Int. Stat. Rev., Vol. 43, No. 1, 1975, pp. 1–44/*Longman Group Ltd/Printed in Great Britain*

The Method of Least Squares and Some Alternatives—Part III[1]

H. Leon Harter
Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio, USA

5. The Modern Era I (1946–1964)

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 $n \geqq 100$ 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 $n^{-\frac{1}{2}}$. For the one-parameter rectangular distribution $f(x) = 1/\lambda$, $0 \leqq x \leqq \lambda$, he finds that the largest observation v "is a sufficient statistic and is evidently the maximum likelihood estimate of λ".

Harald 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 favour 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.

[1] For Part I, see *International Statistical Review* (1974), **42**, 147–174.
For Part II, see *International Statistical Review* (1974), **42**, 235–264.

(See form 1473)

E. J. Gumbel (1946) shows that in a sample of size n (large) the mth observation from one extreme and the kth 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 article.

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 pth and qth 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 $n\{2[(2n-2)!-(n-1)!)^2]/(2n-1)!\}^{\frac{1}{2}}$, which is approximately $n^{\frac{1}{2}}$ 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 (1948*a*) 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 $P_{\varepsilon,n}$-probability point of the parent distribution, where $\sum_{k=(n+1)/2}^{n} C_k^n P_{\varepsilon,n}^k (1-P_{\varepsilon,n})^{n-k} = \varepsilon$ 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 (1948*b*) 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 favourably 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 $x_1, x_2, \ldots, x_k$ 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 ($i = 1, 2, \ldots, n$) in a sample of size n from a normal distribution concerning the mean μ and the variance σ^2, and tabulates results for

$$n = 2, 3, \ldots, 12;\ i = 1, 2, \ldots, [n/2]+1.$$

Not surpisingly, 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 $(x_n-x_1)/2$, yields an unbiased estimator of σ, and studies the distribution of the semirange. (See Note at end)

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 $2 \leq n \leq 10$. (See Note at end)

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 non-normal 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 $N = 3$, $\bar{r}/\sigma \doteq 2{\cdot}10-0{\cdot}81\ \sigma_r/\bar{r}$ for eighteen discrete unimodal distributions: for $N = 4$ and $N = 5$, respectively, $\bar{r}/\sigma \doteq 2{\cdot}29-0{\cdot}69\ \sigma_r/\bar{r}$ and $\bar{r}/\sigma \doteq 2{\cdot}41-0{\cdot}46\ \sigma_r/\bar{r}$ for five selected populations of extreme form.

John Wilder Tukey (1949*a*, *b*, *c*, *d*) and Theodore E. Harris and Tukey (1949) report on a study of sampling from contaminated distributions. Tukey (1949*a*, *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 (1949*b*) consider the relative efficiencies of estimators obtained from the mean and standard deviation by removing the extreme γ per cent 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 neighbour 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 discarding 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.

Alexander McFarlane Mood (1950) deals with regression and linear hypotheses in Chapter 13 (pp. 289–315) of his textbook. He includes a section (13.8) on the method of least squares, in which he states (p. 311): "The primary reason that the method of least squares is commonly used for curve fitting is merely that it leads to a simple linear system of equations for determining the coefficients. To determine the coefficients by minimizing, say, the sum of the absolute deviations, or the sum of the fourth powers of the deviations, would ordinarily be much more troublesome. It just happens that the form of the normal distribution is such that the sum of squares of deviations from the regression function is to be minimized to determine the coefficients in the regression function. If, for example, the points . . . were supposed to be deviations from a regression line with a probability distribution other than a normal distribution, then it would be appropriate to determine estimates of α and β [the regression coefficients] by maximizing the likelihood defined by that distribution. Even here, though, the method of least squares is commonly used in practice to avoid algebraic and arithmetic difficulties, and this is, of course, good and sufficient reason. The theoretical advantages of the principle of maximum likelihood over the principle of least squares may become unimportant when it comes to a matter of choosing, say, between a 40-hour and a 10-hour computation." [This conclusion should be re-examined in view of the great reduction in computing times that has occurred as a result of the development of modern high-speed computers during the past quarter-century.] In Chapter 16 (pp. 385–418), the author presents a number of distribution-free methods based on order statistics. Such a method for linear regression, developed jointly by G. W. Brown and the author, is outlined in sections 13.8 and 13.9 and is discussed in greater detail by Brown and Mood (1951) [q.v.].

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, $N(0, 1)$.

James Blaine Scarborough (1950) discusses the normal law of error and the principle of least squares in Chapter XIV. He points out that all measurements are subject to three kinds of errors: constant or systematic errors, mistakes or blunders, and accidental errors. The mathematical theory of errors deals only with accidental errors, which are those whose causes are unknown and indeterminate. They are usually small, and follow the laws of chance. Often [but not always] they are well approximated by the normal law of error, which the author discusses at length, along with the associated principle of least squares, which leads to choice of the arithmetic mean as the best average of a number of observations, with weights proportional to the square of the precision indices $h_i = 1/\sigma_i\sqrt{2}$. In Chapter XV, the author discusses the precision of measurements, which can be estimated by the [root] mean square error [standard deviation σ], the probable error (P.E.), or the average [absolute] error. In Art. 141, he gives the following rule [cf. Wright (1884)] for rejection of outlying observations: "Find the mean of all the measurements (including the "wild" one) and find the residual for each. Compute the P.E. of a single measurement. . . . Reject any measurement whose residual exceeds 5 times the P.E. of a single measurement." In Chapter XVI, which deals with empirical formulas, the author compares three fitting methods: the graphic method (method of selected points), the method of averages, and the method of least squares. He does not recommend the method of selected points except for obtaining approximate values or in cases where the results obtainable by this method are as accurate as the data used. The method of averages involves grouping the residual equations into as many groups as there are constants in the assumed formula, with each group containing as nearly as possible the same number of residuals, setting the sum of the residuals in each group equal to zero, and solving the resulting equations. The residual equations can be grouped in several ways, each giving a different result. The best formula is obtained by grouping in consecutive order of the values of [one of] the independent variable[s]. The author states that when the number of residual equations is large enough to allow three or more to each group, the method of averages can be depended upon to give good results, but that otherwise we should always use the method of least squares, which gives only one formula, which is always the best possible [if the errors follow the normal law]. He discusses least squares regression in detail, including computational procedures, weighted residuals, non-linear formulas, both variables subject to error, and finding the best type of formula. He does not mention alternatives based on non-normal error laws.

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 ($x_1 \leqq x_2 \leqq x_3$), 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.

S. I. Zukhovitskii (1950) develops a procedure, based on the Fourier descent method, for finding the best approximation (in the Chebyshev sense of minimizing the maximum error) for a system of incompatible linear equations in the de la Vallée Poussin (non-degenerate) case.

D. H. Bhate (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 $y_i, z_{1i}, z_{2i}, \ldots, z_{ki}$ with $i = 1, 2, \ldots, n$. Then the coefficients α_r are estimated by the numbers $\tilde{\alpha}_r$ such that $\underset{z_{ri} < \tilde{z}_r}{\text{median}} (y_i - \tilde{\alpha}_0 - \Sigma \tilde{\alpha}_r z_{ri}) = \underset{z_{ri} > \tilde{z}_r}{\text{median}} (y_i - \tilde{\alpha}_0 - \Sigma \tilde{\alpha}_r z_{ri}) = 0$, $r = 1, 2, \ldots, k$, 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 $x_1, x_2, \ldots, x_n$ are the order statistics of a sample of size $n \leqq 30$ from a population which is (1) rectangular or (2) normal. He tabulates 3-decimal-place percentage points, corresponding to cumulative probability $\alpha = 0{\cdot}005$, 0·01, 0·02, 0·05, 0·1 (0·1) 0·9, 0·95, 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 and 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.

S. I. Zukhovitskii (1951*a*) describes a method of successive approximations for computing the best solution (in the sense of Chebyshev) of a set of incompatible linear equations

$$\sum_{j=1}^{n} a_{ij}\xi_j = b_i \; (i = 1, 2, \ldots, m)$$

where $m > n$, i.e. the set of ξ's which minimizes $\max_i \left| \sum_{j=1}^{n} a_{ij}\xi_j - b_i \right|$. Zukhovitskii (1951*b*) gives an analogous procedure for finding the centre of the smallest sphere containing m

given points, i.e. the point of least deviation (in the sense of Chebyshev) from the given points.

J. H. Cadwell (1952) 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 (1952*a*, *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.

Tsuruchiyo 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.

A. S. Shteinberg (1952*a*, *b*) discusses various methods of solving the problem of the best uniform approximation (i.e., the best approximation in the Chebyshev sense) for a system of incompatible linear equations. One of these methods is the Fourier descent method, for which the procedure is far from uniquely determined.

Ralph Hoyt Bacon (1953) reviews the familiar method of least squares for fitting a straight line to a set of observed points, and discusses variations of it, including the problem of unequal weights and the minimization of residuals other than the vertical. If the ordinates of the observed points are known exactly but the abscissas are subject to error, or if the errors of the ordinates are negligible compared with those of the abscissas, one minimizes the horizontal deviations. If both variables are subject to error, the procedure is more complicated, and the author refers the reader to the treatment of Deming (1943), which he says is the best of which he is aware. He also explains methods for estimating the adequacy of the fit obtained and for comparing the results of several experiments. In an appendix, he gives some of the principles of mathematical statistics underlying these methods.

J. H. Cadwell (1953) presents a method for evaluating the probability density function of the rth 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 $2 \leqq n \leqq 17$ and w_1 most efficient for $18 \leqq n \leqq 31$.

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 over the interval $(-c, +c)$, where c is known, the likelihood function is a constant, and there is no unique maximum likelihood solution, one admissible solution being just as likely as another. [If c is unknown, the maximum likelihood solution is found by minimizing the maximum residual.]

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 per cent and 1 per cent 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 per cent 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)].

Zukhovitskii (1953) presents an algorithm for numerical evaluation of the vector x for which the residual m-vector $r = Ax-b$ has the least possible value for its longest component, given a set of m equations in n unknowns $Ax = b$, where A is an $m \times n$ matrix, $m>n$, b is an m-vector and x is an n-vector. He considers both the non-degenerate (de la Vallée Poussin) case in which the Haar condition (that all n-rowed determinants in A be non-zero) is satisfied and the degenerate case in which this condition is not fulfilled. He devises computational procedures for each case and works out a numerical example of each.

Shmuel Agmon (1954) notes that in various numerical problems one is confronted with the task of solving a consistent system of linear inequalities. He discusses several methods of solving such a system, with emphasis on the relaxation method. He points out that in addition one sometimes has to minimize a linear form, e.g. in linear programming. As another example where this is necessary he mentions the problem of finding the polynomial of best approximation (in the Chebyshev sense of minimizing the maximum deviation) of degree less than n corresponding to a discrete function defined in $N[>n]$ points. He points out that one can, at the expense of increasing considerably the number of unknowns and inequalities in the equivalent system, use the duality principle to reduce this problem to the solution of a system of inequalities involving no minimization. T. S. Motzkin and I. J. Schoenberg (1954) give further theoretical results on the relaxation method, without any mention of the applications discussed by Agmon.

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 $\beta_2 = 1{\cdot}0\,(0{\cdot}2)\,2{\cdot}0\,(0{\cdot}5)\,5{\cdot}0)\,(1{\cdot}0)\,9{\cdot}0$; 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 $x_1, \ldots, x_n$ 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 mth 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.

Morris Morduchow (1954) shows, under fairly general conditions, that the standard deviation of the residuals of the straight line fitting a set of points with uniformly spaced abscissas by the method of averages is at most $2/\sqrt{3}$ times as great as the standard deviation of the residuals of the least squares line. In the course of the analysis, he proves theorems of significance in the practical use of the method of averages.

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 standardized 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.

