NBS REPORT

5068

FITTING AN ARBITRARY FUNCTIONAL RELATIONSHIP BY LEAST SQUARES WITH ALL OF THE VARIABLES SUBJECT TO ERROR

by

Kenneth A. Norton

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
BOULDER LABORATORIES
Boulder, Colorado

U. S. DEPARTMENT OF COMMERCE
Sinclair Weeks, Secretary
NATIONAL BUREAU OF STANDARDS
A. V. Astin, Director

THE NATIONAL BUREAU OF STANDARDS

The scope of the scientific program of the National Bureau of Standards at laboratory centers in Washington, D. C., and Boulder, Colorado, is given in the following outline:

Washington, D. C.

Electricity and Electronics: Resistance and Reactance. Electron Tubes. Electrical Instruments. Magnetic Measurements. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Engineering Metrology.

Heat and Power. Temperature Measurements. Thermodynamics. Cryogenic Physics. Engines and Lubrication. Engine Fuels.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Radiation Instruments Branch.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Fluid Mechanics. Engineering Mechanics. Mass and Scale. Capacity, Density, and Fluid Meters. Combustion Controls.

Organic and Fibrous Materials. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Organic Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure.

Building Technology. Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. Components and Techniques. Digital Circuitry and Devices. Digital Systems. Analog Systems. Application Engineering.

Boulder, Colorado

BOULDER LABORATORIES

F. W. Brown, Director

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Radio Propagation Physics. Upper Atmosphere. Ionospheric Research. Regular Propagation Services. Sun-Earth Relationships.

Radio Propagation Engineering. Frequency Utilization. Tropospheric Propagation.

Radio Standards. High Frequency Electrical Standards. Radio Broadcast Service. High Frequency Impedance Standards. Calibration Center. Microwave Physics. Microwave Circuit Standards.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
8300-00-9083	April 30, 1957	5068

FITTING AN ARBITRARY FUNCTIONAL RELATIONSHIP BY LEAST SQUARES WITH ALL OF THE VARIABLES SUBJECT TO ERROR

by

Kenneth A. Norton

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
BOULDER LABORATORIES
Boulder, Colorado

TABLE OF CONTENTS *

* This is a partial draft. Sections 9-17 are not yet available.

TABLE OF CONTENTS

FITTING AN ARBITRARY FUNCTIONAL RELATIONSHIP BY LEAST SQUARES WITH ALL OF THE VARIABLES SUBJECT TO ERROR

by

Kenneth A. Norton
National Bureau of Standards
Boulder, Colorado

1. INTRODUCTION

Although too often forgotten by those attempting to apply the method of least squares, it has been emphasized by most authors [1, 2, 3, 4, 5, 6, 7] throughout the development of the theory

---

1/ C. F. Gauss, "Theoria Motus Corporum Coelestium," (Hamburg, 1809), Art. 179.

2/ C. H. Kummell, "Reduction of Observation Equations which Contain More than One Observed Quantity," The Analyst (Des Moines), vol. 6, no. 4, July, 1879, pp. 97-105.

3/ H. S. Uhler, "Method of Least Squares and Curve Fitting," Jour. Optical Society of America, vol. 7, pt. 2, Nov., 1923, pp. 1043-1066.

4/ D. V. Lindley, "Regression Lines and the Linear Functional Relationship," Supplement to the Journal of Royal Stat. Soc., vol. 9, no. 2, 1947, pp. 218-244.

5/ Abraham Wald, "The Fitting of Straight Lines if Both Variables are Subject to Error," Annals. of Math. Stat., vol. XI, Sept., 1940, pp. 284-300.

6/ M. S. Bartlett, "Fitting a Straight Line when Both Variables are Subject to Error," Biometrics, vol. V, no. 3, 1949, p. 207-212.

7/ W. E. Deming, "Statistical Adjustment of Data," John Wiley and Sons, Inc., New York, 1943.

that it is necessary to assign, or in some way determine, appropriate relative weights for each of the coordinates determining the location of each of the points, before it is possible to determine consistent estimates of the parameters determining the line, curve, or surface which is to be fitted to these points. The precision of consistent estimates increases with the number of observations, and would give the population mean values if the number of observations were infinite. An additional requirement for obtaining consistent estimates of the parameters by the method of least squares is that the expected values of the errors of the observations be equal to zero, i.e., the average of an infinite number of these observational errors must be taken to be equal to zero. A solution is presented in this paper, with all of the variables subject to error, for this general least squares problem of determining consistent estimates of the parameters of a functional relationship which is expected to fit exactly the average of an infinite number of observations. A general method for weighting the data is presented, and examples are given to illustrate the effects of varying the statistical characteristics of the observed data. These examples demonstrate the necessity, if consistent estimates of the parameters are to be obtained, of modifying the original design of the experiment in many cases in order to obtain data in the particular form required by the statistical model described in this paper. It cannot be emphasized too strongly that the use of statistical models not representative of the experimental data cannot yield consistent estimates of the parameters of the functional relationship actually describing these data. Thus, only to the extent that the experimenter can demonstrate the validity of the statistical model used, can he expect to obtain by its use consistent estimates of the parameters.

The sophisticated scientist, recognizing at the outset that all of his conclusions based on measurements are only relatively correct, turns to mathematical statistics as a means of assigning quantitative probabilities to the degree of his belief. Throughout this paper an attempt is made to determine confidence bands corresponding to specified probabilities for the various statistics determined by least squares. In most cases only approximate confidence bands are now known for many of the important least squares statistics, and the development of more precise solutions would involve a very considerable complication of the analysis. Fortunately, great accuracy in the absolute magnitudes of the confidence bands is usually of secondary importance since probabilities are usually assigned to the proposed significance levels in a somewhat subjective and arbitrary manner. It is important, however, that the variation in the magnitudes of these

confidence bands with relevant statistical parameters, such as the numbers of observations and the numbers of parameters fitted, be relatively correct, and every effort has been made to ensure this result. It is important to note in this connection that the determination of only approximate confidence bands for the statistics of a model actually representative of the experimental data is much more desirable than the determination of exact confidence bands for statistics determined from a simpler statistical model which is not representative of the data.

In this paper each of the n points to be fitted in a k dimensional space is considered to be represented by the means, variances and covariances of a series of observations on each of its k variables. Thus the n points in our two dimensional examples (k = 2) are considered to be samples from n separate and independent bivariate distributions. Estimated values of the parameters (two means, two variances and one covariance) defining each of these distributions, and thus each of the points, are considered to be known on the basis of these measurements. Aside from being more general than the usual textbook expositions, in which the expected values of these variances and covariances are assumed to be the same for all n points, it is believed that the above point of view is often more realistic since this is the way experimental data frequently present themselves to the analyst. For example, the individual coordinates of each of the points to be fitted are often the means of samples from populations with different variances when obtained by the same experimenter in the same laboratory, and will even more likely be so when these points have been obtained as the result of observations by different experimenters in different laboratories.

