LA-3763

C.3

LOS ALAMOS SCIENTIFIC LABORATORY
of the
University of California
LOS ALAMOS • NEW MEXICO

Using Nonlinear Least-Squares Methods for Quantal Response and Sensitivity Analyses, Minimum Chi-Square Estimation, and Differential Equations

UNITED STATES
ATOMIC ENERGY COMMISSION
CONTRACT W-7405-ENG. 36

Printed in the United States of America. Available from
Clearinghouse for Federal Scientific and Technical Information
National Bureau of Standards, U. S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

LA-3763
UC-32, MATHEMATICS
AND COMPUTERS
TID-4500

LOS ALAMOS SCIENTIFIC LABORATORY
of the
University of California
LOS ALAMOS • NEW MEXICO

Report written: August 1, 1967

Report distributed: November 17, 1967

Using Nonlinear Least-Squares Methods for Quantal Response and Sensitivity Analyses, Minimum Chi-Square Estimation, and Differential Equations

by

Roger H. Moore
R. Keith Zeigler

USING NONLINEAR LEAST-SQUARES METHODS FOR QUANTAL RESPONSE AND SENSITIVITY ANALYSES, MINIMUM CHI-SQUARE ESTIMATION, AND DIFFERENTIAL EQUATIONS

by

Roger H. Moore and R. Keith Zeigler

ABSTRACT

A number of special statistical and mathematical techniques reduce to solving equations equivalent to those required by general least-squares procedures. In the statistical category are quantal response and sensitivity analyses and minimum chi-square estimation of parameters; in the mathematical, is the problem of determining constants for differential equations whose solutions are relations among observable variables. This report contains demonstrations of the equivalences and provides examples to which the methods were applied.

INTRODUCTION

Computational procedures in statistics, as in many other fields, are often straightforward but tedious. Modern computers, along with simplified programming methods, make possible application and re-evaluation of many hitherto complicated techniques. Nonlinear least-squares methods, often attributed to Gauss, were thoroughly outlined early in the 19th century but did not come into widespread use until the middle 1950's. An early computer program in this area by the authors of this report was reported elsewhere.[1]

Since the program's first availability, a number of computer installations have made it a part of their libraries, and it has been used routinely by many nonprofessional programmers as a research tool. This broad application has led the authors to consider many new uses for the program. This report discusses some of these new applications and indicates how three general problems, basically unrelated to least squares, can be solved by least-squares methods.

QUANTAL RESPONSE AND SENSITIVITY ANALYSES*

1. Background.

Quantal response and sensitivity methods are used in many scientific fields. They are concerned with the statistical analysis of data obtained by subjecting a test item to a known level of stimulus (an insect is given a certain amount of poison, an explosive is dropped a predetermined number of feet) and noting whether the test item responds (dies, explodes) to that level or does not respond (lives, does not explode). A series of such experiments is carried out on different test items

---

*Based on "Multivariate Quantal Response Analysis Using Regression Methods," by R. K. Zeigler and R. H. Moore, presented at the 126th Annual Meeting of the American Statistical Association, Los Angeles, California, August 15 - 19, 1966.

until several levels of stimulus are encountered at which some respond and some do not.

Perhaps the best-known technique is that of probit analysis.[2] The "up-and-down" method[3] is also widely used and has the advantage (compared with probit analysis) of being noniterative. These methods are both designed to estimate parameters by the method of maximum likelihood, an idea still employed in fairly recent investigations.[4] It is appropriate to mention still another estimating proedure,[5] that is designed for small samples and based on a stochastic process concept rather than maximum likelihood.

Maximum likelihood equations derived for evaluating quantal response data can be solved by nonlinear regression formulation, and the "treatment" to which the experimental units are subjected may be a mixture of individual treatments. These concepts will be illustrated by five examples, three of the univariate type and two bivariate.

2. Maximum Likelihood Estimation.

It is the following line of reasoning that leads to the maximum likelihood equations. Each test item is assumed to be randomly selected from some larger population. It is further assumed that each item has a level of stimulus below which it will not respond and above which it will respond. This level may be called the threshold of the item. Finally, the thresholds of the population of items are presumed to be distributed according to a density function $f(t; \alpha)$ where α is a vector of k parameters which are of interest. Thus, if a randomly selected item is subjected to a level of stimulus, say x, the probability that it will respond may be written

$$\Pr(t < x) = \int_{-\infty}^{x} f(t; \alpha)\, dt = F(x; \alpha). \tag{1}$$

It is common, and sometimes reasonable, to assume that the threshold density is normal, so that

$$\Pr(t < x) = \int_{-\infty}^{x} (2\pi)^{-1/2}\sigma^{-1} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]dt, \tag{1'}$$

and μ and σ are the parameters to be estimated. It is not our intent to discuss the choice of the appropriate form of Eq. (1) for a particular application. Probits,[2] normits,[6] and logits[7] all have their advocates and users.