A. Charnes, W. W. Cooper and R. O. Ferguson (1955) propose the use of linear programming as an alternative to least squares in estimating executive compensation. The method which they propose involves minimizing the sum of the absolute values of the deviations of salaries of individual executives from the salary levels deemed appropriate for executives occupying their respective positions in the job hierarchy, when the salaries are expressed as linear functions of the rankings of the individuals with regard to certain characteristics believed to influence their performance on the job, subject to certain other constraints, e.g. maximum and minimum compensation.

John T. Chu (1955*a*) 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 (1955*b*) shows that, under very general conditions, $\text{var}\,\tilde{x} \geqq \{4\,[f(\xi)]^2\,(2n+3)\}^{-1}$, 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 *m*th range [quasi-range] in terms of Bessel functions of the third kind, and integrates by parts to obtain the distribution of the *m*th range.

Max Halperin, Samuel W. Greenhouse, Jerome Cornfield and Julia Zalokar (1955) tabulate, to three significant figures, the upper and lower 5 per cent and 1 per cent points of the studentized maximum absolute deviate $d = \max_i | x_i - \bar{x} |/s$, where the x_i ($i = 1, \ldots, k$) 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

$$k = 3\,(1)\,10\,(5)\,20\,(10)\,40,\ 60 \text{ and } m = 3\,(1)\,10\,(5)\,20\,(10)\,40,\ 60,\ 120.$$

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.

Cecil Hastings, Jr. (1955) gives an elementary discussion (with graphical illustrations) of best fit in the Chebyshev sense (minimum absolute error). He also gives an iterative procedure for finding the straight line which best fits a finite number of points, as well as an iterative procedure for finding the curve of a given parametric form which best fits a continuous curve over an interval.

E. Ya. Remez (1955*a*, *b*) considers the problem of determining that curve f out of a given family F which has smallest maximum deviation from a given set S, for the cases in which F consists of all straight lines or of all straight lines through the origin. When the set S is finite, inspection of its graph gives the points of maximum deviation, and the author gives an elementary computational procedure for finding the coefficients in the equation of f.

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 per cent, 0·5 per cent, 1·0 per cent, 2·5 per cent, 5·0 per cent and 10·0 per cent points and the median (50 per cent point) of w/s to five decimal places for samples of size $n = 3$ from a normal population.

J. Topping (1955), in a monograph on errors of observation and their treatment, distinguishes between accidental and systematic errors, considers estimation of the error in repeated observations of the same quantity and in compound quantities (products, quotients, sums and differences of quantities subject to error). He introduces frequency distributions (binomial, Poisson, normal, and others) and measures of central tendency (mean and median) and of dispersion (range, mean deviation, and standard deviation). His treatment of the theory of error is based entirely on the normal law of error and the method of least squares; though he admits that the former (and hence also the latter) is not universally valid, he does not offer any alternatives.

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 and 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 $f((y \mid x)$, 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 per cent 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 and Hartley (1954)], but corrected by using a better approximation.

Sabri Ergun (1956) adapts the method of least squares to the fitting of families of straight lines having a common slope or a common intercept, and gives applications to the analysis of data relating to various physiochemical phenomena.

A. A. Goldstein (1956) gives an algorithm for the method of descent in complex domains and its application to the minimal approximation of over-determined systems of linear equations, and shows its relation to linear programming.

Akio Kudô (1956) proposes a new criterion for the rejection of outlying observations. Given three sets of independent observations $\{x\}$: (i) n_1 from $N(m_i, \sigma^2)$, $i = 1, 2, \ldots, n_1$; (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

$$H_0\colon m_1 = m_2 = \ldots = m_n = m^{(2)},\ H_i\,(i \neq 0)\colon m^{(2)} = \text{each } m_j \text{ (except } m_i) = m_i - \Delta.$$

The author presents a decision procedure for which: $\Pr(\text{acc. } D_0 \mid H_0) = 1-p$, $\Pr(\text{acc. } D_i \mid H_i)$ is maximized for $i \neq 0$. The optimum decision procedure involves

$$x_M = \max\{x_1, x_2, \ldots, x_i, \ldots, x_{n_1}\};$$

$\bar{x}$, 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)].

Remez (1956*a*) gives a concise summary of methods previously proposed for finding the best solution (in the Chebyshev sense of minimizing the maximum error) of an incompatible set of linear equations. He previews a forthcoming monograph [Remez (1957)] which presents an effective new numerical method of prevailing deviations for the solution of the problem. Remez (1956*b*) considers the problem of uniqueness or multiplicity of solutions of the Chebyshev problem for a system of incompatible linear equations.

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.

S. I. Askovitz (1957) presents a simple graphic technique for obtaining without calculation the best fitting straight line according to the least squares criterion, provided the points are equally spaced horizontally. A graphic measure of residual variability (mean deviation of the points from the fitted line) is also derived.

Juan Bejar (1957) gives a method, similar to linear programming, to determine the regression line $y = a+bx$ such that $\Sigma \mid y_i-a-bx_i \mid$ is a minimum or the regression plane $z = a+bx+cy$ such that $\Sigma \mid z_i-a-bx_i-cy_i \mid$ 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 rth 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 $n \geqq 9$ and W_2 is more efficient than W_1 for $n \geqq 17$.

Allen A. Goldstein, Norman Levine and James B. Hereshoff (1957) note that several methods of constructing best approximations of continuous functions have been published, but that, on the other hand, the important problem of solving an overdetermined system of linear equations in the Chebyshev sense has not been given the attention it deserves, though an unwieldy and not widely known algorithm has been given by de la Vallée Poussin (1911). This algorithm is based on the theorem, proved by de la Vallée Poussin, that the minimal maximum residual M of a system of m linear equations in n unknowns $(m>n)$ is equal to the largest of the minimal maximum residuals $M_1, M_2, \ldots$ corresponding to the various systems which can be formed by taking the equations $n+1$ at a time. The best approximation for a system of $n+1$ equations in n unknowns can easily be found by solving a system of $n+1$ equations in $n+1$ unknowns, assuming that none of the determinants of any selection of n equations vanishes. At worst the method of de la Vallée Poussin requires the solution of $\binom{m}{n+1}$ systems of $n+1$ equations in $n+1$ unknowns. Even though considerable savings can be realized by a good first approximation, the computation is impractical for large systems. The authors discuss two new methods. They consider functions M and S defined by

$$M = \sup_i \mid \varepsilon_i \mid,\ S = \sum_{i=1}^{m} \varepsilon_i^{2q}\ (q \text{ integral})$$

where $\varepsilon_i = \sum_{j=1}^{m} a_{ij}x_j-b_i$. They note that for sufficiently large q the minimizing vector for S lies arbitrarily close to the minimizing vector for M. To find these vectors they apply the method of descent; more specifically, they minimize S by moving along the largest component of the gradient. They give numerical examples of the minimization of S. They write (p. 343): "The direction of descent to minimize M is more critical. From a first approximation x_{j1}, the maximum $n+1$ residuals are selected and the best approximation x_{j2} for this set is found. The points x_{j1} and x_{j2} are joined by a straight line and M – now a function of one variable – is minimized [this involves solution of at most $\binom{m}{2}$ systems of two linear equations in two unknowns] along this line yielding x_{j3}. The maximum $n+1$ residuals of x_{j3} are selected and a 'best approximation' x_{j4} for this set is found. We join x_{j3} and x_{j4} by a straight line and minimize along this line obtaining x_{j5}, etc."

B. I. Harley and E. S. Pearson (1957) 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 and 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).

J. W. Head and G. M. Oulton (1957) consider a situation in which one has n experimentally determined pairs of values of an independent variable x and a dependent variable y, and has reason to expect that a known function Y of x and y is a polynomial in some other known function X of x and y. They show geometrically how the goodness of fit of the least-squares polynomial is affected when its degree is raised from 1 to 2, the number n of pairs of points being 5. The results suggest that the five points are not usually placed so that the fit is appreciably better for $m = 2$ than for $m = 1$, and that it is seldom useful to fit a non-linear curve to experimental results, at least when $n = 5$.

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.

Remez (1957) discusses at length the problem of best approximation in the sense of Chebyshev (minimizing the maximum error), with emphasis on computational methods. He considers both the general case of approximation to a continuous function over an interval and the special case of approximation to a finite set of points. He summarizes the contributions to the latter case of Laplace, Fourier, Kirchberger, Goedseels, and de la Vallée Poussin spanning the period from the late eighteenth to the early twentieth century, as well as the more recent ones of the Kiev school of mathematicians, especially Zukhovitskii, Shteinberg, and himself. For the solution of finite systems of linear equations he presents the α-algorithm of successive weighted quadratic approximations, the method of successive equalizing descents (*a*) under the Haar-de la Vallée Poussin determinantal condition and (*b*) when this condition does not hold, and the new method of prevailing deviations previewed in his earlier paper [Remez (1956*a*)]. He gives numerical examples of the use of the various methods.

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 $\alpha_4 = 2{\cdot}19, 2{\cdot}14, 1{\cdot}8, 1{\cdot}19, 1$), 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 $w = 0\,(0{\cdot}05)\,7{\cdot}65$.

J. L. Walsh and T. S. Motzkin (1957) point out that polynomials $p_n(x)$ of given degree $n\ (\geq 0)$ of best approximation to a given function $f(x)$ on a real finite point set

$$E\colon (x_1, x_2, \ldots, x_m)$$

are important in numerical computation and have various properties in common, especially those relating to oscillation of the difference $f(x) - p_n(x)$ on E. They state without proof some new results on the totality of such polynomials, where approximation is measured according to any of the classical deviations, such as that of Chebyshev (minimax) or that of least pth powers.

Askovitz (1958) presents a mathematical model which corresponds to the problem of determining the straight line best fitting a set of points according to the least squares criterion. The model consists essentially of a bar in equilibrium under the action of suitably arranged springs. By applying certain relationships between centroids of points and vectors in equilibrium, two points can be located which determine the correct line. A graphical method can thus be developed for drawing the line without any calculations. The abscissas of the observed points need not be equally spaced as for the method given by the author in an earlier paper [Askovitz (1957)]. The author illustrates the technique by an example involving three points.

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), $i = 1, 2, \ldots, n$, 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 $k \doteq 0{\cdot}27n$. 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 $i = 1, 2, \ldots, n-1$.

Ward Cheney and Allen A. Goldstein (1958) elaborate the paper of Zukhovitskii (1951*a*), supplying certain missing details. They write (p. 233): "Consider a system of equations (1) $(A^i, x) = b_i\ (1 \leqq i \leqq m)$ in which $A^i = (A^i_1, A^i_2, \ldots, A^i_n) \in E_n$ [Euclidean n-space], $x = (x_1, x_2, \ldots, x_n) \in E_n$, $b = (b_1, b_2, \ldots, b_m) \in E_m$, where $m > n$, and where $(A^i, x) = \sum_{j=1}^{n} A^i_j x_j$. The restriction is imposed that each set of n rows of the matrix $A = (A^i_j)$ be of rank n. This is termed the *Haar* condition, although it is believed that priority belongs to de la Vallée Poussin [(1911)]. . . . Define *residual functions* (2) $R^i(x) = (A^i, x) - b_i\ (1 \leqq i \leqq m)$. The number $F(x) = \max_{1 \leqq i \leqq m} |R^i(x)|$ is called the *deviation* of system (1) at the point x. The problem of solving (1) approximately in the sense of Tchebycheff [Chebyshev] is that of obtaining a point $\bar{x}$ in E_n which minimizes F. The number $F(\bar{x})$ is known as the *minimum deviation* of system (1), and $\bar{x}$ is termed a *minimax solution*. As is shown in [a paper by Haar (1918)], the Haar condition guarantees the uniqueness of $\bar{x}$." The authors give the problem a geometic interpretation first pointed out by Fourier (1824), prove a theorem stated by Zukhovitskii (1951*a*), and make several relevant remarks. In a related paper [Goldstein and Cheney (1958)], the same authors treat the three problems of the solution of consistent linear equations and inequalities and the Chebyshev approximation of inconsistent linear equations from a unified geometric standpoint, and present an algorithm for their solution. They reinterpret each problem as one of finding the lowest points (if any exist) of a polytope in a Euclidean space and then combine the techniques of steepest descent and elimination of variables to work downward from vertex to vertex. They note that the same viewpoint has been exploited in the algorithms of Fourier (1824) and of de la Vallée Poussin (1911) and that other algorithms have been given by Motzkin and Schoenberg (1954) and by Goldstein *et al.* (1957). After first outlining the algorithm and then stating it in detail, they consider the special case of n variables and $n+1$ equations or inequalities, prove the finiteness of the algorithm, and discuss relationships among the three above problems and linear programming. In particular, on page 425, they state and prove the following theorem: "The following three problems are equivalent in the sense that each may be reduced to problems of the other types: (1) The Tchebycheff problem of finding x in E_n which minimizes $\max_{1 \leq i \leq m} |(A^i, x) - b_i|$. (2) The linear inequality problem of finding x in E_n such that $(A^i, x) \leq b_i$, $(1 \leq i \leq m)$. (3) The linear programming problem of finding x in E_n which minimizes (L, x) subject to $(A^i, x) \leq b_i$, $(1 \leq i \leq m)$ [where (L, x) is an arbitrary linear objective function]."