Many analysts have been disturbed,[4, 8] when using linear regression theory to fit a straight line to a set of n points, X_i, Y_i, to find that a different line is determined when all of the variance is assigned to the values of Y_i than when all of the variance is assigned to the values of X_i. In those cases where no independent information is available as to the relative weights to assign to the Y_i and X_i observations, Wald[5] has developed a method for determining not only consistent estimates of the parameters defining the line, but also

---

[8] Joseph Berkson, "Are There Two Regressions," Jour. Amer. Stat. Assoc., vol. 45, no. 250, June, 1950, pp. 164-180.

confidence intervals for these parameters and estimates of the variances of the Y_i and X_i observations, but only on the hypotheses (1) that the expected values of the variances of X_i and Y_i are the same for all n points, (2) that the unknown errors are sufficiently small so that the classification of the X_i (or Y_i) into two groups in accordance with their magnitudes will be the same as the corresponding classification of the unknown population mean values X_{io} (or Y_{io}), (3) that X_i and Y_i are normally distributed, and (4) that their covariances are zero. Only the first two of the above conditions are required for obtaining consistent estimates of the parameters, while the other two conditions are required in order to permit the assignment of probabilities to the confidence intervals. Bartlett [6] has shown how to improve the efficiency of Wald's method, and Bennett and Franklin [9] have described a method of analysis for fitting a straight line involving a statistical model which is in many respects the same as the model considered in this paper. The present paper deals initially with the case in which independent estimates are available to the analyst of the variances and covariances of each of the X_i and Y_i observations. We will see that, when all of this information is independently available, either by measurement or assumption, tests can be made either (1) of the validity of the assumed functional relationship or (2) for the presence of random "systematic" errors. The designation "systematic" error is used consistently throughout this paper to refer to that component of error of a particular observation point which is not reduced by making repeated observations of its coordinates. If the tests indicate that the assumed functional relationship is compatible with the observed data, consistent estimates of its parameters can be obtained. These estimated parameters are, of course, derived from samples of observed populations of the two random variables X_i and Y_i, and are consistent estimates of the "true" values of the parameters only to the extent that none of these observed populations have any constant bias in their means relative to the "true" values of these random variables. It is shown that a constant bias of this kind cannot be detected by least squares.

If the tests indicate the presence of random "systematic" errors, methods are given for including their effects in the fitting. It is shown, however, that the proper inclusion of these effects of such random "systematic" errors usually cannot be determined from a

---

[9] Carl A. Bennett and Normal L. Franklin, "Statistical Analysis," John Wiley and Sons, 1954, p. 463. These methods appear to be largely drawn from J. W. Tukey, "Components in Regression," Biometrics, 7, 1951, pp. 33-69.

statistical analysis of the experimental data alone and must be based on ad hoc assumptions supplied by the experimenter. In particular, when fitting a straight line, the variance arising from these random "systematic" errors cannot, by statistical analysis alone, be subdivided into its three components: (1) a possible systematic error variance of X_i, (2) a possible systematic error variance of Y_i, and (3) the covariance of these two possible sources of random systematic error.

Berkson[8] has proposed a most desirable experimental procedure for taking data intended to be fit to a linear functional relationship with both variables subject to error. He calls this procedure a controlled experiment and shows that great simplification in the statistical analysis results from its use. In fact it can be shown that the analysis of measurements made by this procedure with both variables subject to error can be reduced to a problem involving only one variable subject to error. This procedure and method of analysis are described, and a brief indication given of how its great advantages may be extended to the general problem of fitting an arbitrary functional relationship with all of the variables subject to error.

It is shown for all of the statistical models discussed in this paper how to "adjust" the experimental observations to those particular values which jointly have a minimum weighted mean square deviation relative to all of the experimental data and of the assumed functional relationship. In the absence of systematic errors, these "adjusted" values will also be consistent estimates of the population mean values of the n experimental points. Thus, by directing the attention of the experimenter to these estimated errors (random plus "systematic") thus determined for each coordinate of each of his observed points, it often makes possible a better understanding of the nature and source of these errors.

Methods are given for calculating (1) confidence bands for the least squares estimates of each of the parameters of the functional relationship considered independently, (2) elliptical confidence regions for two or more of these parameter estimates considered jointly, (3) confidence regions for the fitted functional relationship, and (4) confidence regions for future values predicted by the use of the fitted functional relationship.

The primary purpose of this paper is tutorial, although some results are presented which are believed to be new, and errors in previous solutions are corrected. By a slight modification in approach,

some of the complexity of the usual expositions has been avoided, while at the same time other aspects of the problem have been generalized.

Many of our results are contained in the general solution of this problem obtained by Kummell in 1879 [2] but, possibly because he presented no illustrative examples, the full significance and importance of Kummell's results appear to have been overlooked by subsequent authors.

An effort is made in the following presentation to point out the wide variety of solutions to be expected on the basis (a) of different experimental information or of different initial assumptions relative to the weights to be assigned to the individual observed points or (b) of different assumptions relative to the form of the functional relationship fitted to the data. Thus it cannot be emphasized too strongly that the method of least squares should not be used blindly by the experimenter in the hope that somehow this will provide in some magic way a best solution. Instead, careful consideration should be given first to the form of the function to be fitted to the data, and then to the nature and reliability of each experimental point. A solution by least squares does make possible the efficient use of all of the information available to the experimenter, but it can never give results which are any better than the assumptions and experimental data used.

Some experimenters use least squares only in those cases where it is quite clear from a plot of their data that the assumption under consideration (linearity of the assumed relation, for example) is well established and would not use the method when the points are widely scattered. It seems to the author that this latter attitude reveals a lack of appreciation of the scope of the method, since it is frequently in just those cases where the data are widely scattered that the method of least squares is most useful by providing a significant quantitative evaluation of the reality of the assumed relations. It is important to remember that the results of many, even properly conducted, physical experiments yield widely scattered data because of the impossibility of controlling the influence of many of the variables, and this results in the introduction of large errors. It is quite clear that these uncontrollable experiments should receive as much, if not more, careful attention from the analyst as those experiments for which carefully controlled conditions are feasible. In this paper

a more precise estimate of the standard errors is given which allows for the second order effects which arise with widely scattered data.

Even granting the possibility of obtaining the above information by the application of the method of least squares to the analysis of a set of experimental data, the question as to when the complexities of analysis by this method are justified is a difficult one, and presumably must finally be conceded to be a matter of judgment with the individual analyst. In any case, before deciding this question, the experimenter should know what kinds of additional information are likely to become available from the application of the method, and it is the purpose of this paper to supply that information in detail for the particular case of fitting a linear functional relationship to a set of experimentally determined points. Although most of the detailed examples presented are confined to an assumed linear relationship, the theory presented is developed for the general case of an arbitrary functional relationship, and the normal equations given provide a solution for this general case with all of the variables subject to error.

In the case of fitting data to a straight line, it is shown that the entire effect of introducing errors in the independent as well as in the dependent variable enter into the solution for the parameters by way of second order terms in the residuals and terms involving the differentiation of the weights. The resulting effects on the estimated parameters is often quite large. The necessity for retaining these additional terms makes our solution inherently more complex than the usual solutions in which only the dependent variable is assumed to be subject to error. Unfortunately the determination of the sampling distribution for the estimated parameters is also inherently more complicated in the general case, and precise analytical expressions for these sampling distributions are not yet available. Nevertheless it seems quite clear that it is more desirable, when both variables are subject to error, to obtain consistent estimates of the parameters and only rough estimates of their errors by the methods described in this paper than to follow the current practice of arbitrarily suppressing the influence of the errors of the independent variables, thus obtaining incorrect estimates of the parameters and confidence regions for these incorrect estimates which are precise only on the false assumption that the errors of the independent variable may be suppressed.

The ultimate aim of most analyses by least squares is the prediction of one variable from the observed values of another, and it is shown in this paper that the estimated functional relationship should be used for this purpose. An example of the problem of prediction is the following. Suppose n pairs of mean values X_i, Y_i (i = 1 to n) are determined by averaging m_i observations on n different populations which have n different unknown population mean values X_{io}, Y_{io}. It is supposed that X_i and Y_i are normally distributed about X_{io} and Y_{io} with unknown variances $\sigma^2_{\epsilon i}/m_i$ and $\sigma^2_{\eta i}/m_i$, respectively, and that $Y_{io} = \alpha + \beta X_{io}$ where α and β are unknown constants. A new mean value X_j is determined by averaging m_j observations of X from the j^{th} population with unknown population mean values X_{jo}, Y_{jo} which are also assumed to be related by $Y_{jo} = \alpha + \beta X_{jo}$. It is desired to estimate the population mean value Y_{jo} by means of the observed mean value X_j, and it is shown that a consistent estimate of Y_{jo} is $a + b X_j$ where a and b are the least squares estimates of α and β, approaching these values in the limit as all of the m_i (i = 1 to n) approach infinity.