The usual procedure leading to the likelihood estimates is to select a set of levels of stimulus $x_1, x_2, \ldots, x_i, \ldots, x_m$. A number of items, say n_i, are tested at the level x_i, and r_i of them respond. If the probability of a single item responding at x_i is $P_i(\alpha) = F(x_i; \alpha)$, then $p_i = r_i/n_i$ is an estimate of $P_i(\alpha)$. Since the tests are assumed to be independently performed, the probability of obtaining the m results $p_1, p_2, \ldots, p_i, \ldots, p_m$ can be expressed as

$$P = \prod_{i=1}^{m} \binom{n_i}{r_i} P_i^{r_i}(1 - P_i)^{n_i - r_i}, \tag{2}$$

where it must be remembered that P_i is a function of the parameters of the threshold density function.

The likelihood equations which must be solved are (See Ref. 2, Appendix II.)

$$\frac{\partial \ln P}{\partial \alpha_j} = \sum_{i=1}^{m} \frac{n_i(p_i - P_i)}{P_i(1 - P_i)} \frac{\partial P_i}{\partial \alpha_j} = 0, \qquad j = 1, 2, \ldots, k. \tag{3}$$

These equations generally are nonlinear in the parameters and are commonly solved by iterative methods. The straight-line fitting involved in probit, normit, and logit analysis is a device to provide initial estimates of the parameter vector α and to make each iteration somewhat more palatable. That such methods were and are effective, however, cannot be overlooked or overstated. A great deal of work was expended in preparing auxiliary tables for use when the calculations were being performed by hand or on a conventional desk calculator.

The variances and covariances of the maximum likelihood estimates may be obtained in the usual manner by inverting the matrix A whose element in the jth row and j'th column is

$$a_{jj'} = -E\left(\partial^2 \log P/\partial\alpha_j \partial\alpha_{j'}\right). \tag{4}$$

In practice, of course, this matrix is estimated by evaluating these second partial derivatives using the maximum likelihood estimates of the parameters.

3. Nonlinear Regression Estimation.

In the formulation of the general regression problem, it is common to assume that the data follow the model

$$y_i = g(z_i;\beta) + e_i, \quad i = 1, 2, \ldots, M, \tag{5}$$

where the y_i are observed random variables, z_i is a vector of associated known mathematical variables, β is a vector of unknown parameters to be estimated, and e_i is a random variable whose expected value is zero and whose variance is σ_i^2. Any two of the random variables, say $e_{i'}$ and $e_{i''}$ ($i' \neq i''$), are assumed to have zero covariance. If there are K parameters in the vector β, they may be estimated by minimizing the weighted sum of squares

$$Q = \sum_{i=1}^{M} W_i\left[y_i - g(z_i;\beta)\right]^2 \tag{6}$$

by solving the normal equations

$$\frac{\partial Q}{\partial \beta_j} = -2\sum_{i=1}^{M} W_i\left[y_i - g(z_i;\beta)\right]\frac{\partial g(z_i;\beta)}{\partial \beta_j} = 0, \qquad j = 1, 2, \ldots, K \tag{7}$$

for the parameters. The Gauss-Markoff theorem states that, if $g(z_i;\beta)$ is linear in β and $W_i = 1/\sigma_i^2$, the best linear unbiased estimate of β is obtained from the solution of Eqs. (7). Consequently, even when the model is nonlinear in the parameters, it seems to be standard practice to set the weight equal to the reciprocal of the variance.

The usual procedure for estimating the covariance matrix for the estimates of the regression parameters is to invert the matrix B whose elements are

$$b_{jj'} = \sum_{i=1}^{M} W_i \frac{\partial g(z_i;\beta)}{\partial \beta_j}\frac{\partial g(z_i;\beta)}{\partial \beta_{j'}}, \qquad j,j' = 1, 2, \ldots, K. \tag{8}$$

As with the maximum likelihood covariance matrix, the elements $b_{jj'}$ in practice are evaluated at the estimates of the parameters before the matrix is inverted.

Much effort in recent years has been devoted to obtaining improved methods of minimizing Eq. (6) when the parameters appear nonlinearly.[8,9] This effort has been fostered by dissatisfaction with the basic Gauss linearization procedure, and it has been stimulated by the availability of computers which make such investigations feasible. Most computer facilities, therefore, have available nonlinear regression programs designed to solve Eqs. (7). Useful programs of this type allow W_i to be a known function of the parameters, so that the values of the weights may be modified as the iterations proceed.

4. Equivalence Demonstration.

The equivalence of the likelihood equations (3) and the normal equations (7) is made clear by the three equivalences: $\left[P_i(\alpha) \longleftrightarrow g(z_i;\beta)\right]$, $[p_i \longleftrightarrow y_i]$, and

$$\left[W_i \longleftrightarrow \frac{n_i}{P_i(\alpha)\ (1 - P_i(\alpha))}\right].$$

The equivalence of the covariance matrices derives from expressions (4) and (8), and is seen by writing

$$\frac{\partial^2 \log P}{\partial\alpha_j\,\partial\alpha_{j'}} = -\sum_{i=1}^{M}\frac{n_i}{P_i(\alpha)[1 - P_i(\alpha)]}\frac{\partial P_i(\alpha)}{\partial\alpha_j}\frac{\partial P_i(\alpha)}{\partial\alpha_{j'}} + \text{terms involving } (p_i - P_i(\alpha)). \tag{9}$$

When the negative expectation of Eq. (9) is obtained, the terms involving $(p_i - P_i(\alpha))$ are zero for large samples.