Chan-Hui Chou (1958) gives useful ideas on the limitations of the least squares method in chemical calculations and the pitfalls encountered in technical work. He points out that the method of least squares should be used to determine the best set of values for the constant coefficients of a prescribed form of an equation, not to decide which form of an equation should be used to represent the given set of data. The choice of an equation form should come before application of the method of least squares. If this is not possible, the choice should be made by using statistical methods (significance tests and/or confidence procedures) in conjunction with the method of least squares, not by the latter method alone.

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 rth 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.

Otto J. Karst (1958) develops a method for finding a straight line of best fit to a set of two-dimensional points such that the sum of the absolute values of the vertical deviations of the points from the line is a minimum. If the line is constrained to pass through any designated point, one application of a numerical procedure leads to the solution, which may be a unique line or a sheaf of lines. Otherwise, iteration of the procedure is necessary to find the line or lines of least deviations.

James E. Kelley, Jr. (1958) applies linear programming to curve fitting. He writes (pp. 15–16): "The following approximation problem is of recurring interest: Given a set of real-valued functions $f(x)$, $g_j(x)$, $(0 \leq j \leq n)$, continuous on the closed interval S of the real axis, find an n-dimensional real vector $\alpha = (\alpha_j)$ such that the norm of

$$(1)\ \delta(\alpha, x) = f(x) - \Sigma_j \alpha_j g_j\ (x \text{ in } S)$$

is a minimum. . . . Of most practical interest in the present context are norms of the form (2) $L_p(\alpha) = \left[\int_S |\delta(\alpha, x)|^p dx\right]^{1/p}$ where p is a positive integer. The case $p = 2$, which is simply the 'least squares' problem, has been adequately solved by classical means. . . . The general case has recently been treated by a method of steepest descent [Goldstein *et al.* (1957)]. Except perhaps for $p = 2$, the case $p = \infty$ is probably most interesting. In this case, minimizing (2) is equivalent to minimizing the functional (3) $\lambda(\alpha) = \max_S |\delta(\alpha, x)|$. . . . In the present paper, we will consider how linear programming may be used to minimize (3) in the case that the functions $f(x)$, $g_j(x)$ $(0 \leq j \leq n)$ are given at a finite set of points, x_i, $(1 \leq i \leq m)$ in S". He applies the simplex method to a dual formulation of the problem, and constructs an algorithm for which n need not be specified in advance, but can be chosen so as to meet a preassigned tolerance.

Yu V. Linnik (1958), in his introduction, gives an outline of problems in least squares and some typical examples, together with a brief historical outline from the time of Legendre (1805) [1806], Gauss (1809), and Laplace (1812). In Chapters I–III he outlines the concepts of algebra, probability theory, and mathematical statistics needed to follow his presentation of the method of least squares in succeeding chapters. Chapter IV includes a section on the rejection of outlying observations, for which the author recommends the criterion of Grubbs (1950), and Chapter VII includes a section dealing specifically with the problem of linear regression. Chapter XIII deals with the work of Wald (1940) on the problem of regression when both variables are subject to error and on the method of averages. In Chapter XIV, the author gives miscellaneous additional results. Included are sections on the role of the normal law in the theory of least squares and on Cauchy's method of interpolation as a substitute for the method of least squares. The author states that, because the calculation is very simple, Cauchy's method is preferable to the method of least squares in many cases, even though it results in some loss of accuracy.

Bernard Ostle and J. M. Wiesen (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

$$f(x) = c\,(1+|x-\theta|^k)^{-h}, \quad -\infty<x<\infty,$$

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.

Forman S. Acton (1959) devotes the first chapter of his book on the analysis of straight-line data to the choice of a model. In Chapter 2, he discusses the classical model: x known without error; variance of y constant. In addition to the analytical method of least squares, he discusses the graphical least-squares fitting procedure of Askovitz (1957), applicable when the x values are equally spaced, and the method of Nair and Shrivastava (1942). He devotes much space to confidence limits on the slope and the intercept. In addition to methods based on the mean and the standard deviation, he presents methods based on various functions of order statistics (median, midrange, range, etc.) and distribution-free methods. In Chapter 5, he considers regression with both x and y subject to error, including maximum likelihood estimation of the regression coefficients. In Chapter 9, he deals with the use of transformations to stabilize the variability or to make the model additive. In Chapter 10, he discusses the rejection of unwanted data. In particular, he presents the Grubbs (1950) and Dixon (1951) criteria for the rejection of outliers.

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 mth range R_m, which is a certain linear transform of the mth 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 mth 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, $W = w/\sigma$, 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 $P = 0{\cdot}0001$, 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, 0·099, 0·9995, and 0·9999; and (3) a table of moments of the range [mean to 10 decimal places (11 significant figures), variance to 10 DP (10 SF), α_3 to 8 DP (8 SF) and α_4 to 7 DP (8 SF)] for samples of size $n = 2\ (1)\ 100$.

Morris Morduchow and Lionel Levin (1959) extend the work of Morduchow (1954) on fitting a straight line by the method of averages to the case of fitting a parabola to a set of n points ($n > 3$) with equally spaced abscissas. The n points are divided into three groups, in consecutive order of the abscissas, containing n_1, n_2, and n_3 points, where $n_1+n_2+n_3 = n$. A case of special interest is symmetric averaging, in which $n_1 = n_3$. As a measure of the greatest possible "inaccuracy" of the method of averages, the authors compute $\phi = (\sigma_s/\sigma_a)_{\min}$, where σ_s and σ_a are the standard deviations of the residuals for the method of least squares and the method of averages, respectively, for various groupings. For equal or almost equal size of the groups, ϕ approaches the value $2\sqrt{10/9} \doteq 0{\cdot}702$ asymptotically as $n \to \infty$. The optimum grouping is that with approximately $n_1 = n_3 = n/5$, $n_2 = 3n/5$, for which $1/\phi$ approaches $(5/4)\sqrt{5/6} \doteq 1{\cdot}142$ asymptotically, as compared with $(9/2\sqrt{10}) \doteq 1{\cdot}422$ for groups of equal size and the corresponding factor $2/\sqrt{3} \doteq 1{\cdot}152$ for fitting a straight line [see Morduchow (1954)]. (See Note at end)

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 $\nu \leqq 10$, the distribution of the standardized extreme deviate from the sample mean,

$$u = \max\left[(x_n-\bar{x})/\sigma,\ (\bar{x}-x_1)/\sigma\right]$$

and the corresponding studentized deviate, $t_n = \max\left[(x_n-\bar{x})/s_\nu,\ (\bar{x}-x_1)/s_\nu\right]$, where

$$x_1 \leqq x_2 \leqq \ldots \leqq x_n$$

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 per cent and 1 per cent points of t_n for $n = 2\,(1)\,10,\ 12$ and $\nu = 1\,(1)\,10$, and discusses the method of preparation of this table.

Rider (1959) derives the distribution of the rth quasi-range, $W_r = X_{n-r} - X_{r+1}$, where $x_1 \leqq x_2 \leqq \ldots \leqq x_n$ 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, $\pi^2/6 \doteq 1{\cdot}6449$ for $r = 0$.

Edward L. Stiefel (1959*a*) describes in detail the theoretical and numerical aspects of a linear Chebyshev approximation of a function $f(x)$ by given functions $\phi_1(x), \ldots, \phi_m(x)$ at specified points $x_1, \ldots, x_n$. The theoretical development leads to computational techniques for numerical solution of the problem. The author discusses the relation of this method to least squares and least qth power approximations and the special case of polynomial approximations, and gives the advantages and disadvantages of alternative computational procedures. Stiefel (1959*b*) considers the problem of finding a point $(x_1, \ldots, x_m)$ such that the maximum of the absolute values of the numbers $h_j = a_{jk}x_k + c_j$ ($j = 1, \ldots, n$; $k = 1, \ldots, m<n$) is a minimum. He shows that there is a unique such point if the Haar condition holds, i.e. if all of the m-rowed minors of the matrix (a_{jk}) have rank m. The constructive proof yields an effective computational method for finding this point. The author gives extensions of this result and connections with the least squares solution, and applies the results to the problem of finding polynomial approximations in one or more variables.

Harry Svensson (1959) shows that the expressions for the unknown coefficients in a curve-fitting process by the method of least squares are appreciably simplified if the values of the independent variable are symmetrically spaced about zero or if they can be transformed so as to satisfy this condition. He shows that fitting a third-degree polynomial becomes as simple as fitting a straight line in the general case. Moreover, if the values of the independent variable are equally spaced, so that they can be written as 0, ± 1, ± 2, ..., the curve-fitting process simplifies still more, since the sums of powers of this variable can be tabulated once for all; such a table for up to 50 observations is given. These simplifications are especially valuable if one does not have access to an electronic computer, and is forced to resort to computation by hand or on a desk calculator.

Harvey M. Wagner (1959) notes that in regression problems alternate criteria of " best fit" to least squares are least absolute deviations and least maximum deviations. He points out that linear programming techniques may be used to solve the latter two problems. In particular, if the linear regression relation contains p parameters, he formulates the problem of minimizing the sum of the absolute values of the "vertical" deviations as a p equation linear programming model with bounded variables and the problem of fitting by the Chebyshev (minimax) criterion as a standard-form $p+1$ equation linear programming model.

John Edward Walsh (1959) proposes a large-sample non-parametric criterion for rejection of outlying observations. Let $x_1 \leqq x_2 \leqq \ldots \leqq x_n$ 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.

Joseph L. Walsh (1959) considers the problem of fitting a polynomial

$$p_n(z) = a_0z^n+a_1z^{n-1}+\ldots+a_n$$

of degree n to a given function $f(z)$ defined on a closed bounded set E so as to minimize the sum of the pth powers of the differences $f(z)-p_n(z)$. The case of interest in the present study is the one in which the set E consists of a finite number of real points. He notes that this measure of approximation, most used in the case $p = 2$ (least squares), goes back to Fourier and Gauss, while Chebyshev explicitly used the measure of approximation

$$\max\,[|f(z)-p_n(z)|,\ z \text{ on } E]$$

although Kirchberger made the first general theoretical study of it. In the case $p = 2$, the polynomial is unique. Chebyshev approximation can be considered as the limit of least pth power approximation as $p\to\infty$, and is often denoted by $p = \infty$. The author reviews some results of Walsh and Motzkin (1957) for the cases $p = 1$, $1<p\leqq\infty$, and $0<p<1$, and presents some new results on extremal approximations.

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 $0 \leqq (\bar{x}-H)/H \leqq (b-a)^2/4ab$, 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 $H \leqq g \leqq \bar{x}$, 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, focusing 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 ($i = 1, 2, \ldots, n$, where n is the number of observations) such that $|z_i|>C\sigma$ where $z_i = y_i-\bar{y}$, $\bar{y} = \Sigma y_i/n$. 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.