The above is believed to be a more realistic statement of the problem of prediction than the more usual statement [4, 5, 6] that we wish to know the expected value of Y for a given observed value X_j without specifying that X_j is from the j^{th} population with a fixed, even if unknown, population mean value X_{jo}. For example, if we have estimated the density, β, of a certain kind of steel by measuring the n weights, Y_i, and the n volumes, X_i, of n steel balls made with this steel, we may estimate the weight, Y_j, of a j^{th} steel ball by bX_j where b is the least squares estimate of β obtained on the assumptions (1) that Y_i and X_i are each measured with error and (2) that the true density β is the same for all n + 1 balls, i.e., the assumption that a linear functional relationship exists. Note that the j^{th} steel ball has a fixed true weight Y_{jo} and a fixed true volume X_{jo}. Now the expected value of the observed volume X_j for the j^{th} steel ball is obviously its true volume X_{jo} and it follows that a consistent estimate of the weight Y_{jo} of the j^{th} steel ball (given the j^{th} steel ball and its measured volume X_j) is bX_j. Now consider Lindley's [4] formulation and solution of the problem of prediction. Lindley assumes, in effect, that an infinite population of steel balls exists and that the true volumes X_{io} are normally

distributed* about a mean value X_o with variance σ_o^2; he further assumes that the true density β is the same for all of the steel balls and that the variances of the errors of measurement have the same magnitudes σ_Y^2 for weight and σ_X^2 for volume independent of the size of the balls being measured. He then purports to show that an unbiased estimate of the weight of the j^{th} ball is $b' X_j$ where b' is the regression coefficient of the n random observed values of Y_i on X_i. However, he also shows that b' approaches the constant value $\gamma = (X_o^2 + \sigma_o^2)\beta / (X_o^2 + \sigma_o^2 + \sigma_X^2)$ as n approaches infinity or, more generally, that the average value of b' equals γ where the individual values of b' are determined from many random finite samples with n balls in each sample. Let us assume that γ is determined exactly by determining the regression of Y_i on X_i by measuring an infinite number of steel balls chosen at random from the population. Now select one more steel ball and measure its volume X_j; Lindley states that an unbiased estimate of the weight of this j^{th} steel ball is $Y'_j = \gamma X_j$. But we may measure the volume of this j^{th} steel ball an infinite number of times, and in this way we might hope to determine the true weight Y_{jo}; however, the mean value $\overline{Y'_j} = \gamma \overline{X_j} = \gamma X_{jo} < Y_{jo} = \beta X_{jo}$ and we conclude that Lindley's prediction leads to a biased estimate of the true value of the weight of the j^{th} steel ball. If we had instead used the infinite sample of Y_i, X_i to estimate b by least squares with both variables subject to error, the estimate so obtained would be equal to β, and now the mean value $\overline{Y_j} = \beta \overline{X_j} = \beta X_{jo} = Y_{jo}$, i.e., the use of the average of a large number of measurements of the volume of the j^{th} steel ball would lead in this way to the true weight of this ball. Lindley's expected values are obtained by averaging over the entire population of steel balls; thus he has shown that a second independent set of measurements of n balls chosen at random from the same population would lead, as n approaches infinity, to the same biased estimate, γ, of the density, provided this second set of measurements of volume were made with instruments having the same precision so that σ_X^2 is the same. Only in this most unsatisfactory sense can Lindley claim

---

* Since a volume cannot be negative, this assumption cannot be strictly true, but the distribution of X_{io} can approximate a normal distribution if $\sigma_o << X_o$.

that predictions based on simple regression theory are unbiased; such predictions are not even consistent since they disagree with the average of a large number of measurements made on the j^{th} ball.

Predictions made using the fitted functional relationship will, in most cases, be biased; but they are always at least consistent, and they have the advantage that no special assumptions need be made as to the distribution of the true values X_{io}. Furthermore, the bias of predictions made using the fitted functional relationship will always* be less than the bias of predictions made using simple regression analysis with the errors in the independent variable neglected.

The above discussion of prediction relates to the case in which a functional relationship is assumed to exist. However, it also follows from the above argument that predictions of a dependent variable Y made by the use of regression analysis of data involving one or more independent variables which are measured with error will also be biased. For example, if σ^2_{Xo} denotes the variance of the true values of the independent variable and σ^2_X the variance of the errors of measurement of the independent variable, then an unbiased prediction of Y is given by $(Y - \overline{Y}) = b'(X - \overline{X})(\sigma^2_{Xo} + \sigma^2_X)/\sigma^2_{Xo}$ where b' is the biased slope of the regression of Y on X with the errors in X ignored. In practice the bias correction $(\sigma^2_{Xo} + \sigma^2_X)/\sigma^2_{Xo}$ might be estimated by $S^2_{Xi}/(S^2_{Xi} - s^2_X)$ where S^2_{Xi} is the sample variance of the X_i about their mean value $\overline{X}$ and s^2_X is an estimate of the variance σ^2_X of the errors of measurement of the X_i.

---

* The mathematical proof is not yet available, but this conclusion seems reasonable since the functional relationship analysis is designed to provide a first order correction for this particular bias. In a certain sense predictions made using the fitted functional relationship can actually be shown to be unbiased—see page 2.18.

2. FITTING A STRAIGHT LINE

Let us begin with a precise formulation of the problem. We consider 2n sets of random variables, X_{it}, Y_{it}, t = 1 to m_i, i = 1 to n with $m_i > 1$ and $n > 1$. A random variable is a real variable with an associated probability distribution. The expected values of these random variables are $E(X_{it}) = X_{io}$ and $E(Y_{it}) = Y_{io}$. The expected value is the average of the values in a large sample as the sample size goes to infinity; or, more precisely, the expected value of a random variable is its first moment, i.e., its average value weighted in accordance with its probability distribution. X_{io} and Y_{io} are also called the population mean values of these random variables while the random variables $(X_{it} - X_{io}) = \epsilon_{it}$ and $(Y_{it} - Y_{io}) = \eta_{it}$ are called errors, although a large component of such deviations from the population mean values may arise in some applications from the natural variations of the phenomena under investigation. Regardless of the cause of the deviations, we take the expected values $E(\epsilon_{it}) = 0$ and $E(\eta_{it}) = 0$. The random variables ϵ_{it} and η_{it} within the i^{th} group are assumed to have the same bivariate probability distribution, to be independent and thus uncorrelated, i.e., $E(\epsilon_{it}^2) = \sigma_{\epsilon i}^2$, $E(\eta_{it}^2) = \sigma_{\eta i}^2$, and $E(\epsilon_{it}\,\eta_{it}) = \rho_i\,\sigma_{\epsilon i}\,\sigma_{\eta i}$ while $E(\epsilon_{it}\,\epsilon_{iu}) = 0$, $E(\eta_{it}\,\eta_{iu}) = 0$, and $E(\eta_{it}\,\epsilon_{iu}) = 0$ for $t \neq u$. The variances of ϵ_{it} and of η_{it} are all finite, but may differ from each other and from one group to the next. The observations in the different groups (i = 1 to n) are independent and thus uncorrelated $E(\epsilon_{it}\,\epsilon_{ju}) = 0$ and $E(\eta_{it}\,\eta_{ju}) = 0$ for $i \neq j$. It is anticipated that the general approach here developed may be extended to the case of fitting autocorrelated data, but this is beyond the scope of the present paper. Finally it is assumed that the population mean values X_{io} and Y_{io} are exactly related by the linear relation $Y_{io} = \alpha + \beta X_{io}$, i.e., α and β denote the "true" values of the parameters of this linear relation. By virtue of these assumptions, the model described in this section excludes "systematic" errors by definition. More general models with explicit allowance for random systematic errors are presented in later sections.