5. Univariate Examples.

In the following examples, reference to the LASL program means the computer program used extensively at the Los Alamos Scientific Laboratory for least-squares problems.[1] Results are reported to the same number of figures used in the original examples, even though the LASL program was required to solve the least-squares equations to at least seven, and sometimes eight, significant figures. The convergence criterion in the LASL program when operated on an IBM 7094 (which has slightly more than eight-digit single-precision accuracy) required that the estimated corrections at the final iteration be less than 10^{-6} of the current values of the parameters.

Example 1. Perhaps the best known example (because of its appearance in a pioneering text) of probit analysis is concerned with the effect of a series of concentrations of rotenone sprayed on the chrysanthemum aphis (Ref. 2, pp. 25 - 55). The data are repeated in Table I.

Table I

Toxicity of Rotenone to Chrysanthemum Aphis

Concentration (mg/l)	Log Concentration (x_i)	n_i	r_i
10.2	1.01	50	44
7.7	0.89	49	42
5.1	0.71	46	24
3.8	0.58	48	16
2.6	0.41	50	6
0	--	49	0

As is typical of this type of experiment, the assumption was made that the common logarithms of the thresholds were normally distributed with mean, μ, and standard deviation, σ. The LASL program was given the concentrations of Table I and computed the common logarithms for the remainder of the calculation. The estimates reported in the reference gave (after some manipulation) $\hat{\mu}_F = 0.686(\hat{\sigma}_{\hat{\mu}_F} = 0.0220)$ and $\hat{\sigma}_F = 0.239(\hat{\sigma}_{\hat{\sigma}_F} = 0.0267)$, while the LASL program gave $\hat{\mu}_L = 0.685(\hat{\sigma}_{\hat{\mu}_L} = 0.0221)$ and $\hat{\sigma}_L = 0.237(\hat{\sigma}_{\hat{\sigma}} = 0.0271)$ directly. The discrepancies clearly are small and attributable to differing degrees of precision in the performance of the two methods of calculation.

Example 2. This is an example of a situation in which the values of x_i cannot be preselected. A tabulation of such a set of results[10] is given in Table II. The data come from an experiment concerned with determining the penetration characteristics of a projectile. The penetrating velocities were assumed to be normally distributed as in model (1').

Table II

Velocities and Conditions of Impact of a Given Projectile Fired at a Given Armour Plate

Velocity (f/s)	Condition of Impact
2433	Non-Penetration ($n_1 = 1, r_1 = 0$)
2415	Non-Penetration ($n_2 = 1, r_2 = 0$)
2415	Non-Penetration ($n_3 = 1, r_3 = 0$)
2453	Penetration ($n_4 = 1, r_4 = 0$)
2423	Penetration ($n_5 = 1, r_5 = 0$)

Using initial estimates of $\mu_0 = 2435$ and $\sigma_0 = 17$, the reported results were $\hat{\mu}_G = 2431.6(\hat{\sigma}_{\hat{\mu}_G} = 10.7)$ and $\hat{\sigma}_G = 15.0(\hat{\sigma}_{\hat{\sigma}_G} = 12.5)$ with an estimated covariance of 50.8. The LASL program gave $\hat{\mu}_L = 2431.6(\hat{\sigma}_{\hat{\mu}_L} = 10.6)$ and $\hat{\sigma}_L = 14.9(\hat{\sigma}_{\hat{\sigma}_L} = 12.0)$ with an estimated covariance of 46.6.

Example 3. The up-and-down method[3] is illustrated with a set of "mocked-up" data based on a sample of size 60 from a normal distribution with $\mu = 1.3$ and $\sigma = 0.2$. The observations are repeated in Table III, where the term "normalized height" is used to indicate that these data might have come from an experiment in which explosives are dropped certain distances and a response is an explosion.

Table III

Demonstration "Data"

Normalized Height	Number of Explosions	Number of Non-Explosions
2.0	1	0
1.7	10	0
1.4	18	9
1.1	2	18
0.8	0	2

The results were $\hat{\mu}_D = 1.32(\hat{\sigma}_{\hat{\mu}_D} = 0.035)$ and $\hat{\sigma}_D = 0.17(\hat{\sigma}_{\hat{\sigma}_D} = 0.039)$, while the LASL program gave $\hat{\mu}_L = 1.32(\hat{\sigma}_{\hat{\mu}_L} = 0.035)$ and $\hat{\sigma}_L = 0.17(\hat{\sigma}_{\hat{\sigma}_L} = 0.038)$.