Cheney and Goldstein (1960), following a conjecture by Stiefel (1959*a*), give an (untested) iterative algorithm for Chebyshev approximation. They also consider the problems of approximation by rational functions which are best in the "least squares" sense or in the "uniform" sense.

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 and 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 (1960*a*) defines odd location statistics T and even location-free statistics S by

$$T(x_1+h, x_2+h, \ldots, x_n+h) = T(x_1, x_2, \ldots, x_n)+h, T(-x_1, -x_2, \ldots, -x_n) = -T(x_1, x_2, \ldots, x_n),$$
$$S(x_1+h, x_2+h, \ldots, x_n+h) = S(x_1, x_2, \ldots, x_n), S(-x_1, -x_2, \ldots, -x_n) = S(x_1, x_2, \ldots, x_n)$$

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 and 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 (1960*b*) proves that if the distribution is symmetric about θ and ET exists, then $E\{T \mid S = s\} = \theta$, together with a multivariate extension useful in obtaining unbiased estimators of

$$\theta: e.g., (R_2M_1+R_1M_2)/(R_1+R_2),$$

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 (1960*a*) 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 $\alpha_4 = 9$, 3, 2·19, 2·14, 1·8 and 1·61 respectively. He finds that the adequacy of the asymptotic formula increases with α_4. Rider (1960*b*) 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.

Stiefel (1960) points out that there are two essentially different methods of attacking the Chebyshev problem of solving an inconsistent system of linear equations

$$(1)\ \eta_j = \sum_{k=1}^{m} a_{jk}x_k + c_j = 0,\ j = 1, 2, \ldots, n$$

in such a way that the "solution" x_k minimizes (2) $\zeta = \max | \eta_j |, j = 1, 2, \ldots, n$. The "*minimizing methods*" start with a trial solution x_k and use an algorithm for diminishing ζ during every step of an appropriate computational routine. Zukhovitskii (1951*a*) published such an algorithm requiring only a finite number of steps to solve the problem. Another approach is based on theorems of de la Vallée Poussin (1911), who used a "*reference*", i.e. a set of $(m+1)$ equations η_ρ chosen from among the given equations (1). The solution of the reduced Chebyshev problem (3) Minimize $\zeta^* = \text{Max} | \eta_\rho |$ can be computed explicitly, and ζ^* is then a *lower* bound of the minimal deviation ζ corresponding to the solution of the unrestricted problem (1), (2). Therefore "*maximizing methods*" are available whose strategy is to replace the reference by another one so that the reference-deviation ζ^* is raised. In one earlier paper [Stiefel (1959*b*)], the author published such an algorithm, called the *exchange-method*, which also reaches the final solution in a finite number of steps. In another [Steifel (1959*a*)], he advocated the construction of "*minimax-methods*" furnishing a lower and an upper bound for the desired minimum (2) and closing the gap between them during the computation, and Cheney and Goldstein (1960) constructed such an algorithm. In the present paper, Stiefel shows that many of the algorithms for solving the Chebyshev problem are closely related to the methods of linear programming, the main result being that Zukhovitskii's algorithm and the exchange method are completely equivalent to the simplex method of linear programming, and are duals of each other. He also shows that the exchange method, properly adapted to the computational routines for linear programming, is more economical than the simplex method.

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). (See Note at end)

J. Descloux (1961) generalizes the exchange algorithm developed by Stiefel (1959*a*) for discrete linear Chebyshev approximation in the nondegenerate case (i.e., when the Haar condition is satisfied) to include the degenerate case. He first demonstrates some properties of the approximations in the degenerate case, then presents a generalized exchange algorithm. He also gives some numerical examples.

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 centre 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 (1961*a*) derives locally best tests, based on the sample skewness $\alpha_3 = \sqrt{\beta_1}$ and the sample kurtosis α_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 (1961*b*) 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, in which he discusses trimming and Winsorization, to the relation between rejection and estimation.

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."

Walter D. Fisher (1961) reviews the formulation as a problem in linear programming of the statistical problem of fitting a multiple linear regression by minimizing the sum of absolute deviations from the regression function. He traces the history of the problem, beginning with the work of Fourier [(1824)] and including also the contributions of Edgeworth ([1887*c*, 1888], 1923), Rhodes (1930), Singleton (1940), Charmes *et al.* (1955), and Karst (1958).

Philip George Guest (1961), in his book on numerical methods of curve fitting, deals with the treatment of observations of a single variable (Part I), the fitting of straight lines (Part II), and the fitting of polynomials and other curves (Part III). Part I includes a discussion of the arithmetic mean, the variance, and other moments, also the range and its use in estimating standard deviation, and a section (3.7) on the rejection of outlying observations. In the latter, he cites the discussions of Brunt (1917) and Jeffreys (1939) [1948] and lists the tables designed for testing an outlying observation in the following cases: (*a*) Population mean and standard deviation known [Pearson and Hartley (1954), Table 24]; (*b*) Population standard deviation known [ibid., Table 25]; (*c*) Standard deviation estimated independently [ibid., Table 26; Nair (1952)]; (*d*) Standard deviation estimated from data set being tested [Grubbs (1950)]; (*e*) Range used instead of standard deviation [Dixon (1950, 1951)]. He also mentions the account of (*c*) and (*e*), with examples, given by Proschan (1953). In Parts II and III, he gives a thorough treatment, with numerical examples, of calculating schemes for fitting straight lines, polynomials, and other curves by the method of least squares, but does not mention alternative methods.

Harter (1961) gives examples of the use of tables of percentage points of the range [Harter and 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 and] 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.

Charles Lawrence Lawson (1961) makes numerous contributions to the theory of linear least maximum (Chebyshev) approximation; only those dealing with approximation to a finite set of real valued points will be summarized here. In Chapter 5 he establishes the following relationship between weighted *p*th power approximation and least maximum approximation. Under the hypothesis of relative unisolvence (which he defines in section 2.9), a least maximum approximator over a finite set of points is the limit of an infinite sequence of weighted least *p*th power approximators (for fixed finite $p>1$) where the weights are defined recursively. He writes (p. 6): "A satisfactory method for computing the least maximum approximator in a unisolvent family for a real valued function on a finite point set is the exchange algorithm which is based upon ideas introduced by Remez. . . . Stiefel (1960) has shown that this approximation problem can be formulated as a linear program and that the exchange algorithm can be interpreted as a variant of the simplex algorithm which takes advantage of special properties of this linear program. Computational algorithms can also be based upon . . . our results in Chapter 5. Where the exchange algorithm is applicable it has the advantage of reaching the solution after a finite amount of computation; however, when the function to be approximated is . . . vector valued the approximation problem is no longer a linear program and the exchange algorithm is not applicable." In Chapter 6, the author discusses his experience in computing a number of vector valued approximations using an algorithm based on the results of Chapter 5, including comparisons with a report by Goldstein *et al.* (1957) of computational experience with a different algorithm.

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_v of population variance in the denominator by the pooled estimate

$$s^* = \{[(n-1)\, s^2 + v s_v^2]/(n+v-1)\}^{\frac{1}{2}},$$

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 (1961*a*) 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 (1961*b*) 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 and Chandra Sekar (1936), the latter based on the work of Thompson (1935), and tabulates their 5 per cent and 1 per cent critical values for regression on m variables ($m = 1, 2, 3$).

(See Note at end)

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.

S. Vajda (1961), in his book on mathematical programming, includes a chapter on selected applications, which contains a section (8.4) on approximation to a function so as to minimize the largest (absolute) deviation. He works out a simple numerical example involving fitting a polynomial of degree 0, 1, 2, or 3 to four points by the minimax method and, for the sake of comparison, by the method of least squares. Obviously, both methods give the same result (the cubic curve passing through all four points) for degree 3. For degree 0, the minimax method yields the line y = midrange (y_i) and the method of least squares, the line y = mean (y_i) where the y_i ($i = 1, 2, 3, 4$) are the ordinates of the given points.

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 $x_1, x_2, \ldots, x_n$ from a number x is a minimum when x is the median of the x_i ($i = 1 \ldots, n$).

Odoardo Cucconi (1962) proposes a criterion for the rejection of outlying observations from a k-dimensional distribution ($k = 1, 2, 3, \ldots$) 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

$$[N/(N-1)] \sum_{i=1}^{k} \sum_{j=1}^{k} (\Delta_{ij}/\Delta)\, ({}_s x_i - m_i)\, ({}_s x_j - m_j),\; s = 1, 2, \ldots, N,$$

where ${}_r x_i$ ($r = 1, 2, \ldots, N$; $i = 1, 2, \ldots, k$) is the ith coordinate of the rth 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 ${}_r x_i$ and ${}_r x_j$ from their respective means m_i and m_j, and Δ_{ij} is the cofactor of c_{ij}. In order to test the hypothesis H_0 that the sth ($s = 1, 2, \ldots, N$) 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 $\alpha = 0{\cdot}05$, $0{\cdot}01$; $k = 1, 2, 3, 4$; and various values of N ranging from 5 to 100. (See Note at end)

Harter (1962), as part of a study of the ratio of two ranges not otherwise relevant to the present topic, tabulates (to 8 DP) the probability density function of the standardized range $W = w/\sigma$ 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 4 DP table of Sibuya and Toda (1957) at intervals of 0·05 in W. Harter's values for W a multiple of 0·05, when rounded to 4 DP, agree with those of Sibuya and 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: $Y_i = \alpha + \beta X_i + \varepsilon_i$ against H_1: $Y_i = \phi(X_i) + \varepsilon_i$, $i = 0, 1, \ldots, n$, where α, β and ϕ are unspecified and $\phi(x)$ is a non-linear convex function, the ε_i are independent identically distributed random variables with median zero and a continuous density function $f(\varepsilon)$ 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 and 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.

Günter Meinardus (1962) states that most of the important results on the theory of Chebyshev approximation can be deduced from two simple theorems in analysis, and that the theoretical development is advantageous because it makes possible a clear exposition of construction procedures. He describes simple construction procedures, and gives some numerical examples. He considers both the case of approximation to a continuous function and that of approximation to a finite set of points. In the latter case, which is the one of interest in the present study, he makes extensive use of the results of Stiefel (1959*b*).

J. E. L. Peck (1962) extends the usual least-squares procedure for curve fitting by the use of orthogonal polynomials to the situation where the desired curve is constrained to pass through certain fixed points, e.g. the origin. He states that his purpose is to show that polynomial curve fitting with such constraints is almost as easy as when there are no constraints; the same algorithm is used, but it must be applied twice, the second time with a different set of data. (See Note at end)

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 and 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 $z_i = (y_i - \dot{y})/a_{i|n}$ against i, where the y_i $(i = 1, 2, \ldots, n)$ are the residuals, $\dot{y}$ is their median, and $a_{i|n}$ is the standard normal deviate corresponding to the cumulative probability of the ith 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.

Wagner (1962) proposes a curvilinear regression model that does not require assuming that the regression functions have specific mathematical forms, but only that they are monotone or concave. He uses linear programming methods to fit the regression functions according to the criteria of minimal sum of absolute deviations and minimal maximum deviation, and sketches the alteration needed to provide a least squares fit.

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 misbehaviour of the data has been diagnosed, it is important, they say, to deal with it before seeking out other sorts of misbehaviour. 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 residual 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.

