Report No. 163, prepared under ONR Contract N00014-67-A-0112-0055, NR-042-067, Department of Statistics, Stanford University. AD 714813. (TE,AV, RL,SM)

Birnbaum, Allan; Laska, Eugene; Meisner, Morris (1971). Optimally robust linear estimators of location. Journal of the American Statistical Association 66, 302-310. (TE,AV,RL)

Duckworth, F. C. (1971). On the Limitations of the Method of Least Squares for the Treatment of Certain Types of Data. RD B N-1971, Berkeley Nuclear Laboratories, Central Electricity Generating Board, Berkeley, England. (TE, LR,LS)

Edwards, Arthur L. (1971). FITTER, FUNGUS, FUNNY: Linear Least-Squares Curve Fitting of Weighted Tabular Data. UCID-30017, Lawrence Radiation Laboratory, University of California, Livermore. (TE,LR,LS)

Johns, M.V.,Jr. (1971). Non-Parametric Estimation of Location. Technical Report

No. 41, Department of Statistics, Stanford University, Stanford, California. (TE,TD,ML,MD,SM)

Rummel, R. E.; Burdick, R. G. (1971). A Graphical Method of Removing Outlier Values from Analytical Data. Report of Investigations 7472, Bureau of Mines, U. S. Department of the Interior, Washington, D. C. PB 197023. (TE,TD)

Sherroop, Tom; Winter, William R. (1971). A Statistical Technique for Computer Identification of Outliers in Multivariate Data. NASA TN D-6472. National Aeronautics and Space Administration, Washington, D. C. (TE,TD)

Caso, John (1972). Robust Estimation Techniques for Location Parameter Estimation of Symmetric Distributions. Unpublished M. S. thesis (GSA/Math/72-3), Air Force Institute of Technology. (TE,AV,ML,DA,MM,EL,HU,HD,SM)

Eisenhart, Churchill (1972). The development of the concept of the best mean of a set of measurements from antiquity to the present day. Journal of the American Statistical Association 67 (to appear). (TE,AV,AM,MG,HD,MR,LR,LS,EA, LF,ML)

Additional References

Neyman, J.; Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A, 175-240. (TE,AV,DI,MR,RA)

Pearson, Egon S. (1932). The percentage limits for the distribution of range in samples from a normal population. (n≤100). Biometrika 24,404-417. (TE,DI, RA)

McKay, A. T.; Pearson, E. S. (1933). A note on the distribution of range in samples of n. Biometrika 25, 415-420. (TE,DI,RA)

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. Biometrika 27, 466-471. (TE,AW,AM,ST,TD,KC,AK)

Nair, K. R.; Shrivastava, M. P. (1942). On a simple method of curve fitting. Sankhyā 6, 121-132. (TE,LR,LS,AS)

Daly, Joseph F. (1946). On the use of sample range in an analogue of Student's t-test. Annals of Mathematical Statistics 17, 71-74. (TE,AW,AM,AA)

Lord, E. (1947). The use of range in place of standard deviation in the t-test. Biometrika 34, 41-67; corrigenda, 39(1952), 442. (TE,AW,AM,AA)

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

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

GR geometric range

GM geometric mean

GR Grubbs' rule (for rejection of outliers)

GS Grubbs' criterion (for rejection of outliers)

HA Hodges' alternative (to Hodges-Lehmann estimator)

HC Hoyleaureich's criterion (for rejection of outliers)

HL Hodges-Lehmann estimator

HM harmonic mean

HO Hogg's estimator

HU Huber's estimator

IC Irwin's criterion (for rejection of outliers)

IR interquartile range

JA Jeffreys' alternative (to the rejection of outliers)

KC Kudo'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

MK McKay's criterion (for rejection of outliers)

ML [illegible] likelihood

MM [illegible] method [minimize maximum residual]

MO mode

MQ median-quartile average

MR midrange

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)

NW Newcomb's method (of treating outliers)

NR nonlinear regression

NS Nair-Shrivastava method (of curve fitting)

OS order statistics

PA plus approximative methods [best 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

SC Stone's (first) criterion (for rejection of outliers)

SD standard deviation [or variance ≡ $(SD)^2$]

SM Stewart's method (criterion) (for rejection of outliers)

SR studentrange

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