6. Bivariate Examples.

The procedures outlined in Sections 2 and 3 above are not limited to data obtained from simple treatments. A treatment may be a combination of simple treatments, such as two kinds of poison applied as a mixture, or dropping distance, humidity, and temperature used in testing sensitivity of explosives. To apply these procedures, it is necessary to arrive at a formulation involving the combination of treatments that is similar to Eq. (1). Thus, one must be able to write

$$\mathcal{E}(p_i) = f(x_{i1}, \ldots, x_{im}; \alpha_1, \ldots, \alpha_k),$$

$$i = 1, \ldots, n \tag{10}$$

where $\mathcal{E}(p_i)$ is the expected value of the observed fraction responding to the ith treatment combination, $(x_{i1}, \ldots, x_{im})$ is the set of m simple treatment levels that compose the ith treatment combination, and $(\alpha_1, \ldots, \alpha_k)$ are the parameters to be estimated.

Example 4. Consider a portion of the data presented elsewhere (Ref. 11, pp. 534 - 535, Table 2), which are given here as Table IV in a somewhat different form. The expected value of the fractions responding to a combination of levels of the two poisons was equal to one minus the incomplete bivariate normal distribution using the common logarithms of the simple treatment levels; that is,

$$\mathcal{E}(p_i) = 1 - \int_{-\infty}^{\log x_{i1}} ds \int_{-\infty}^{\log x_{i2}} g(s,t)dt, \tag{11}$$

where

$$g(s,t) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{s-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{s-\mu_1}{\sigma_1}\right)\left(\frac{t-\mu_2}{\sigma_2}\right)+\left(\frac{t-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

and where the parameters of Eq. (10) were identified arbitrarily with those of Eq. (11) by the pairs (α_1, μ_1), (α_2, σ_1), (α_3, μ_2), (α_4, σ_2), and (α_5, ρ). The estimates of the five parameters obtained from a LASL least-squares analysis of the data were $\hat{\mu}_{1L} = -0.9949$, $\hat{\sigma}_{1L} = 0.3450$, $\hat{\mu}_{2L} = -0.8333$, $\hat{\sigma}_{2L} = 0.4949$, and $\hat{\rho}_L = -0.7380$. (The estimated standard errors were, respectively, 0.0251, 0.0332, 0.0316, 0.0565, and 0.1125.) The original procedure,[11] using the mixture data, provided an estimate only of the correlation coefficient, $\hat{\rho}_H = -0.74$, with (what appears to be) a standard error of 0.09. Using the LASL program's results, it was found that the weighted sum of squares was 12.17. This was not significantly different from 19 (using chi-square critical values of 8.91 and 32.87 for (24 - 5) = 19 degrees of freedom at the 0.05 level of significance). Consequently, the fit could be considered satisfactory. The last two columns of Table IV give, respectively, the observed and estimated proportions responding to the treatment combinations.

Example 5. This example is provided by a problem involving the quest for an empirical function to describe the probability of survival of humans who had received whole-body irradiation, when no two individuals had identical treatments. Data were available on 104 subjects who had received radiation for a period of 2 weeks or longer.[12] The radiation received was measured in midline rads per week. The observation made on each subject was whether he was alive or dead 2 months after the cessation of treatment. Thus, for each of the 104 subjects, there was available the triplet (r_i, t_i, p_i) where r_i is the midline rads/week, t_i is the total treatment time, and $p_i = 0$ or 1 according to whether the subject was alive or dead 2 months after treatment. For the subjects considered, r_i ranged from 0.3 to 184.8 rads/week and t_i ranged from 2 to 216 weeks.

For various reasons, it was felt that the probability of survival could be expressed in the form

$$P_i = \exp\,[w(r_i)\ t_i]\ , \tag{12}$$

where $w(r)$ was an unknown function of r that had to be determined. To accomplish this, the data were divided into six sets according to r_i lying in the intervals (0-20), (20-40), (40-80), (80-120), (120-160), and (160-200). Each of the sets was fitted to the function

$$P_i = \exp\,(\alpha t_i), \tag{13}$$

with the six values of the estimates of α being approximately -0.0020, -0.0023, -0.0529, -0.1751, -0.1739, and -0.2339. These numbers were associated, respectively, with the numbers 10, 30, 60,