V. G. Ashar and T. D. Wallace (1963) report the results of a small sampling study (50 random samples each of size 20) of minimum absolute deviations (m.a.d.) estimators $\hat{b}_i$ of the parameters β_i $(i = 0, 1, 2)$ in the linear model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, subject to the Boscovich constraint $\bar{y} = \hat{b}_0 + \hat{b}_1 \bar{x}_1 + \hat{b}_2 \bar{x}_2$, where the errors are normally distributed but autocorrelated. They note that if the Gauss-Markov assumptions are satisfied by a set of data, m.a.d. estimators (or any other estimators than least squares) will come off second best, but that, since the solution procedure for m.a.d. estimates turns out to be the standard linear program, one may in certain cases wish to derive m.a.d. estimates because of the ease of placing additional linear constraints upon the parameters. Since their results showed that the efficiency of the empirical m.a.d. estimates was only 48·32 per cent with respect to empirical least squares, 42·56 per cent with respect to theoretical least squares, and 19·50 per cent with respect to theoretical maximum likelihood, they conclude tentatively that one should be prepared to give up considerable efficiency if he wishes to use m.a.d. estimates. They note, however, that Karst found no significant differences between the variances of m.a.d. and least squares estimates for 30 samples and one regression vector. They also conclude that, while ordinary least squares may be highly inefficient for autocorrelated errors, the criterion of minimum squared errors appears to yield considerably more efficient estimates than minimum absolute errors. Cases in which the errors are not normally distributed were not included in their study, so it is not surprising that m.a.d. estimators did not perform well.

Philipp J. Davis (1963) proves (Corollary 7.4.7. p. 140) that every overdetermined system of linear equations has at least one Chebyshev solution (i.e. a solution which minimizes the maximum of the individual discrepancies). He goes on to show that the solution may not be unique; either there is a unique solution or there are infinitely many solutions forming a convex set. We have already seen [Haar (1918), Cheney and Goldstein (1958)] that if the Haar condition is satisfied, the Chebyshev (minimax) solution is unique. Davis shows (Corollary 7.5.4. p. 142) that there is always a unique solution to the problem of minimizing the sum of the pth powers of the absolute deviations for $1 < p < \infty$. (See Note at end)

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

Patrick A. Gainer (1963) notes that it would be desirable to be able to calculate the effect of an additional observation on a previous estimate of the least-squares regression coefficients (or to remove the effect of an observation which has entered into the solution, but appears, upon later examination, to be "wild") without resorting to the lengthy process of matrix inversion. He derives simple recursion formulas for both adding and subtracting observations from the well-known least-squares normal equations by simple applications of matrix algebra. The starting point for the derivation is the matrix formulation of the well-known least-squares solution, and succeeding steps in the derivation do not require any further application of statistical theory. It is assumed that the errors of observation are uncorrelated.

Harter (1963) tabulates (to 8 DP) the probability integral of the standardized rth 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 $P = 0{\cdot}0001$, 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 ith and jth order statistics ($i \leqq j$) 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 $3/\pi \doteq 0{\cdot}955$ relative to the sample mean. (See Note at end)

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 ith 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 and 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 and 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.

T. J. Rivlin (1963) begins by stating the well-known fact that the system of linear equations (1) $\sum_{j=1}^{m} a_{ij}x_j + b_i = 0$ $(i = 1, \ldots, n)$, with $m<n$, has, in general, no solution. If we adopt the vector notation $x = (x_1, \ldots, x_m)$ and set (2) $r_i(x) = \sum_{j=1}^{m} a_{ij}x_j + b_i$ $(i = 1, \ldots, n)$, then a solution of (1) is an x that satisfies $r_i(x) = 0$, $i = 1, \ldots, n$. Since (1) may not have a solution, we relax our requirements and seek, as a generalization, an x which minimizes

$$(3)\ \| r(x) \| = \max_{i=1,\ldots,n} | r_i(x) |.$$

An x^* which satisfies $\| r(x^*) \| < \| r(x) \|$ for all x is called a Chebyshev solution of (1). The author makes a study of Chebyshev solutions and their use in obtaining Chebyshev approximations to a continuous function over an interval and to a discrete set of points. Most of the results which he obtains are not new, but are found, explicitly or implicitly, in the literature. He reviews the work of de la Vallée Poussin (1911), Haar (1918), Stiefel (1959*b*, 1960), and other authors. Among other results he shows that every set of equations of the form (1) has a Chebyshev solution and that if the Haar condition [that every $m \times m$ submatrix of $A = (a_{ij})$ be non-singular] is satisfied, the solution is unique and the minimal deviation $M = \| r(x^*) \|$ is attained for at least $m+1$ values of the index i.

John W. Tukey and Donald H. McLaughlin (1963) discuss trimming and Winsorization. Given n ordered observations $y_1 \leqq y_2 \leqq \ldots \leqq y_n$, their arithmetic mean is

$$\bar{y} = (y_1 + y_2 + \ldots + y_n)/n,$$

their g-times (symmetrically) trimmed mean is $y_{Tg} = (y_{g+1} + y_{g+2} + \ldots + y_{n-g})/(n-2g)$, and their g-times (symmetrically) Winsorized mean is

$$y_{Wg} = (g \cdot y_{g+1} + y_{g+1} + y_{g+2} + \ldots + y_{n-g} + g \cdot y_{n-g})/n.$$

Clearly both y_{Tg} and y_{Wg} pay less attention to extreme values than does $\bar{y}$, but y_{Wg} 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:

$$\mathrm{SSD}_{Tg} = (y_{g+1} - y_{Tg})^2 + (y_{g+2} - y_{Tg})^2 + \ldots + (y_{n-g} - y_{Tg})^2;$$

$$\mathrm{SSD}_{Wg} = g\,(y_{g+1} - y_{Wg})^2 + (y_{g+1} - y_{Wg})^2 + (y_{g+2} - y_{Wg})^2 + \ldots + (y_{n-g} - y_{Wg})^2 + g\,(y_{n-g} - y_{Wg})^2.$$

Wilks (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 $\alpha = 0{\cdot}010, 0{\cdot}025, 0{\cdot}050, 0{\cdot}100$; $k = 1, 2, 3, 4, 5$; and $n = 5\,(1)\,30\,(5)\,100\,(100)\,500$.

A. T. Berztiss (1964) gives a comprehensive survey, with an extensive bibliography, of the least squares fitting of polynomials, with emphasis on methods which are applicable when the data are not equally spaced. Besides the Euclidean (least squares) norm

$$\| E \|_2 = \left(\sum_{i=1}^{N} \varepsilon_i^2 \right)^{1/2},$$

where ε_i ($i = 1, \ldots, N$) is the deviation of the ith observed value from the fitted curve, he mentions the Gershgorin [least absolute values] norm $\| E \|_1 = \sum_{i=1}^{N} | \varepsilon_i |$ and the Chebyshev (minimax) norm $\| E \|_\infty = \max | \varepsilon_i |$. All three are special cases of the general vector norm $\| E \|_p = \left(\sum_{i=1}^{N} | \varepsilon_i |^p \right)^{1/p}$, the best estimate being the one satisfying the requirement

$$\| E \| = \text{minimum}.$$

Berztiss does not recommend the Gershgorin norm for general use, though he admits that it may be suitable in special cases. He states that the least squares and minimax norms are the popular ones. He opines that the Chebyshev criterion may be the one to use if the values can be observed with great accuracy. He notes, however, that the least squares criterion is the simpler to apply, and the least squares estimates are the maximum likelihood estimates if the distribution of errors is normal, which is usually assumed. (See Note at end)

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 and 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. (See Note at end)

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 $z<3{\cdot}41$, representing a truncation of more than 9 per cent.

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. $F(t) = (1-\varepsilon)\,\Phi(t)+\varepsilon H(t)$, $0 \leqq \varepsilon < 1$, 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 $x_1, x_2, \ldots, x_n$ be a random sample of size n, and let the estimator $T = T_n(x_1, x_2, \ldots, x_n)$ 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

$$\rho(t) = t^2/2 \text{ for } |t| < k,\ \rho(t) = k\,|t| - k^2/2 \text{ for } |t| \geqq k,$$

where k and ε are related by $\sqrt{2\pi}\,(1-\varepsilon) = \int_{-k}^{k} e^{-t^2/2}dt + (2/k)e^{-k^2/2}$. He also considers robust estimation of a scale parameter, which he finds more difficult and less satisfactory.

Oldřich Kropáč (1964) notes that, in practical applications of numerical methods, checking operations, which prove the correctness of the calculations just finished, are very important. He deals with the checking of the individual elements of the matrix of normal equations occurring in the smoothing of experimental data by the least squares method. He expresses the checks in matrix notation convenient for use on a digital computer; the checks can be performed in connection with the multiplication of two matrices.

K. V. Mardia (1964) obtains the exact distributions of extremes, ranges and midranges in samples from any multivariate population. He lets $(x_{1j}, \ldots, x_{kj})$, $j = 1, \ldots, n$, be a random sample from a k-variate continuous population with p.d.f. $f(x_1, \ldots, x_k)$, and denotes the minimum and maximum observations of the ith variate by X_i and Y_i respectively, the ith range by $R_i (= Y_i - X_i)$ and the ith midrange by $V_i (= \{Y_i + X_i\}/2)$, where $i = 1, \ldots, k$. He finds the distributions of $(X_1, \ldots, X_k, Y_1, \ldots, Y_k)$, $(R_1, \ldots, R_k)$ and $(V_1, \ldots, V_k)$, which reduce to the classical forms for $k = 1$.

Günter Meinardus (1964), in Part I of his book on approximation of functions, deals with linear approximation. In Section 1, he states the general linear approximation problem as follows (p. 1): "Let R be a normed linear space of elements $f, g, \ldots$ over the field of real or complex numbers, and let the symbol $\| f \|$ denote the norm of f. In addition, let V be a finite dimensional linear space of R. . . . For a given $f \in R$ determine an element $g \in V$ such that $\| g - f \| \leqq \| h - f \|$ for all $h \in V$." Sections 3–7 deal with Chebyshev approximation, which uses the minimax norm. The Haar uniqueness theorem is presented in subsection 3.2. Section 7 deals with numerical methods, and subsection 7.4 is devoted to the discrete case, which is the one of interest in the present study. In particular, the author reviews the work of Stiefel (1959*a*, *b*, 1960), who emphasized that, while the problem is one of linear programming, the exchange method, which solves the dual problem, is usually more effective than the simplex method. He notes that if the space V does not fulfil the Haar condition (all determinants formed from n rows of the $n \times N$ coefficient matrix of an overdetermined system of $N > n$ equations in n unknowns are different from zero), certain difficulties arise in the exchange algorithm and in the methods of linear programming, and the solution may not be unique; this case has been considered by Descloux (1961). Part II deals with non-linear approximation.

Mary G. Natrella (1964) compares the variances of the unique medians of samples of size m (for odd values of m) with those of the pseudo-medians (the averages of the two central observations) for m 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 m 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 and Pearson (1954), and test the accuracy of the approximation used by those authors.

John Rischard Rice (1964), in Chapter 1 of his first of two volumes on the approximation of functions, defines the L_p norm of a continuous function $f(x)$ on the unit interval as

$$L_p(f) = \left[\int_0^1 |f(x)|^p \, dx \right]^{1/p}, \; p \geqq 1,$$

and that of a function $g(x)$ defined on a finite point set $X = \{x_i \mid i = 1, 2, \ldots, m\}$ as

$$L_p(g) = \left[\sum_{i=1}^{m} | g(x_i)|^p \right]^{1/p}.$$

In sections 1–7, he considers the choice of norm, noting that until the advent of high-speed computers, the L_2 norm (least squares) was usually used because of its computational simplicity, even when it was not the most desirable choice. In Chapters 2, 3, and 4, the author considers in some detail the L_2 (least squares), L_∞ (Chebyshev or minimax), and L_1 (least absolute values) norms. He notes that the L_2 approximation is always unique and the L_∞ approximation is unique if the Haar condition is satisfied, but gives no analogous condition for the uniqueness of the L_1 approximation. In Chapter 6, he deals with computational methods. In particular, in section 6–6 he considers the method of descent for polytopes, and points out that there are two approximation problems (Chebyshev approximation and L_1 approximation, both on a finite set) which lead to the problem of finding the lowest point on a polytope. In section 6–8, he considers the method of ascent as an alternative to the method of descent for Chebyshev approximation. In section 6–9, he points out the close relation between some aspects of approximation theory and linear programming. He notes that Stiefel (1960) has shown that the method of descent [Zukhovitskii (1951*a*)] and the exchange method [Stiefel (1959*b*)], which is an ascent method, are duals of one another in the sense of linear programming, so that either can be used, but that the exchange method is computationally more efficient.