Let the i^{th} observation point be defined by the mean values $X_i = \frac{1}{m_i}\sum X_{it}$ and $Y_i = \frac{1}{m_i}\sum Y_{it}$. Our problem is to fit the straight line

$$Y = a + bX \tag{2.1}$$

to the n points, X_i, Y_i. In (2.1) a and b denote the estimated values of the parameters α and β. Unbiased estimates of the variances of X_{it}, Y_{it}, X_i, and Y_i are:

$$s^2_{\epsilon i} = \frac{1}{m_i - 1} \sum_{t=1}^{m_i} (X_{it} - X_i)^2 \qquad s^2_{\eta i} = \frac{1}{m_i - 1} \sum_{t=1}^{m_i} (Y_{it} - Y_i)^2 \tag{2.2}$$

$$s^2_{Xi} = \frac{1}{m_i} s^2_{\epsilon i} \qquad s^2_{Y_i} = \frac{1}{m_i} s^2_{\eta i} \tag{2.3}$$

The estimate from the sample in the i^{th} group of the correlation coefficient, r_i, between X_{it} and Y_{it}, which also characterizes the correlation between the errors in the coordinates X_i and Y_i of the i^{th} observation point, is given by:

$$r_i s_{\epsilon i} s_{\eta i} = s_{\epsilon \eta i} \equiv \left\{ \sum_{t=1}^{m_i} (X_{it} - X_i)(Y_{it} - Y_i) \right\} / (m_i - 1) \tag{2.4}$$

The numerical values of the three coordinates and their variances for the three-point examples discussed in detail in this paper are given in Table 2.1.

TABLE 2.1*

i	m_i	X_i	$s^2_{\epsilon i}$	$s^2_{X_i}$	Y_i	$s^2_{\eta i}$	$s^2_{Y_i}$
1	5	2	5	1	2	10	2
2	5	6	20	4	4	15	3
3	5	8	25	5	8	30	6

* The sample values in this table are unrealistic since, in practice, the probability of finding such round numbers would be near zero; these round numbers were chosen for convenience only. See Section 8 for an example with a more typical set of numbers.

Using three different values of $r_i = 0$, -0.9, and $+0.9$, together with the values given in Table 2.1, we obtain by the methods to be discussed in detail in the sequel, the three solutions shown by the three different lines on Figs. 1, 2 and 3. The lines associated with each of the three points on Fig. 1 extend a distance equal to one standard error and are intended to provide a visual indication of the uncertainties in each of the three observed values. The meaning of the ellipses shown on these three figures will be explained later.

The necessity for estimating the standard errors, s_{Xi} and s_{Yi} associated with each point before applying the method of least squares will undoubtedly appeal to most physicists as a most natural course of action and may lead, in many cases, to greater care in the planning of experiments. The variances, s^2_{Xi} and s^2_{Yi} may be considered to arise from experimental errors in observation of the coordinates of each point, or may be considered to arise from natural variations of these quantities. For example, the population mean of each observation point, X_{io}, Y_{io}, may be thought of as corresponding to a given fixed setting (as of a rheostat, for example) of the experimental system, and X_i, Y_i might then be the averages of a series of m_i observations X_{it}, Y_{it}, taken at that setting. Note that the m_i values observed within the i^{th} group are assumed to be samples from the same statistical population; in the language of the physicist these m_i values are considered to have been obtained under the same experimental conditions. On the other hand, the observed values in the n different groups may each be from statistical populations with different variances as well as different population mean values, and this will often be the case in practice.

Let $V_{Yi} \equiv (Y_i - a - b X_i)$ denote the deviation in the Y direction of the i^{th} point from the fitted line, and let $w(V_{Yi})$ denote the weight assigned to the deviation V_{Yi}. Our present problem is to determine the values of a and b in (2.1) which will minimize, S(a, b), the weighted sum of the squares of the deviations of the points from the line:

$$S(a, b) = [w(V_{Yi}) V^2_{Yi}] \qquad (2.5)^*$$

---

* This dual role of the symbols a and b, i.e., as variables in this expression for S(a, b) and as the particular constants which minimize S, should not lead to as much confusion as the use of two sets of symbols.

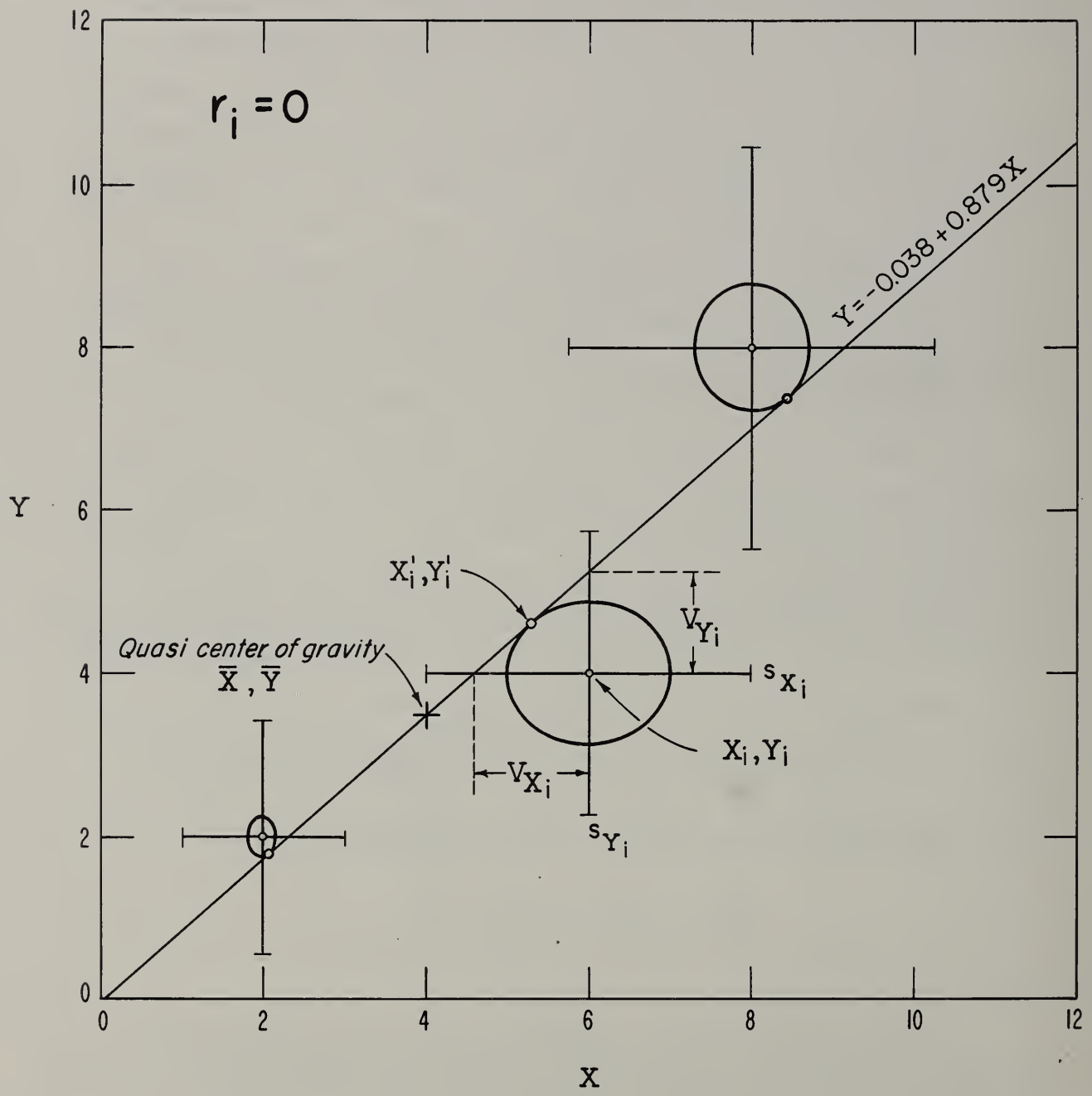
$r_i = 0$
Y=-0.038+0.879X
X_i', Y_i'
V_{Y_i}
s_{X_i}
Quasi center of gravity
$\bar{X}, \bar{Y}$
X_i, Y_i
V_{X_i}
s_{Y_i}
Y
X
0
2
4
6
8
10
12