Table IV

Six-Day Toxicity to Beetles (<u>Tribolium castaneum</u>) of Direct Sprays

Pyrethins, D.D.T., and the Two Together, in Shell Oil P31

Deposit		Treatment Combination					
Insecticide	(mg./10 sq.cm.)	x_{i1}	x_{i2}	n_i	r_i	p_i	$\hat{P}_i$
1.2 o/o w/v pyrethins	2.52	0.03024	0	48	3	0.0625	0.0642
	3.30	0.03960	0	48	3	0.0625	0.1189
	4.25	0.05100	0	50	9	0.1800	0.1943
	5.33	0.06396	0	50	16	0.3200	0.2819
	7.15	0.08580	0	50	20	0.4000	0.4178
	9.53	0.11436	0	50	30	0.6000	0.5613
	12.28	0.14736	0	49	37	0.7551	0.6820
	15.58	0.18696	0	50	35	0.7000	0.7802
2.0 o/o w/v D.D.T.	2.45	0	0.0490	49	8	0.1633	0.1678
	3.18	0	0.0636	50	8	0.1600	0.2315
	4.25	0	0.0850	50	16	0.3200	0.3158
	5.48	0	0.1096	50	21	0.4200	0.3988
	7.24	0	0.1448	50	25	0.5000	0.4952
	9.54	0	0.1908	50	28	0.5600	0.5910
	12.36	0	0.2472	50	35	0.7000	0.6763
	15.54	0	0.3108	50	37	0.7400	0.7448
1.2 o/o pyrethins + 2.0 o/o D.D.T.	2.74	0.02964	0.0494	50	14	0.2800	0.2307
	3.20	0.03840	0.0640	49	18	0.3673	0.3442
	4.10	0.04920	0.0820	50	22	0.4400	0.4839
	5.34	0.06408	0.1068	50	36	0.7200	0.6553
	7.11	0.08532	0.1422	50	42	0.8400	0.8287
	9.60	0.11520	0.1920	50	45	0.9000	0.9454
	12.45	0.14940	0.2490	50	50	1.0000	0.9867
	15.65	0.18780	0.3130	50	50	1.0000	0.9974

100, 140, and 180 which were the midpoints of the arbitrary grouping intervals. A plot of these pairs of numbers on log-log graph paper revealed that a straight line connected them reasonably well. Consequently, it was decided that w(r) might have the form

$$w(r) = \beta_1 r^{(\beta_2)} . \tag{14}$$

Combining Eqs. (12) and (14), the function

$$P_i = \exp \left[\beta_1 r_i^{(\beta_2)} t_i\right] \tag{15}$$

was obtained. The entire set of 104 observations was submitted to the LASL program with the final estimates (and standard errors) of $\hat{\beta}_1 = -0.000241$ (0.000203) and $\hat{\beta}_2 = 1.36$ (0.21). The weighted sum of squares, defined by

$$S = \sum_{i=1}^{104} W_i \left(p_i - \hat{P}_i\right)^2 , \tag{16}$$

where $1/W_i = \hat{P}_i(1 - \hat{P}_i)$ since $n_i = 1$ for all cases, had a value of 122.0. Comparing this at the 0.05 level of significance with the expected value of a chi-square variable with 102 degrees of freedom, S was found not suspiciously large. Hence, the fit was considered satisfactory for these data.

7. <u>Comments.</u>

One immediate conclusion that may be reached is that a reasonably general least-squares computer program can replace several specialized quantal analysis programs, as long as the general program has the capability of convenient specification of the function being fitted. It must also allow the modification of weights at appropriate stages of the computation. However, it is not necessary to specify the algorithm by which the least-squares solution is attained. For instance, it is not necessarily required that the derivatives be analytically computed. Indeed, direct search methods often are successful when conventional procedures fail.

It is worth noting that nothing in the preceding development requires n_i to be larger than one. Thus, levels at which only a single item is tested may be incorporated into the computation. However, experimental data, in order to provide unique parameter estimates, must exhibit at least one "cross-over" level; i.e., there must be at least one level of stimulus at which there is a lack of response which is higher than the lowest level at which response does occur.

The relationship between maximum likelihood and least-squares estimates has been noted previously.[13] However, that discussion seems somewhat clouded by the requirement that the function be linear in the parameters and by the introduction of several forms of functions to be minimized, all of which are given the generic name χ^2.

MINIMUM CHI-SQUARE ESTIMATION*

1. A Problem.

Suppose the data (Ref. 14, p. 439, a portion of Table 30.4.2) appearing in Table V are given.

Table V

Distribution of Mean Temperatures for June in Stockholm, 1841-1940

Degrees Celsius	Years Observed
-12.4	10
12.5-12.9	12
13.0-13.4	9
13.5-13.9	10
14.0-14.4	19
14.5-14.9	10
15.0-15.4	9
15.5-15.9	6
16.0-16.4	7
16.5-	8

It is desired to estimate the mean and standard deviation of the distribution and to test the hypothesis that the data are "normally distributed."

*Based on "Minimum Chi-Square Estimation with Non-Linear-Least Squares Procedures," by Roger H. Moore, presented at the Spring Meeting of the Rio Grande Chapter of the Association for Computing Machinery, Cloudcroft, New Mexico, April 21-22, 1966.

2. Usual Solution.

It may be determined (Ref. 14, p. 440) that "....the exact class intervals are 12.45, 12.95, etc." Hence, one may estimate the parameters by the standard method (See Ref. 15, pp. 14-27) of setting up a frequency distribution, coding the midpoints of the intervals, and correcting the coded estimates to obtain the final estimates. After application of this method to the data of Table V, Table VI is obtained.

Table VI

Computing Layout of Data in Table V

Mid-point of Interval X_i	Coded Score x_i	Frequency f_i	$f_i x_i$	$f_i x_i^2$
11.70	-5	0	0	0
12.20	-4	10	-40	160
12.70	-3	12	-36	108
13.20	-2	9	-18	36
13.70	-1	10	-10	10
14.20	0	19	0	0
14.70	1	10	10	10
15.20	2	9	18	36
15.70	3	6	18	54
16.20	4	7	28	112
16.70	5	8	40	200
17.20	6	0	0	0
Total		100	10	726

The additional interval at either end is introduced to conform to the statement (Ref. 14, p. 438) that "We first assume that the grouping has been arranged such that the two extreme classes do not contain any observed values." Thus, $f_1 = f_{12} = 0$.