John R. Rice and John S. White (1964) report theoretical and experimental results of applying several of the L_p norms ($1 \leqq p \leqq \infty$) to the problem of determining one or two parameters (mean or regression coefficients) from data subject to one of several symmetric error distributions for several sample sizes. They state the following conclusions (p. 255): "There is one main point to be made in this paper: *The L_p norm one should use for smoothing and estimation depends on the distributions of the errors.* Furthermore there is a large variation in the effectiveness of the various norms *and no single norm is good (or even mediocre) in all situations.* For the important problem of smoothing and estimation in the presence of wild points, the L_1 norm appears to be markedly superior among the L_p norms, $1 \leqq p \leqq \infty$. In [Tukey (1962)] it is proposed to treat this situation by means of either trimming (removal of equal numbers of the highest and lowest values) or Winsorizing (replacing the k highest and lowest values by the $(k+1)$st high and $(k+1)$st low values, respectively). One then uses the L_2 norm for the modified data. For normally distributed errors these methods lead to estimates that are nearly as efficient as the L_2 estimates for the unmodified data. For distributions with long tails they are clearly much superior to L_2 estimates. These two methods were not compared with the L_1 estimates in this study, though this would be of interest. It should be noted that L_1 estimates from Winsorized and unmodified data are identical. The extension of trimming and Winsorization to smoothing and estimation involving several parameters is not simple while L_1 estimation is naturally defined in this case. There is yet another criterion for defining a 'best' estimate, one which is apparently unrelated to a norm. This is the criterion of *Maximum Likelihood,* see [Cramér (1946)]. The maximum likelihood estimate of a^* in [$y_i = a^*+e_i$, $i = 1, 2, \ldots, m$, where the e_i are random errors] depends on the distribution of the errors. If we have $1 \leqq p \leqq \infty$ and $dF(x)/dx = [c]\, e^{-|x|^p}$ [where c is a normalizing constant] then the L_p estimate of α^* is a maximum likelihood estimate of α^*. This relationship holds in the limit $p = \infty$. . . .While it is important to understand the relationship between the distributions $dF(x)/dx$ of the erors and the proper L_p norm, there is a serious difficulty in actual problems. In this paper it is assumed that $dF(x)/dx$ is known. In real problems one cannot determine $dF(x)/dx$ with any precision. It has already been pointed out that the presence of wild points indicates a long tail on $dF(x)/dx$. In real problems, the best one can usually do is determine if $dF(x)/dx$ has a long tail, if it is sharply truncated or if it lies somewhere in between."

Alex Rosengard (1964*a, 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 (1964*b*) 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 centre 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.

References (Glossary of Code Letters, page 41)

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**, 366–369. (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, MO, SD, RA, QN, LR, NR, LS, LF, ML)

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, AV, DI, AM, RA)

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)

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, AV, DI, AM, 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, MO, OS)

Eisenhart, Churchill; Deming, Lola S.; Martin, Celia S. (1948*a*). 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. (1948*b*). 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)

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, TO, 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, WM, 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. (1949*a*). 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. (1949*b*). The Truncated Mean in Moderately Large Samples. Memorandum Report No. 32, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE, AV, AM, RL, DA, WM).

Tukey, John W. (1949*c*). Scaling by and for Percentiles and Exponential Averages. Memorandum Report No. 33, Statistical Research Group, Princeton University, Princeton, New Jersey. (TE, AV, RL)

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

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

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

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

Mood, Alexander McFarlane (1950). *Introduction to the Theory of Statistics*. McGraw-Hill Book Co., Inc., New York-Toronto-London. (TE, AV, DI, LR, AM, MD, SD, IR, OS, LS, ML, BM)

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)

Scarborough, James B. (1950). *Numerical Mathematical Analysis* (Second Edition). Johns Hopkins Press, Baltimore. First edition, 1930; fourth edition, 1958. (TE, AV, AM, DI, SD, AD, LR, NR, LS, MA, SP, TO, WC)

Seth, G. R. (1950). On the ditribution 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)

Zukhovitskii, S. I. (1950). Certain Problems in the Theory of Chebyshev Approximations (Russian [or Ukrainian?]). Doctoral dissertation, Kiev University. (TE, LR, MM)

Bhate, D. H. (1951). 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. Univ. of California Press, Berkeley-Los Angeles. (TE, AV, MD, LR, OS, BM)

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)

Zukhovitskii, S. I. (1951*a*). An algorithm for the solution of the Čebyšev approximation problem in the case of a finite system of incompatible linear equations (Russian). *Doklady Akademii Nauk SSSR* (N.S.), **79**, 561–564. (TE, LR, MM)

Zukhovitskii, S. I. (1951*b*). An algorithm for finding the point of least deviation (in the sense of P. L. Čebyšev) from a given system of *m* points. (Ukrainian. Russian summary). *Dopovidi Akademii Nauk Ukrainskoi RSR*, **1951**, 404–407. (TE, AV, MD, MM)

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

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

Hald, A. (1952*b*). *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, BT)

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 Biometric Research*. John Wiley & Sons, Inc., New York. (TE, AV, DI, AM, MD, RA)

Shteinberg, A. S. (1952*a*). On the problem of the best uniform approximation for systems of incompatible linear equations and on the method of equation-gradient corrections (Ukrainian. Russian summary). *Dopovidi Akademii Nauk Ukrainskoi RSR*, **1952** (3), 167–174. (TE, LR, MM)

Shteinberg, A. S. (1952*b*). On Certain Methods of Solving the Problem of Best Uniform Approxmation for a System of Incompatible Linear Equations (Russian [or Ukrainian?]). Doctoral dissertation, Kiev University. (TE, LR, MM)

Bacon, Ralph Hoyt (1953). The "best" straight line among the points. *American Journal of Physics*, **21** (6), 428–446. (TE, LR, LS)

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 American 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)

Youden, W. J. (1953). Sets of three measurements. *The Scientific Monthly*, **77** (?), 143–147. (TE, TO, BT)

Zukhovitskii, S. I. (1953). On the best approximation, in the sense of P. L. Chebyshev, of a finite system of incompatible linear equations (Russian). *Matematicheskii Sbornik*, **75** [N.S. 33] (2), 327–342. (TE, LR, MM)

Agmon, Shmuel (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, **6**, 382–392. (TE, LR, MM)

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)

Morduchow, Morris (1954). Method of averages and its comparison with the method of least squares. *Journal of Applied Physics*, **25** (10), 1260–1263. (TE, LR, LS, MA)

Motzkin, T. S.; Schoenberg, I. J. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, **6**, 393–404. ([TE, LR], MM)

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)

Zitek, 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)

Charnes, A.; Cooper, W. W.; Ferguson, R. O. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, **1**, 138–151. (TE, LR, LF)

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

Chu, John T. (1955*b*). 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 absolute deviate in normal samples. *Journal of the American Statistical Association*, **50**, 185–195. (TE, OS, EX, TO, NC)

Hastings, Cecil, Jr. (1955). *Approximations for Digital Computers*. Princeton University Press, Princeton, N.J. ([TE, LR], MM)

Remez, E. Ya. (1955*a*). Method of grapho-analytical solution of certain problems of the Chebyshev approximation (Russian). *Ukrainskii Matematicheskii Zhurnal*, **7** (1), 71–90. (TE, LR, MM)

Remez, E. Ya. (1955*b*). On a grapho-analytic solution of certain problems of Chebyshev approximation (Russian). *Doklady Akademii Nauk SSSR*, **101** (3), 409–412. (TE, LR, MM)

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

Topping, J. (1955). *Errors of Observation and their Treatment*. The Institute of Physics, London. (TE, AV, DI, LR, NR, AM, MD, RA, AD, SD, LS)

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 programa [illegible] 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 [illegible]ion 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)

Ergun, Sabri (1956). Application of the principle of least squares to families of straight lines. *Industrial and Engineering Chemistry*, **48** (11), 2063–2068. (TE, LR, LS)

Goldstein, A. A. (1956). On the Method of Descent in Complex Domains and its Application to the Minimal Approximation of Overdetermined Systems of Linear Equations. Convair Astronautics, Mathematics Pre-print No. 1. (TE, LR, MM)

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

Remez, E. Ya. (1956*a*). On effective solution of a system of incompatible linear equations according to Chebyshev's principle of best uniform approximation. (Ukrainian. Russian summary.) *Dopovidi Akademii Nauk Ukrainskoi RSR*, **1956** (4), 315–320. (TE, LR, MM)

Remez, E. Ya. (1956*b*). Problems of uniqueness of multiplicity of solutions of the Chebyshev problem for a system of incompatible linear equations and the concept of a normal Chebyshev solution. (Russian.) *Ukrainskii Matematicheskii Zhurnal*, **8** (1), 34–53. (TE, AV, MM)

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)

Askovitz, S. I. (1957). A short-cut graphic method for fitting the best straight line to a series of points according to the criterion of least squares. *Journal of the American Statistical Association*, **52**, 13–17. (TE, DI, LR, AD, LS)

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)

Ghosal, 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)

Goldstein, Allan A.: Levine, Norman; Hereshoff, James E. [illegible] On the "best" and "least *Q*-th" approximation of an overdetermined system of linear equations. *Jour[illegible] of the Association for Computing Machinery*, **4**, 341–347. (TE, LR, MM, LP)

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

Head, J. W.; Oulton, G. M. (1957). Fitting curves to experimental data by least squares: An examination of the method of plotting experimental results. *Aircraft Engineering*, **24**, 268–270. (TE, LR, LS)

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)

Remez, E. Ya. (1957). *General Computational Methods of Chebyshev Approximation. The Problems with Linear Real Parameters.* (Russian.) Izdatel'stvo Akademii Nauk Ukrainskoi RSR, Kiev. English translation (1962) by U.S. Joint Publications Research Service. U.S. Atomic Energy Commission, Division of Technical Information, Oak Ridge, Tenn. (TE, LR, MM)

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)

Walsh, J. L.; Motzkin, T. S. (1957). Polynomials of best approximation on a real finite point set. *Proceedings of the National Academy of Science*, **43**, 845–846. (TE, LR, MM, LP)

Askovitz, S. I. (1958). Centroids, vectors, and least squares. *American Journal of Physics*, **26** (3), 164–168. (TE, LR, LS)

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)

Cheney, Ward; Goldstein, Allen A. (1958). Note on a paper by Zuhovickiĭ concerning the Tchebycheff problem for linear equations. *Journal of the Society for Industrial and Applied Mathematics*, **6** (3), 233–239. (Convair Astronautics, Mathematics Pre-print No. 8, 1957.) (TE, LR, MM)

Chou, Chan-Hui (1958). Least squares. *Industrial and Engineering Chemistry*, **50** (5), 799–802. (TE, LR, LS)

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

Goldstein, Allen A.; Cheney, Ward (1958). A finite algorithm for the solution of consistent linear equations and inequalities and for the Tchebycheff approximation of inconsistent linear equations. *Pacific Journal of Mathematics*, **8**, 415–427. (Convair Astronautics, Mathematics Pre-print No. 7, 1957.) (TE, LR, MM)

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].

Karst, Otto J. (1958). Linear curve fitting using least deviations. *Journal of the American Statistical Association*, **53**, 118–132. (TE, LR, LF)

Kelley, James E., Jr. (1958). An application of linear programming to curve fitting. *Journal of the Society for Industrial and Applied Mathematics*, **6** (1), 15–22. (TE, LR, LS, LP, MM)

Linnik, Yu. V. (1958). *Method of Least Squares and Principles of the Theory of Observations*. (Russian.) Fizmatgiz, Moscow. English translation by Regina C. Elandt (edited by N. L. Johnson). Pergamon Press, New York-Oxford-London-Paris, 1961 (TE, LR, NR, LS, MA, CM, TO, GS).