Figure 2.1

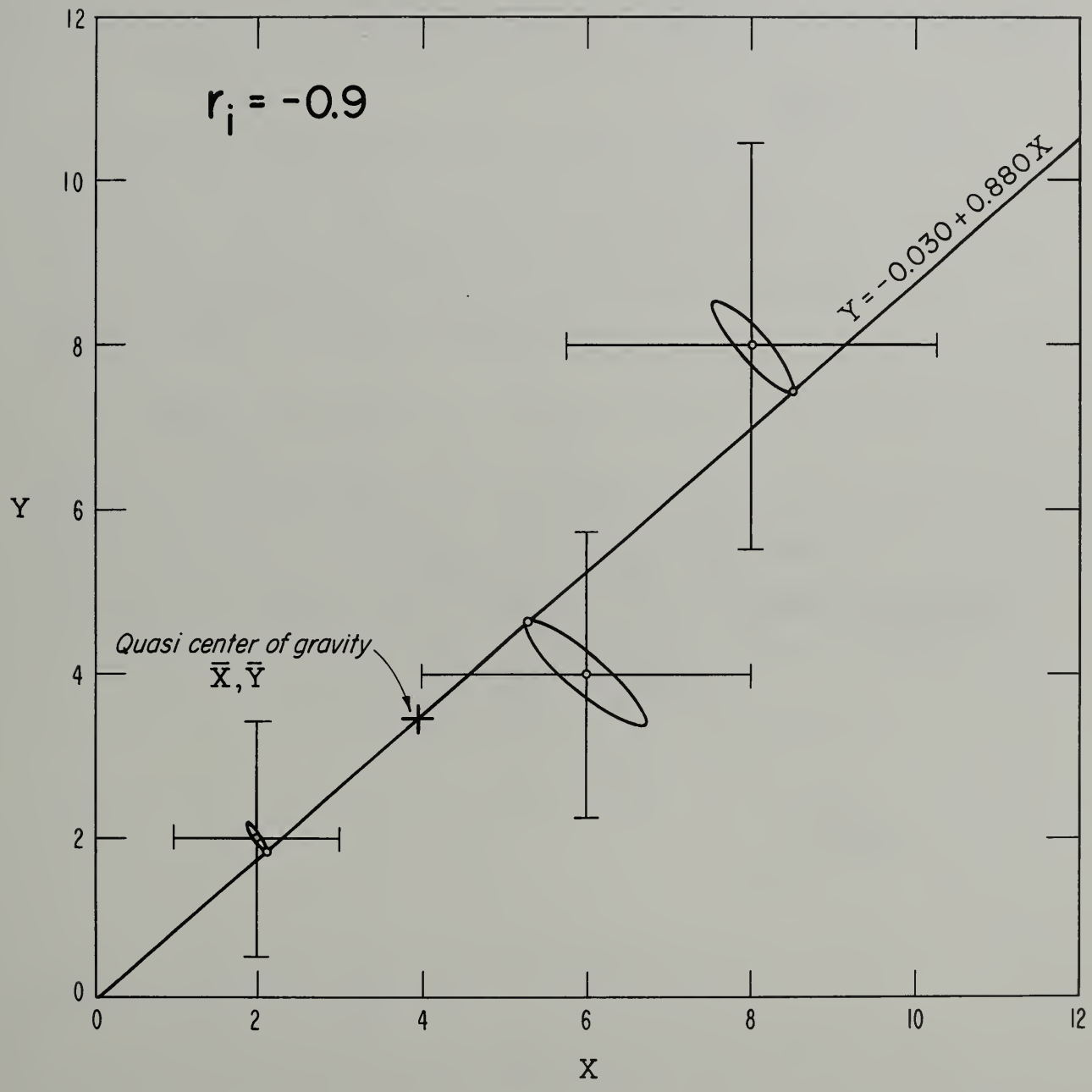
$r_i = -0.9$
Y = -0.030 + 0.880 X
Quasi center of gravity
$\bar{X}, \bar{Y}$
Y
X
0
2
4
6
8
10
12

Figure 2.2

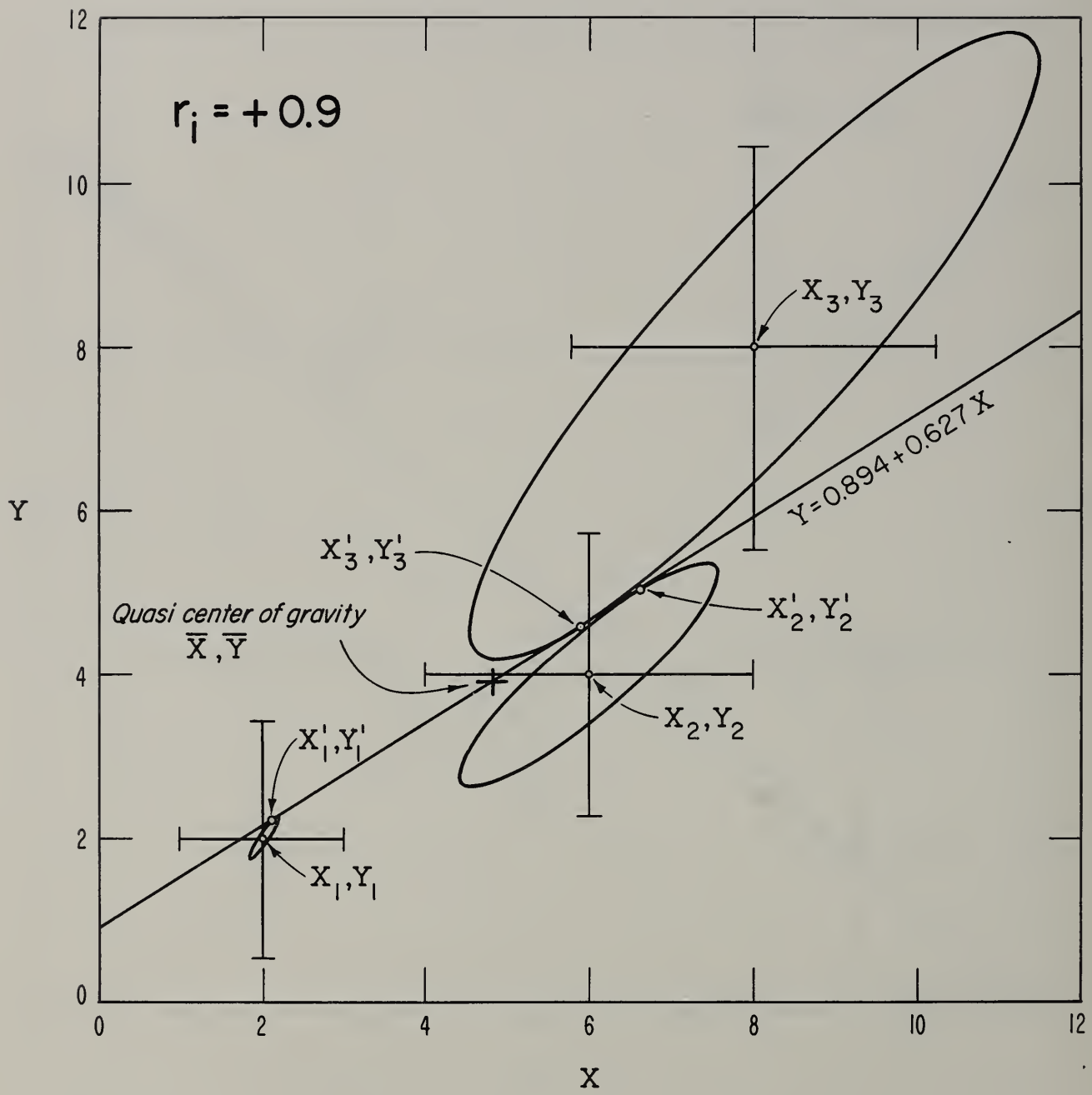
$r_i = +0.9$
X_3, Y_3
$Y = 0.894 + 0.627X$
X'_3, Y'_3
X'_2, Y'_2
Quasi center of gravity
$\overline{X}, \overline{Y}$
X_2, Y_2
X'_1, Y'_1
X_1, Y_1
Y
X
0
2
4
6
8
10
12