Letting $i' = 0.50$ be the interval width and $X_0 = 14.20$ correspond to $x = 0$, this layout provides for the following computations:

$$N = \Sigma f_i = 100$$

$$\bar{x} = (\Sigma f_i x_i)/N = 10/100 = 0.10$$

$$\bar{X} = i'\bar{x} + X_0 = (0.50)(0.10) + 14.20 = 14.25$$

$$s_x^2 = \left[\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2/N\right]/(N - 1)$$

$$= [726 - (10)^2/100]/99$$

$= 725/99 = 7.3232$

$s_x = 2.706$

$s_X = i's_x = (0.50)(2.706) = 1.353.$

It must be emphasized at this point that the parameters should be estimated from the ungrouped data if they are available.

The question of whether the data have come from a normal distribution may be answered by performing a goodness-of-fit test (Ref. 15, pp. 226-227). This is accomplished by comparing the observed frequencies with those obtained theoretically under the assumption of normality. Table VII is then established.

Table VII

Computational Layout for Goodness-of-Fit of Data in Table V

Upper Endpoint E_i	$\frac{E_i - \bar{X}}{s_X}$	Theoretical Frequency F_i	Observed Frequency f_i	$\frac{(f_i - F_i)^2}{F_i}$
12.45	-1.330	9.18	10	0.07
12.95	-0.961	7.65	12	2.47
13.45	-0.591	10.90	9	0.33
13.95	-0.222	13.48	10	0.90
14.45	0.148	14.67	19	1.28
14.95	0.517	13.87	10	1.08
15.45	0.887	11.50	9	0.54
15.95	1.256	8.30	6	0.64
16.45	1.626	5.25	7	0.58
∞	∞	5.20	8	1.51
		100.00	100	9.40

The statistic obtained by summing the last column of the table is to be compared with a critical value obtained from a chi-square distribution with (10 - 3) = 7 degrees of freedom since there are 10 intervals, nine of which have independent probabilities associated with them, and two parameters have been estimated from the data. For example, if the 5% level of significance is chosen, the critical value is 14.07. Since 9.40 is less than 14.07, the hypothesis of normality is accepted.

Thus, the problem posed earlier is answered. However, some questions remain: Would the conclusion be the same if the ungrouped parameter estimates had been used? Suppose 100 (instead of 99) had been used in the computation of s_X? Should Sheppard's corrections (Ref. 14, p. 438) have been applied? Are these estimates "best"?

3. Minimum Chi-Square Solution.

Often, it is not possible to recover the raw data, and one must rely on grouped data. Sometimes, as in sorting or sieving operations, there is no chance at all to use individual measurements. How, then, may estimates be obtained?

One way of doing this is to obtain values, say $\bar{X}^*$ and s^*, that minimize the sum of the entries in the final column of Table VII. These are called minimum chi-square estimates and are detailed elsewhere (Ref. 14, pp. 424-441).

To obtain these estimates, one must minimize

$$\chi^2 = \sum_i \frac{\left[f_i - F_i(\mu, \sigma)\right]^2}{F_i(\mu, \sigma)}$$

by finding the values of μ and σ for which $\partial\chi^2/\partial\mu$ and $\partial\chi^2/\partial\sigma$ are both zero. These values may be denoted by $\bar{X}^*$ and s^*.

That solving such equations is not necessarily a simple task may be demonstrated by noting, for instance, that

$$\frac{\partial\chi^2}{\partial\mu} = -\sum_i \left[2\,\frac{(f_i - F_i)}{F_i} + \frac{(f_i - F_i)^2}{F_i^2}\right] \frac{\partial F_i}{\partial\mu} .$$

4. The Least-Squares Approach.

The way out of this dilemma lies in applying the observation (Ref. 14, p. 425) that it can be "...shown for large N the influence of the second term within the brackets becomes negligible." Then it becomes a matter of solving the simpler system

$$\frac{\partial\chi^2}{\partial\mu} = -2\sum_i \frac{(f_i - F_i)}{F_i}\,\frac{\partial F_i}{\partial\mu} = 0$$

$$\frac{\partial\chi^2}{\partial\sigma} = -2\sum_i \frac{(f_i - F_i)}{F_i}\,\frac{\partial F_i}{\partial\sigma} = 0$$

for μ and σ. This technique is called (Ref. 15, p. 426) the "modified χ^2 minimum method."