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)

Acton, Forman S. (1959). *Analysis of Straight-Line Data*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London. (TE, AV, DI, LR, AM, MD, MR, SD, RA, LS, NS, TO, GS, DC)

Geffroy, Jean (1959). Contribution à la théorie des valeurs extrêmes. 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 (1959). Die Verwendung der Variationsbreite bei der Schätzung der Standardabweichung und bei der Varianzanalyse ungeordneter Blockanlagen, sowie eine weitere Vereinfachung des Hartleyschen Verfahren zur direkten Bestimmung von Grenzdifferzen. *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 1 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].

Morduchow, Morris; Levin, Lionel (1959). Comparison of the method of averages with the method of least squares: fitting a parabola. *Journal of Mathematics and Physics*, **38**, 181–192. (TE, LR, NR, MA, LS)

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)

Stiefel, Edward L. (1959*a*). Numerical methods of Tchebycheff approximation. *On Numerical Approximation* (Proceedings of a Symposium, Mathematics Research Center, U.S. Army, University of Wisconsin, April 21–23, 1958; edited by R. E. Langer), pp. 217-232. University of Wisconsin Press, Madison. (TE, LR, LS, LP, MM)

Stiefel, E. (1959*b*). Über diskrete und lineare Tschebyscheff-Approximationen. *Numerische Mathematik*, **1**, 1–28. (TE, LR, LS, MM)

Svensson, Harry (1959). A useful simplification of the method of least squares. *Science Tools – The LKB Instrument Journal*, **6** (1), 1–3. (TE, LR, LS)

Wagner, Harvey M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, **54**, 206–212. (TE, LR, LS, LF, MM)

Walsh, J. E. (1959). Large sample non-parametric rejection of outlying observations. *Annals of the Institute of Statistical Mathematics* (Tokyo), **10**, 223–232. (TE, OS, TO, WR)

Walsh, Joseph L. (1959). On extremal approximations. *On Numerical Approximation* (Proceedings of a Symposium, Mathematics Research Center, U.S. Army, University of Wisconsin, April 21–23, 1958; edited by R. E. Langer), pp. 209–216. University of Wisconsin Press, Madison. (TE, LR, LS, LF, LP, MM)

Weiler, 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, WC, EM, BC, ST, TM, JA, GS, CT, AR)

Cheney, E. W.; Goldstein, A. A. (1960). Machine methods for Chebychev approximations. *Information Processing* (Proceedings of the International Conference on Information Processing, UNESCO, Paris, 15–20 June 1959), pp. 100–101. UNESCO, Paris; R. Oldenbourg, München; Butterworths, London. (TE, LR, LS, MM)

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *Annals of Mathematical Statistics*, **31**, 385–391. (TE, AV, DI, AM, WM, 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 A6–A8].

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

Hogg, Robert V. (1960*b*). 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. (1960*a*). 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. (1960*b*). Variance of the median of samples from a Cauchy distribution. *Journal of the American Statistical Association*, **55**, 322–323. (TE, AV, MD)

Stiefel, E. (1960). Note on Jordan elimination, linear programming and Tchebycheff approximation. *Numerische Mathematik*, **2**, 1–17. (TE, LR, MM)

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, DD, 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)

Descloux, J. (1961). Dégénérescence dans les approximations de Tschebyscheff linéaires et discrètes. *Numerische Mathematik*, **3** (3), 180–187. (TE, LR, MM)

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. (1961*a*). 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. (1961*b*). 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, WR, FC)

de Finetti, Bruno (1961). The Bayesian approach to the rejection 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)

Fisher, Walter D. (1961). A note on curve fitting with minimum deviations by linear programming. *Journal of the American Statistical Association*, **56**, 359-362. (TE, LR, LF)

Guest, P. G. (1961). *Numerical Methods of Curve Fitting*. Cambridge University Press, London-New York. (TE, AV, DI, LR, NR, AM, SD, RA, LS, TO, DC, GS, NC)

Harter, H. Leon (1961*a*). 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 (1961*b*). 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, WA, AM, MD, MA, ML, M4)

Lawson, Charles Lawrence (1961). Contributions to the Theory of Linear Least Maximum Approximation. Ph.D. Dissertation, University of California, Los Angeles. (TE, LR, NR, MM, LP, LS)

Quesenberry, C. P.; David, H. A. (1961). Some tests for outliers. *Biometrika*, **48**, 379–387. (TE, TO, TM, NC, GS, KC, AR)

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

Sen, Pranab Kumar (1961*b*). 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)

Vajda, S. (1961). *Mathematical Programming*. Addison-Wesley Publishing Company, Inc., Reading, Mass.-London. (TE, LR, NR, LS, MM)
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**, 131–139. (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. Bošković 1961*, pp. 19-25. Beograd-Zagreb-Ljubljana. (TE, AV, WA, AM, MD, LR, LF, LS)
Harter, H. Leon (1962). Percentage Points of the Ratio of Two Ranges and Related Tables. ARL 62-378. Wright-Patterson AFB. AD 286848. (TE, DI, RA) [See also Harter (1970), Volume 1: Chapter 1, Section 1 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)
Meinardus, Günter (1962). Über Tschebyscheffsche Approximationen. *Archive for Rational Mechanics and Analysis*, **9** (4), 329–351. (TE, LR, NR, MM)
Peck, J. E. L. (1962). Polynomial curve fitting with constraint. *SIAM Review*, **4** (2), 135–141. (TE, LR, NR, LS)
Rosengard, Alex (1962). Études 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**, 87–89. (TE, AV, DI, AM, MD, RA)
Siddiqui, M. M. (2962). 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, MR, RL, DA, WM, RA, TO, WI, TF)
Wagner, Harvey M. (1962). Non-linear regression with minimal assumptions. *Journal of the American Statistical Association*, **57**, 572–578. (TE, NR, LS, LF, MM)
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)
Ashar, V. G.; Wallace, T. D. (1963). A sampling study of minimum absolute deviations estimators. *Operations Research*, **11**, 747–758. (TE, LR, LS, LF)
Davis, Philipp J. (1963). *Interpolation and Approximation*. Blaisdell Publishing Co., Waltham, Mass. (TE, LR, LP, MM)
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)
Gainer, Patrick A. (1963). A Method for Computing the Effect of an Additional Observation on a Previous Least-Squares Estimate. NASA Technical Note D-1599. National Aeronautics and Space Administration, Washington. (TE, LR, LS)
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, RL, HL, TO, WI, AR)
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, FC, LA)
Lieberman, Gerald J.; Miller, Rupert G., Jr. (1963). Simultaneous tolerance intervals in regression. *Biometrika*, 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)
Rivlin, T. J. (1963). Overdetermined systems of linear equations. *SIAM Review*, **5**, 52–66. (TE, LR, MM)
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, WM)
Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhyā* (A) **25**, 407–426. (TE, TO, IC, TM, NC, DC, GS, FC, MW)
Berztiss, A. T. (1964). Least squares fitting of polynomials to irregularly spaced data. *SIAM Review*, **6** (3), 203–227. (TE, LR, NR, LS, LF, LP, MM)
Chew, 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, WR, FC)
Derman, 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, WA, 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, FC)
Huber, Peter J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101. (TE, AV, AM, MD, RL, WM, DA, HL, HU, DI, LR, LS, TO, AR)
Kropáč, Oldřich (1964). Kontrolní operace při vyrovnávání experimentálních údajů metodou nejmenších čtverců. (Russian and English summaries.) *Zpravodaj VZLU* (Letnany), **1** (43), 11–16. (TE, LR, NR, LS)
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)
Meinardus, Günter (1964). *Approximation von Funktionen und ihre Numerische Behandlung*. Springer-Verlag, Berlin-Heidelberg-New York. Expanded English translation (1967), *Approximation of Functions: Theory and Numerical Methods*, by Larry L. Schumaker. (TE, LR, NR, MM)
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, FC)
Pearson, E. S.; Stephens, M. A. (1964). The ratio of range to standard deviation in the same normal sample. *Biometrika*, **51**, 484–487. (TE, DI, SD, RA)
Rice, John R. (1964). *The Approximation of Functions, Volume I: Linear Theory*. Addison-Wesley Publishing Co., Reading, Mass.-Palo Alto-London. (TE, LR, NR, LS, LF, LP, MM)
Rice, John R.; White, John S. (1964). Norms for smoothing and estimation. *SIAM Review*, **6** (3), 243–256. (TE, LR, LS, LF, LP, MM, ML, TR, WI)
Rosengard, Alex (1964*a*). Indépendance limite uniforme de la moyenne et des valeurs extrêmes d'un échantilonl. *Comptes Rendus de l'Académie des Sciences* (Paris), **258**, 5786–5788. (TE, AV, AM, EX)
Rosengard, Alex (1964*b*). Loi-limite jointe d'un quantile et de la moyenne d'un échantillon. *Comptes Rendus de l'Académie des Sciences* (Paris), **259**, 2759–2761. (TE, AV, AM, QN)
Rosengard, Alex (1964*c*). Indépendance limite uniforme d'un quantile 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. (1964). A note on estimation from a Cauchy sample. *Journal of the American Statistical Association*, **59**, 460–463. (TE, AV, AM, MD, DA)

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)
BF Bartlett's (method of) fitting (straight lines)
BM Brown-Mood estimators (of regression parameters)
BT best two (out of three)
CC Chauvenet's criterion (for rejection of outliers)
CM Cauchy's method (of interpolation)
CT (Bliss)-Cochran Tukey criterion (for rejection of outliers)
CU Cucconi's criterion (for rejection of outliers)
DA discard averages (trimmed means)
DC Dixon's criterion (for rejection of outliers)
DD discard deviation
DH differences at half range
DI dispersion (measures of)
EA equal areas (under joint p.d. curve) (Laplace's "most advantageous method")
EM Edgeworth's modification (of Stone's second criterion)
EX extremes (largest and smallest values in sample)
FC Ferguson's criterion (for rejection of outliers)
GA Gastwirth estimators
GC Glaisher's criterion (for rejection of outliers)
GE geometric midrange
GG geometric range
GM geometric mean
GR Goodwin's rule (for rejection of outliers)
GS Grubbs' criterion (for rejection of outliers)
HA Hodges' alternative (to Hodges-Lehmann estimator)
HC Heydenreich's criterion (for rejection of outliers)
HL Hodges-Lehmann estimator
HM harmonic mean
HO Hogg's estimator
HS Hulme-Symms alternative (to the rejection of outliers)
HU Huber's estimator
IC Irwin's criterion (for rejection of outliers)
IR interquartile range
JA Jeffreys' alternative (to the rejection of outliers)
KC Kudô's criterion (for rejection of outliers)
LA Laurent's analogue (of Thompson's criterion)
LC linear combinations (of order statistics)
LD largest (absolute) deviation
LF least (sum of absolute) first (powers) [Laplace's "method of situation"]
LN least number of deviations (least sum of zero powers)
LP least (sum of) *p*th (powers of absolute values)
LR linear regression
LS least squares

LW linearly weighted means
MA method of averages
MC Merriman's criterion (for rejection of outliers)
MD median
MG method of group averages
MK McKay's criterion (for rejection of outliers)
ML maximum likelihood
MM minimax method (minimize maximum residual)
MO mode
MQ median-quartile average
MR midrange
MS method of successive differences
MT median and two other order statistics
MW multivariate Wilks' criterion (for rejection of outliers)
MZ Mazzuoli's criterion (for rejection of outliers)
M4 maximum (sum of) fourth powers (of p.d.f. of errors)
NC Nair's criterion (for rejection of outliers)
NM Newcomb's method (of treating outliers)
NR nonlinear regression
NS Nair-Shrivastava method (of curve fitting)
OM Ogrodnikoff's method (of treating outliers)
OS order statistics
PA plus approximative méthode (most approximative method)
PC Peirce's criterion (for rejection of outliers)
PM power means
QA quadratic average (mean)
QD quartile deviation (semi-interquartile range)
QM quasi-midrange (quasi-median)
QN quartiles
QR quasi-range
RA range
RC Rohne's criterion (for rejection of outliers)
RL robust estimators of location
RM range method
RS robust estimators of scale
SC Stone's (first) criterion (for rejection of outliers)
SD standard deviation [or variance $\equiv$ $(SD)^2$]
SM Stewart's method (criterion) (for rejection of outliers)
SP (method of) selected points
SR semirange
ST Student's rule (for rejection of outliers)
SW Switzer's estimator
S2 Stone's second criterion (for rejection of outliers)
TC Tippett's criterion (for rejection of outliers)
TD transformation of data (and choice of model)
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
TR trimming
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