Figure 2.3

Following Gauss, the square brackets [ ] are used throughout this paper to denote the sum of the n values so enclosed and corresponding to the n independent points to which the functional relationship is to be fitted. It will be convenient to define the weight of a deviation to be the reciprocal of its variance as estimated by means of the following particular approximate formula for this variance:

$$\frac{1}{w(V_{Yi})} \equiv s^2_{V_{Yi}} \equiv s^2_{Yi} - 2br_i s_{Yi} s_{Xi} + b^2 s^2_{Xi} \equiv s^2_i \equiv \frac{1}{w_i} \qquad (2.6)^*$$

The above approximate expression for the variance follows directly from the above expression for V_{Yi} and the law for the propagation of variance.† It will become evident as we proceed that the definition of the weight in terms of this particular approximation to the variance will lead to consistent estimates of the parameters α and β. With a and b now representing the particular values of these variables which minimize S(a, b), (2.5) may be expressed:

$$S(a, b) = [w(V_{Yi})(Y_i - a - bX_i)^2] = [(V_{Yi}/s_i)^2] = \text{minimum} \qquad (2.7)$$

The alert reader will raise the question as to why we have chosen to measure our deviations, V_{Yi} in the Y direction, and the answer is that the direction chosen is immaterial since S is invariant to a homogeneous strain, translation, or rotation of the coordinate axes; the logical necessity for this invariance in least squares was pointed out by Roos. 10/ In particular, it is easy to show that $[w(V_{Xi})V^2_{Xi}]$ is identical to (2.7); thus $V_{Xi} \equiv (X_i + \frac{a}{b} - \frac{1}{b} Y_i) = -\frac{1}{b} V_{Yi}$ and

$$\frac{1}{w(V_{Xi})} \equiv s^2_{V_{Xi}} = s^2_{Xi} - \frac{2}{b} r_i s_{Yi} s_{Xi} + \frac{1}{b^2} s^2_{Yi} = \frac{s^2_i}{b^2} = \frac{1}{b^2 w_i} .$$

---

* Note that s^2_i is finite and positive except in the trivial cases $r_i = \pm 1$ and $b s_{Xi} = r_i s_{Yi}$.

† Terms involving the variances s^2_a and s^2_b of the random variables a and b were omitted in deriving (2.6).

10/ C. F. Roos, " A General Invariant Criterion of Fit for Lines and Planes where all Variates are Subject to Error, " Metron, Feb., 1937.

The proof of the more general invariance of S stated above is given in Section 3. Throughout the remainder of this and the following sections the symbol w_i will be understood to represent $w(V_{Y_i})$ and the symbol s_i^2 will be understood to represent $s^2_{V_{Yi}}$.

Since S is to be a minimum for the least squares determination of the values of a and of b, it follows that the partial derivatives of S with respect to a and b must both be equal to zero. We will consider first the partial derivative with respect to a:

$$\frac{1}{2}\ \frac{\partial S}{\partial a} = -\left[w_i(Y_i - a - bX_i)\right] = 0 \qquad (2.8)$$

$$a = \frac{[w_i Y_i]}{[w_i]} - b\,\frac{[w_i X_i]}{[w_i]} \equiv \overline{Y} - b\overline{X} \qquad (2.9)$$

The point $\overline{X}$, $\overline{Y}$ is what Deming [7] calls the quasi center of gravity.

The first step in the series of calculations required for our solution is the determination of the n values of w_i corresponding to the n pairs of observations. The reader will detect a logical difficulty in our development at this point, since he is asked to use a formula for w_i which involves the value of the so far unknown constant b. This is indeed a logical difficulty, but is overcome in practice simply by using an estimated value, say b_0, in (2.6) for the initial determination of w_i and later, if necessary, repeating the entire set of calculations with a better estimate of b obtained from the first set of calculations. The general conditions under which this iterative process will be convergent have not been studied, but no difficulty is anticipated in most practical applications. An estimated value of the quasi center of gravity may now be determined.

Substituting the value of a obtained from (2.9) into (2.1), we obtain:

$$(Y - \overline{Y}) = b(X - \overline{X}) \qquad (2.10)$$

It will be convenient now to choose a new set of coordinates, $x = X - \overline{X}$ and $y = Y - \overline{Y}$ with their origin at the quasi center of gravity; then (2.10) and (2.7) become:

$$y = bx \tag{2.11}$$

$$S(b) = [w_i(y_i - bx_i)^2] = \left[\frac{(y_i - bx_i)^2}{s_{Yi}^2 - 2br_i s_{Yi} s_{Xi} + b^2 s_{Xi}^2} \right] = \text{minimum} \tag{2.12}$$

By setting the derivative of S with respect to b equal to zero, we obtain an expression which may be solved for b:

$$\frac{1}{2} \frac{\partial S}{\partial b} = -[w_i x_i(y_i - bx_i)] + [w_i^2(r_i s_{Yi} s_{Xi} - bs_{Xi}^2)(y_i - bx_i)^2] = 0 \tag{2.13}$$

This may be expressed as a quadratic in b if, as before, we set b equal to an estimated value b_o in the weighting factor after multiplying each of the first set of terms by $w_i(s_{Yi}^2 - 2br_i s_{Yi} s_{Xi} + b^2 s_{Xi}^2) \equiv 1$.

$$b^2[w_i^2 s_{Xi} x_i(s_{Xi} y_i - r_i s_{Yi} x_i)] - b[w_i^2(s_{Xi}^2 y_i^2 - s_{Yi}^2 x_i^2)] - [w_i^2 s_{Yi} y_i(s_{Yi} x_i - r_i s_{Xi} y_i)] = 0 \tag{2.14}$$

$$b = \frac{[w_i^2(s_{Xi}^2 y_i^2 - s_{Yi}^2 x_i^2)] + \sqrt{[w_i^2(s_{Xi}^2 y_i^2 - s_{Yi}^2 x_i^2)]^2 + 4[w_i^2 s_{Xi} x_i(s_{Xi} y_i - r_i s_{Yi} x_i)] \cdot [w_i^2 s_{Yi} y_i(s_{Yi} x_i - r_i s_{Xi} y_i)]}}{2[w_i^2 s_{Xi} x_i(s_{Xi} y_i - r_i s_{Yi} x_i)]} \tag{2.15}$$

If the plus sign be changed to minus in front of the radical, the result is the slope of the worst fitting line among those which pass through $\overline{X}$, $\overline{Y}$, i.e., that which maximizes S.