It happens that these are precisely the kind of equations that must be solved in a general least-squares problem when it is recognized that the weights in such a problem may be regarded as the reciprocals of the function being fitted. To see this, consider a simple two-parameter least-squares problem where the function to be minimized is

$$Q(\beta_1, \beta_2) = \sum_i W_i \left[y_i - g(x_i; \beta_1, \beta_2)\right]^2.$$

The normal equations are of the form

$$\frac{\partial Q}{\partial \beta_j} =$$

$$-2 \sum_i W_i \frac{\partial g(x_i; \beta_1, \beta_2)}{\partial \beta_j} \left[y_i - g(x_i; \beta_1, \beta_2)\right] = 0,$$

$$j = 1, 2.$$

That these have the same form as those of the minimum chi-square method may be seen by using f_i for y_i, F_i for $g(x_i; \beta_1, \beta_2)$, and $1/F_i$ for W_i. The trick lies in having a nonlinear least-squares computer program that allows one to modify the weights at each iteration. One such program containing this provision is available.[1]

When this approach is applied to the data in Table V, the estimates $\bar{X}^* = 14.23$ and $s^* = 1.512$ are obtained and Table VIII may be derived.

Table VIII

Results of Applying Modified Minimum Chi-Square to Data in Table V

Upper Endpoint	F_i	f_i	$\frac{(f_i - F_i)^2}{F_i}$
12.45	11.91	10	0.31
12.95	7.89	12	2.15
13.45	10.41	9	0.19
13.95	12.34	10	0.44
14.45	13.13	19	2.63
14.95	12.53	10	0.51
15.45	10.73	9	0.28
15.95	8.24	6	0.61
16.45	5.68	7	0.30
∞	7.15	8	0.10
	100.01	100	7.52

It will be noted that the sum of the entries in the last column is considerably smaller than that obtained using the frequency distribution estimates and shown in Table VII. Indeed, this is slightly better than the $\chi^2 = 7.86$, along with a mean of 14.23 and a standard deviation of 1.574, reported with the original data (Ref. 14, p. 439).

5. Conclusion.

When a general least-squares program is available, it is suggested that minimum chi-square estimates are as easy to obtain as any other type. This removes the problem of subjectively choosing the proper computing procedure when data are already grouped. Indeed, there are some sets of data which demand this kind of treatment because individual measurements simply are not available. In addition, the minimum chi-square method provides the proper statistic for testing the hypothesis of normality in an unequivocal manner that gives every consideration to the hypothesis itself. Finally, the method may be applied to any form of the F_i that may be required. It is satisfactory even when F_i and its partials are not expressible in closed form and must be obtained numerically.

DIFFERENTIAL EQUATIONS*

1. A Problem.

To avoid circumlocution, consider the following: The time when a shock wave in Plexiglas passes a known distance is measured. Table IX displays a set of data obtained from such an experiment.

Table IX

Measured Values of Distance vs Time for a Shock Wave in Plexiglas

t(usec)	x(mm)
0.720	5.105
1.468	10.185
2.235	15.263
3.031	20.350
3.846	25.441

*Although they clearly are applicable to the situations, the methods discussed in this section make use of an example and data kindly supplied by B. Hayes, Group GMX-8, Los Alamos Scientific Laboratory.

It is desired to determine the instantaneous velocity for any distance traversed by the shock wave.

For any one-dimensional wave propagating in the positive x direction, one can write

$$\frac{\partial G}{\partial t} + f(t)G = 0 \tag{17}$$

as a solution to the wave equation, where G is the amplitude function and f(t) is the decrement. Specializing to the shock velocity, Eq. (17) can be written

$$\frac{du}{dt} + f(t)u = 0 \tag{18}$$

with $u = dx/dt$ the velocity of the wave front. A solution for expression (18) is

$$u = \eta e^{-\int_{t_0}^{t} f(w)dw}, \tag{19}$$

and the problem is to determine the function

$$x = x(t) = \eta \int_{t_0}^{t} e^{-\int_{t_0}^{z} f(w)dw} dz \tag{20}$$

from a given tabular set of x's and corresponding t's. The assumption that the function is well-behaved in being continuous and in having continuous derivatives is required.

2. An "Easy" Solution.

One approach is to assume that f(t) is a constant and that $t_0 = 0$, so that $x(0) = t(0) = 0$. Then the velocity expression can be written

$$u = u_0 e^{-\alpha t} \tag{21}$$

where $u_0 = \eta = dx/dt\big|_{t=0}$. This results in the equation

$$x = \frac{u_0}{\alpha}(1 - e^{-\alpha t}) . \tag{22}$$

For a given set of data, then, it is a simple matter to apply nonlinear least-squares methods and obtain estimates of u_0 and α. Applied to the data in Table IX and using initial estimates for u_0 and α of 7.0 and 0.07, respectively, this method gives final estimates of $\hat{u}_0 = 7.15$ and $\hat{\alpha} = 0.0415$ from which the minimized sum of squared deviations (observed minus calculated distances) is found to be 1.688×10^{-3}.

3. A More General Solution.

There are occasions in which the function f(t) must take on forms other than the constant. For example, if $f(t) = \alpha + \beta t$, the nonlinear contribution to the shock can be assessed. That is, if $\beta = 0$, normal attenuation occurs as the shock wave passes through the Plexiglas. If $\beta \neq 0$, the value of β is an indication of the amount of distortion as the shock wave propagates. These considerations lead to the modification of Eq. (20) into

$$x = x(t) = \int_0^t e^{-\alpha z - \beta z^2} dz, \tag{23}$$

an expression that is not integrable in closed form. However, the conditions imposed earlier allow differentiation under the integral sign (See Reference 16, pp. 167-169). It is therefore a simple matter to treat the problem as a least-squares problem, using some numerical integration scheme to obtain the quantities required to perform the fit.