Additional References added in Proof (see supplementary pages 43 and 44)

Bartlett, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, **5**, 207–212. (TE, LR, BF, MA, NS, LS)

Plackett, R. L. (1949). A historical note on the method of least squares. *Biometrika*, **36**, 458–460. (TE, LR, LS)

Madansky, Albert (1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, **54**, 173–205. (TE, LR, LS, ML, MG, NS, BF)

Birnbaum, Allan (1961). Some theory and techniques for robust estimation (preliminary report) (abstract). *Annals of Mathematical Statistics*, **32**, 622. (TE, AV, RL, LC, OS)

Teicher, Henry (1961). Maximum likelihood characterization of distributions. *Annals of Mathematical Statistics*, **32**, 1214–1222. (TE, AV, AM, DI, SD, ML)

Dixon, W. J. (1962). Rejection of observations. *Contributions to Order Statistics* (edited by Ahmed E. Sarhan and Bernard G. Greenberg), pp. 299–342. John Wiley & Sons, Inc., New York–London. (TE, AV, AM, MD, DI, SD, RA, TO, DC, GS, IC, MK, NC, TM)

Rice, John R. (1962). Tchebycheff approximation in a compact metric space. *Bulletin of the American Mathematical Society*, **68**, 405–410. (TE, LR, NR, MM)

Descloux, Jean (1963). Approximations in L^p and Chebyshev approximations. *SIAM Journal*, **11**, 1017–1026. (TE, LR, NR, MM, LP)

Freeman, Harold (1963). *Introduction to Statistical Inference*. Addison-Wesley Publishing Company, Inc., Reading, Mass.–Palo Alto–London. (TE, AV, AM, MD, LR, LS, LF, ML)

Laska, Eugene M. (1963). A General Theory of Robustness. Doctoral dissertation, New York University. University Microfilms, Inc., Ann Arbor, Mich. (TE, AV, DI, AM, RL, RS, OS)

Box, G. E. P.; Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society* (B), **26**, 211–243; discussion, 244–252.

Chipman, John S.; Rao, M. M. (1964). The treatment of linear restrictions in regression analysis. *Econometrica*, **32**, (1–2), 198–209. (TE, LR, LS)

Additions in Proof to Part III

M. S. Bartlett (1949) presents and illustrates a modification of the method of Wald (1940) for fitting a straight line when both variables are subject to error. This modification involves the use of three groups (containing, as nearly as possible, equal numbers of points) instead of two, the result being the line through the centro[illegible], the middle group with slope equal to that of the line joining the centroids of the two extreme groups, not the latter line itself, as proposed by Nair and Shrivastava (1942). The author shows that his method has, in general, greater efficiency than either of the other two methods.

R. L. Plackett (1949) summarizes the justifications by Laplace (1823), Gauss (1823) and Markoff (1900) of the method of least squares. He suggests that Gauss was the first to justify least squares as giving those linear estimates which are unbiased of minimum variance. He offers the opinion that Laplace and Gauss proved quite different theorems about least squares, that Gauss' justification is preferable, and that Markoff, who refers to Gauss' work, may perhaps have clarified assumptions there but proved nothing new. He points out that Gauss' [second] proof is valid for all values of the sample size n, entirely free from any assumption of normality.

Albert Madansky (1959) considers the situation where X and Y are related by $Y = \alpha+\beta X$, where α and β are unknown and where we observe X and Y with error, i.e. we observe $x = X+u$ and $y = Y+v$. He assumes that $Eu = Ev = 0$ and that the errors (u and v) are uncorrelated with the true values (X and Y). He points out that in order to use standard techniques (least squares or maximum likelihood) to estimate α and β, one needs additional information about the variances of the errors. He explores other methods, including various forms of the method of group averages, such as those of Nair and Shrivastava (1942) and Bartlett (1949), for which he studies the efficiency when various proportions of the observations are included in the groups.

Allan Birnbaum (1961) outlines some theory and techniques for robust estimation. He lets $f(x, \theta, \gamma)$ be the density function of sample point x, depending on real parameter θ and any (nuisance) parameter γ of specified ranges. For any given estimator $\theta^* = \theta^*(x)$ of θ, he lets $r(\theta, \gamma)$ denote the mean-squared-error (m.s.e.), and defines admissibility of θ (possibly in a restricted class of estimators) as usual. He calls θ^* robust (over the specified range of γ) if for each γ' there exists a corresponding estimator, admissible when $\gamma = \gamma'$ is known, with m.s.e. $\underline{r}(\theta, \gamma')$ such that $\underline{r}(\theta, \gamma')/(\theta, \gamma')$ is near unity for all θ. He calls an estimator *admissibly robust* if it is simply admissible. He characterizes admissibly robust invariant estimators of a location parameter θ as having a modified Pitman structure, and discusses problems of computing r's and attainable robustness. Under restriction to unbiased estimates linear in ordered observations, he characterizes admissibly robust estimators and illustrates fairly tractable theoretical and computational methods.

Henry Teicher (1961) proves three theorems characterizing translation parameter families and scale parameter families of absolutely continuous distributions by the maximum likelihood estimators of their parameters. If, for all random samples of sizes two and three from a translation parameter family $F(x-\theta)$ of absolutely continuous distributions with p.d.f. $f(x)$ lower semi-continuous at $x = 0$, a maximum likelihood estimator of θ is the sample arithmetic mean, then $F(x)$ is a normal distribution with mean zero. If, for samples of all sizes from a scale parameter family $F(x/\sigma)$, $\sigma > 0$, of absolutely continuous distributions with p.d.f. $f(x)$ continuous on $(0, \infty)$ and $\lim_{y \downarrow 0} [f(\lambda y)/f(y)] = 1$ for all $\lambda > 0$, a maximum likelihood estimator of σ is the sample arithmetic mean, then $F(x)$ is the exponential distribution, i.e. $f(x) = e^{-x}$, $x>0$ and $f(x) = 0$, $x \leq 0$. If, for samples of all sizes from a scale parameter family of absolutely continuous distributions $F(x/\sigma)$, $\sigma>0$, with p.d.f. $f(x)$ continuous on $(-\infty, \infty)$ and $\lim_{y \to 0} [f(\lambda y)/f(y)] = 1$ for all $\lambda > 0$, a maximum likelihood estimator of σ is

$$(n^{-1} \Sigma_{i=1}^{n} x_i^2)^{1/2},$$

then $F(x)$ is the normal distribution with mean zero and variance one.

W. J. Dixon (1962) discusses rejection of observations, tests for outliers, analysis of contaminated data, and departures from normality. After some general discussion, he enumerates some sampling assumptions [fixed sample size n; a proportion $1-\gamma$ of the observations come from $N(\mu, \alpha^2)$ and the remainder from either $N(\mu+\lambda\sigma, \sigma^2)$ or $N(\mu, \lambda^2\sigma^2)$; observations comprise one sample of size n or k samples of r_i $(i = 1, \ldots, k)$, with μ known or unknown, σ known or unknown, and (if σ is unknown) an independent estimate s available or not]. He then lists various criteria for rejection of outliers [χ^2-test, extreme deviation, range, modified F-tests, various ratios of ranges and subranges, and criteria based on k samples of equal size n]. He reproduces tables of percentiles for the various criteria. He then considers the power of the criteria and their performance in various situations, sampling from a contaminated population, attempts to remove bias, and the effects of contamination on mean, median, variance and range. He closes with a section on problems needing further research.

John R. Rice (1962) defines a new type of Tchebycheff approximation, the strict approximation, for a function of several variables defined on a finite point set. He shows that the strict approximation is unique, and describes a method of ascent algorithm of the 1 for 1 exchange type for determining it.

Jean Descloux (1963) shows that the strict approximation defined by Rice (1962) for functions of several variables defined on a finite set is the limit of the best approximation in L^p when $p \to \infty$. He gives a construction of the strict approximation which is slightly different from that of Rice but entirely equivalent to it. In either case, the idea is to select among best approximations one which can be considered 'best among the best'.

Harold Freeman (1963) shows [Example (*b*), page 255] that the maximum likelihood estimator of the location parameter of the Laplace (double exponential) distribution is the value (the sample median) from which the sum of the absolute deviations of the sample values is a minimum. He does not apply this principle to the problem of linear regression, which he treats only by the method of least squares. In Section 27-7 (pp. 280–281), he shows the relation between best linear unbiased estimators, obtainable either by the method of Lagrange multipliers or by application of the Gauss–Markov theorem [see Gauss (1823), Markov (1900) and Plackett (1949)], and linear regression. Letting x be a random variable with expectation defined at each value of t by the linear regression function $E(x|t) = \alpha+\beta t$, he shows that the conventional least-squares operational technique on observed data $(x_1, t_1), \ldots, (x_n, t_n)$ yields the best linear unbiased estimator of $\alpha+\beta t$. He says this is probably the strongest justification for the extensive use of least squares in regression analysis, since it involves no assumptions of normality and can readily be extended to multiple linear regression.

Eugene M. Laska (1963) notes that, according to the extended definition of robustness proposed by Tukey (1960), a robust unbiased estimator is one which is unbiased for all members of a class of alternative distributions which are specified as being possibly true and has variance fairly close to the smallest variance that could be obtained by any unbiased estimator under each of these alternatives. He studies this definition in detail for the special cases of linear unbiased estimators and unbiased estimators (not restricted to be linear) of location and scale parameters, and presents examples in the case where the underlying distribution is known or assumed to be either normal or double exponential. The robust estimators are functions of the ordered observations.

G. E. P. Box and D. R. Cox (1964) note that in the analysis of data it is often assumed that observations $y_1, y_2, \ldots, y_n$ are independently normally distributed with constant variance and with expectations specified by a model linear in a set of parameters θ. They make the less restrictive assumption that such a normal, homoscedastic, linear model is appropriate after some suitable transformation has been applied to the y's. They make inferences about the transformation and about the parameters of the linear model by computing the likelihood function and the relevant posterior distribution. They discuss the relation of their methods to earlier procedures for finding transformations, and illustrate their methods with examples. The authors and discussants cite the work of other authors on ways to recognize the need for a transformation prior to fitting a linear model and on the choice of a transformation when one is needed.

John S. Chipman and M. M. Rao (1964) develop a method for least squares estimation of the regression coefficients in multiple linear regression problems when they are assumed to be subject to a set of linear restrictions. The simplest type of linear restriction is the specification that one or more regression coefficients are equal to zero or to some other constant; by suitable transformation of variables, any linear restriction can be reduced to this type.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

(6) The Method of Least Squares and Some Alternatives. ~~Parts I-VI~~ Part III.

5. TYPE OF REPORT & PERIOD COVERED

(9) Journal Article ~~s~~

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

(10) H. LEON HARTER

8. CONTRACT OR GRANT NUMBER(s)

In-House

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Air Force Flight Dynamics Laboratory/FBRD
Air Force Wright Aeronautical Laboratories (AFSC)
Wright-Patterson AFB, OH 45433

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

~~61102F~~ (16) AF-7071
(17) 707102~~02~~

11. CONTROLLING OFFICE NAME AND ADDRESS

(11) 1975

12. REPORT DATE

~~Aug 1974 - Apr 1976~~

13. NUMBER OF PAGES

226 (12) 240p.

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

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

18. SUPPLEMENTARY NOTES

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

012 070

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

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

bpg

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

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

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

UNCLASSIFIED

19. Uniform distribution
Outliers
Robust estimation
Adaptive procedures

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

1-4 5 R
6