When $[w_i^2 s_{Xi}^2 y_i^2]$ is either very much less than or very much greater than $[w_i^2 s_{Yi}^2 x_i^2]$, the second term under the radical in (2.15) will be smaller than the first term, and we may obtain the following series expansions for b by expanding the radical:

$$b = \frac{[w_i^2 s_{Yi} y_i (s_{Yi} x_i - r_i s_{Xi} y_i)]}{[w_i^2 (s_{Yi}^2 x_i^2 - s_{Xi}^2 y_i^2)]} \left\{ 1 - \frac{[w_i^2 s_{Xi} x_i (s_{Xi} y_i - r_i s_{Yi} x_i)] \cdot [w_i^2 s_{Yi} y_i (s_{Yi} x_i - r_i s_{Xi} y_i)]}{[w_i^2 (s_{Yi}^2 x_i^2 - s_{Xi} y_i^2)]^2} + \ldots \right\} \tag{2.16}$$

where $$[w_i^2 s_{Xi}^2 y_i^2] << [w_i^2 s_{Yi}^2 x_i^2] \tag{2.17}$$

$$b = \frac{[w_i^2 (s_{Xi}^2 y_i^2 - s_{Yi}^2 x_i^2)]}{[w_i^2 s_{Xi} x_i (s_{Xi} y_i - r_i s_{Yi} x_i)]} \left\{ 1 + \frac{[w_i^2 s_{Xi} x_i (s_{Xi} y_i - r_i s_{Yi} x_i)] \cdot [w_i^2 s_{Yi} y_i (s_{Yi} x_i - r_i s_{Xi} y_i)]}{[w_i^2 (s_{Yi}^2 x_i^2 - s_{Xi} y_i^2)]^2} - \ldots \right\} \tag{2.18}$$

where $$[w_i^2 s_{Xi}^2 y_i^2] >> [w_i^2 s_{Yi}^2 x_i^2] \tag{2.19}$$

In the limit as $\sigma^2_{\epsilon i} = 0$ or $\sigma^2_{\eta i} = 0$, we obtain the following two solutions from (2.16) and (2.18), respectively:

$$b = \frac{\left[\frac{y_i x_i}{2}\right]}{\left[\frac{x_i}{2}\right]} \quad \text{(where } \sigma^2_{\epsilon i} = 0\text{)} \qquad (2.20)$$

$$b = \frac{\left[\frac{y_i^2}{2}\right]}{\left[\frac{y_i x_i}{2}\right]} \quad \text{(where } \sigma^2_{\eta i} = 0\text{)} \qquad (2.21)$$

Fortunately in some cases all of the bivariate distributions associated with the n observation points may be considered to be samples from statistical populations with the same variance. In these cases it will be advantageous to pool the sample variances associated with the observations on each of the coordinates, and in this way obtain improved estimates which also have the advantage of being independent of i. These pooled estimates may be calculated from the following equations:

$$s^2_\epsilon = \frac{[(m_i - 1)\, s^2_{\epsilon i}]}{[m_i] - n}; \qquad s^2_\eta = \frac{[(m_i - 1)\, s^2_{\eta i}]}{[m_i] - n} \qquad (2.22)$$

$$s^2_{Xi} = s^2_\epsilon / m_i \quad ; \qquad s^2_{Yi} = s^2_\eta / m_i \qquad (2.23)$$

The decision to pool the sample variances in the manner indicated above should be made, of course, only after tests have been made to see whether the n individual sample variances may reasonably be considered to be from the same parent population. A good review of such tests is given in a publication of the Office of Scientific Research and Development[11] and on page 196 of reference 9.

---

[11] Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis, "Techniques of Statistical Analysis," Chapter 15, McGraw-Hill Book Co., Inc., 1947.

A pooled estimate of the correlation coefficient, r, must also be obtained. Such a pooled estimate of r may be obtained directly from:

$$r = \frac{[\Sigma_t (X_{it} - X_i)(Y_{it} - Y_i)]}{\sqrt{[\Sigma_t (X_{it} - X_i)^2][\Sigma_t (Y_{it} - Y_i)^2]}} \qquad (2.24)$$

It is often better, however, to make use of the individual sample values of r_i so that they also may be tested to find out whether they are probably from the same population. For this purpose use may be made of a characteristic of sample correlation coefficients for normally distributed data discovered by Fisher. [12/] Fisher found that the statistic:

$$z_i = \frac{1}{2}[\log_e(1 + r_i) - \log_e(1 - r_i)] \qquad (2.25)$$

is distributed almost normally with variance $(m_i - 3)^{-1}$. Snedecor [13/] discusses the use of this statistic for testing whether the individual values of r_i are from the same population, and presents a very convenient graphical method for converting r_i to z_i and vice versa. Since z_i is approximately normally distributed, a pooled estimate r may be obtained from the weighted average value:

$$\bar{z} = \frac{[(m_i - 3)\, z_i]}{[m_i - 3]} \qquad (2.26)$$

which is then converted by means of the following relation to the required pooled estimate of r:

$$r = \tanh\left\{\bar{z} - \frac{r\left[\frac{m_i - 3}{m_i - 1}\right]}{2\,[m_i - 3]}\right\} \qquad (2.27)$$

The above relation allows for a small bias in the distribution of z; this transcendental equation for r may be solved by an iterative process, using initially an estimate of r.

[12/] R. A. Fisher, "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," Metron, vol. 1, no. 4, 1921.

[13/] George W. Snedecor, "Statistical Methods," Iowa State College Press, Fifth Edition, 1956, page 175.

When it is feasible to pool the variances and covariances as indicated above, it is no longer necessary to use an estimated value, b_o, to obtain a solution, and the above equations for the slope simplify to the following:

$$b = \frac{-[m_i x_i^2] + (s_\epsilon^2 / s_\eta^2)[m_i y_i^2] + \sqrt{[m_i\{x_i^2 - (s_\epsilon^2 / s_\eta^2) y_i^2\}]^2 + 4(s_\epsilon / s_\eta)[m_i\{(s_\epsilon / s_\eta) x_i y_i - r x_i^2\}] \cdot [m_i\{x_i y_i - r(s_\epsilon / s_\eta) y_i^2\}]}}{2(s_\epsilon / s_\eta)[m_i\{(s_\epsilon / s_\eta) x_i y_i - r x_i^2\}]} \qquad (2.28)$$

$$b = \frac{[m_i y_i x_i] - r(s_\epsilon / s_\eta)[m_i y_i^2]}{[m_i x_i^2] - (s_\epsilon^2 / s_\eta^2)[m_i y_i^2]} \left\{1 - \ldots\right\} \qquad \text{where } s_\epsilon^2 [m_i y_i^2] << s_\eta^2 [m_i x_i^2] \qquad (2.29)$$

$$b = \frac{[m_i y_i^2] - (s_\eta^2 / s_\epsilon^2)[m_i x_i^2]}{[m_i y_i x_i] - r(s_\eta / s_\epsilon)[m_i x_i^2]} \left\{-1 + \ldots\right\} \qquad \text{where } s_\eta^2 [m_i x_i^2] << s_\epsilon^2 [m_i y_i^2] \qquad (2.30)$$

$$b = \frac{[m_i y_i x_i]}{[m_i x_i^2]} \qquad \text{where } \sigma_\epsilon^2 = 0 \qquad (2.31)$$

$$b = \frac{[m_i y_i^2]}{[m_i x_i y_i]} \qquad \text{where } \sigma_\eta^2 = 0 \qquad (2.32)$$

If we set m_i equal to a constant in the last two of the above equations, we obtain the pair of solutions arising from regression analysis. It is important to note that these are very special cases and will represent the proper solution for a linear functional relation between two observed variables if, and only if, either one or the other of the conditions $\sigma_\epsilon^2 = 0$ or $\sigma_\eta^2 = 0$ can be established.