The foregoing discussion was based on the assumption that time is the independent variable and that distance is the dependent variable. There are reasons for making the reverse assumptions: (1) velocity at some specified distance, rather than time, is usually required; and (2) expansion of the exponential function for hand computations then results in all terms being positive, rather than alternately positive and negative, thereby aiding certain calculations. Thus, Eq. (18) can be rearranged to read

$$\frac{d(\ln u)}{dx} + g(x) = 0 \tag{24}$$

with the result that

$$t = \frac{1}{u_0} \int_0^x e^{\int_0^z g(w)dw} dz,$$

where $u = dx/dt$ and $u_0 = dx/dt\big|_{x=t=0}$.

To illustrate: The assumption that $g(x) = \alpha$, leads to the final form

$$t = \frac{1}{\alpha u_0} \left(e^{\alpha x} - 1\right) \tag{25}$$

to be fitted.

Similarly, the assumption that $g(x) = \alpha + \beta x$, leads to

$$t = \frac{1}{u_0} \int_0^x e^{\alpha z + \beta z^2} dz . \tag{26}$$

Using initial estimates of u_0, α, and β of 7.15, 0.0415, and 0.0, respectively, final least-squares estimates were $\hat{u}_0 = 7.22$, $\hat{\alpha} = 0.00816$, and $\hat{\beta} = 0.0000782$. Although $\hat{\beta}$ is small, its estimated standard deviation, 0.0000305, is such that acceptance of the hypothesis that $\beta = 0$ is marginal at the 10% level of significance. Certainly, further experimentation is indicated by this result for one to be completely comfortable with acceptance of that hypothesis.

4. Remarks.

Experimental data are often obtained for which the only known relation among the variables is some differential equation. When the equation can be reduced to the form

$$F(dy/dx, x; \alpha_1, \alpha_2, \ldots, \alpha_k) = 0 , \tag{27}$$

the function to be fitted can be written, in its most general form, as

$$y = \int_{u_1(x;\ \alpha)}^{u_2(x;\ \alpha)} f(t, x; \alpha)\, dt + \text{const.} \tag{28}$$

In expressions (27) and (28), α denotes the vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ whose elements are the parameters to be estimated. Even though exact integration may be impossible, the problem yields to straightforward nonlinear least-squares methods which employ numerical integration techniques to evaluate the function (28) and its derivatives.

REFERENCES

1. Moore, R. H., and Zeigler, R. K., "The Solution of the General Least Squares Problem with Special Reference to High-Speed Computers," Los Alamos Scientific Laboratory Report LA-2367, 1959.
2. Finney, D. J., Probit Analysis, 2nd Ed., University Press, Cambridge, 1952.
3. Dixon, W. J., and Mood, A. M., "A Method for Obtaining and Analyzing Sensitivity Data," Am. Stat. Assoc. J. 43, 109 (1948).
4. Ashford, J. R., and Smith, C. S., "General Models for Quantal Response to the Joint Action of a Mixture of Drugs," Biometrika 51, 413 (1964).
5. Brownlee, K. A., Hodges, J. L., Jr., and Rosenblatt, M., "The Up-and-Down Method with Small Samples," Am. Stat. Assoc. J. 48, 262 (1953).
6. Berkson, J., "Tables for Use in Estimating the Normal Distribution Function by Normit Analysis," Biometrika 44, 411 (1957).
7. Berkson, J., "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function," Am. Stat. Assoc. J. 48, 565 (1953).
8. Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares," Technometrics 3, 269 (1961).
9. Marquardt, D. W., "An Algorithm for Least Squares Estimation of Non-Linear Parameters," SIAM J. 11, 431 (1963).
10. Golub, A., and Grubbs, F. E., "Analysis of Sensitivity Experiments When the Levels of Stimulus Cannot be Controlled," Am. Stat. Assoc. J. 51, 257 (1956).
11. Hewlett, P. S., and Plackett, R. L., "Statistical Aspects of the Independent Joint Action of Poisons, Particularly Insecticides. II. Examination of the Data for Agreement with the Hypothesis," Ann. Appl. Biol. 37, 527 (1950).
12. Langham, W. H., Los Alamos Scientific Laboratory, Private communication, 1966.
13. Berkson, J., "Estimation by Least Squares and by Maximum Likelihood," In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, J. Neyman, Ed., U. California Press, Berkeley, 1956, pp. 1-11.
14. Cramér, H., Mathematical Methods of Statistics, Princeton U. Press, Princeton, 1951.
15. Dixon, W. J., and Massey, F. J., Jr., Introduction to Statistical Analysis, 2nd Ed., McGraw-Hill, New York, 1957.
16. Sokolnikoff, I. S., and E. S., Higher Mathematics for Engineers and Physicists, 2nd Ed., McGraw-Hill, New York, 1941.