In some cases, even though $\sigma^2_{\epsilon i}$ and $\sigma^2_{\eta i}$ may differ from point to point, the analyst may know a priori that the n bivariate distributions are characterized (a) by a common value ρ for their population correlation coefficients and (b) by a common value C^2 for the ratio $(\sigma^2_{\epsilon i}/\sigma^2_{\eta i})$ between their variances. In this case the following weighting factor may be used:

$$w_i = (m_i - 1)/\hat{\sigma}^2_{\eta i}(1 - 2b\hat{\rho}\hat{C} + b^2\hat{C}^2) \tag{2.33}$$

Here $\hat{\sigma}^2_{\eta i}$, $\hat{\rho}$, and $\hat{C}$ represent maximum likelihood estimates as derived in Appendix I for the case of sampling from bivariate normal populations; the factor $(m_i - 1)$ in (2.33) was used instead of the usual m_i in order to correct for the small sample bias characteristic of the maximum likelihood estimates. In the above case the expression for the slope becomes:

$$b = \frac{[w_i(\hat{C}^2 y_i^2 - x_i^2)] + \sqrt{[w_i(\hat{C}^2 y_i^2 - x_i^2)]^2 + 4[w_i(\hat{C}^2 x_i y_i - \hat{\rho}\hat{C} x_i^2)]\cdot[w_i(x_i y_i - \hat{\rho}\hat{C} y_i^2)]}}{2[w_i(\hat{C}^2 x_i y_i - \hat{\rho}\hat{C} x_i^2)]} \tag{2.34}$$

The factor $(1 - 2b\hat{\rho}\hat{C} + b^2\hat{C}^2)$ in (2.33) does not vary with i, and thus may be factored out in determining $\overline{X}$, $\overline{Y}$, and b. Thus an iterative process is not required for obtaining the solution in this case; however, unless the magnitudes of ρ and C are known a priori, the iterative process described in Appendix I will be required for determining the n values of $\hat{\sigma}^2_{\eta i}$.

In the special case where $\hat{\rho} = 0$, (2.34) reduces to a solution obtained by Kummell 2/ in 1879 and rediscovered by Gini 14/ in 1921. The worst fitting line obtained by reversing the sign

---

14/ Corrado Gini, "Sull' interpolazione di una retta quando i valori della variabile indipendente sono affeti da errori accidentali," Metron, vol. 1, no. 3, (1921), pp. 63-82.

before the radical in (2.34) is perpendicular to the best fitting line if, and only if, $(\hat{C}^2 - 1)[w_i x_i y_i] - \hat{\rho}\ \hat{C}\ \{[w_i(x_i^2 - y_i^2)]\} = 0$.

The locations (X'_i, Y'_i) along the fitted lines on Figs. 2.1, 2.2, and 2.3 are called the "adjusted values" of these points, and formulas for their determination are given in the following section of this paper. The case shown on Fig. 2.3 is of particular interest since the line obtained, which actually passes above the point for i = 1, would probably not have been anticipated as a possibility by an analyst not equipped with the present theory. Note also in this case that the order of ascending magnitude, X_1, X_2 and X_3 is changed to X'_1, X'_3 and then X'_2 in the case of the "adjusted" values, and this illustrates the difficulty involved in the a posteriori ordering of the data as required in the Wald[5] and Bartlett[6] methods of analysis.

Table 2.2 gives the values of a and b as calculated for the general case illustrated on Figs. 2.1, 2.2, and 2.3, and for ten other special cases. Each of the pairs of values of a and b shown in Table 2.2 are least squares solutions for the parameters of the lines fitting the same three points of Table 2.1 but with weights as described in the left-hand column of Table 2.2. Thus Table 2.2 provides examples of the effects of the several components in the weighting factor. Note in particular the large influence of r_i. Since the weights depend on a knowledge of the variances and covariances of a series of observations of the coordinates of the points, it is only when estimates of all of these variances and covariances are available that we may derive consistent conclusions from our least squares solution.

TABLE 2.2

No.	Method of Weighting	r_i	a	b
1	General Case (Table 2.1) Fig. 2.1	0	-0.038	+0.879
2	Weight independent of i (Pooled variance; $m_i = 5$)	0	-0.647	+0.996
3	General Case (Table 2.1) Fig. 2.2	-0.9	-0.030	+0.880
4	Weight independent of i (Pooled variance; $m_i = 5$)	-0.9	-0.657	+0.998
5	General Case (Table 2.1) Fig. 2.3	+0.9	+0.894	+0.627
6	Weight independent of i (Pooled variance; $m_i = 5$)	+0.9	-0.129	+0.899
7	$s^2_{Xi} = 0$; s^2_{Yi} (from Table 2.1)	0	+0.151	+0.811
8	$s^2_{Xi} = 0$; Weight independent of i (Pooled variance for s^2_Y; m_i constant)	0	-0.286	+0.929
9	$s^2_{Yi} = 0$; s^2_{Xi} (from Table 2.1)	0	-0.125	+0.938
10	$s^2_{Yi} = 0$; Weight independent of i (Pooled variance for s^2_X; m_i constant)	0	-1.077	+1.077
11	$\sigma^2_{\epsilon i} = \sigma^2_{\eta i}$; $\rho = 0$; w_i by (2.33) and (I-16)	0	-0.015	+0.877
12	$\sigma^2_{\epsilon i} = \sigma^2_{\eta i}$; $\rho = -0.9$; w_i by (2.33) and (I-16)	-0.9	-0.104	+0.893
13	$\sigma^2_{\epsilon i} = \sigma^2_{\eta i}$; $\rho = +0.9$; w_i by (2.33) and (I-16)	+0.9	0.423	+0.767

When Bartlett's test * was applied to the variance estimates, $s^2_{\epsilon i}$ and $s^2_{\eta i}$, in Table 2.1, it was found that variations from the pooled estimates larger than those in Table 2.1 would be expected with random normally distributed samples of this size with probabilities of 0.57 and 0.33, respectively. Thus, in the absence of physical reasons for expecting different variances, it appears that these might well be samples from the same population, and thus may be pooled to obtain an improved estimate. The pooled variances of Table 2.1 are $s^2_\epsilon = 16.666$ and $s^2_\eta = 18.333$. These are the values used in obtaining the solutions given in Table 2.2 for this case for the three assumed values of r.

Our general formula (2.15) for the slope of the least squares straight line is quite complicated, but it should be noted that this complication is inherent to the problem and arises from the use of a very large amount of detailed information, i.e., 5 numerical values of X_i, Y_i, s^2_{Xi}, s^2_{Yi}, and r_i associated with each point. It is thus natural to expect greater complication when more information is taken into account. In the special cases for which a smaller number of different data are necessary to describe the information available, the solutions (e.g., (2.20), (2.21), (2.28), (2.31), (2.32), and (2.34)) are correspondingly less complex. Furthermore, the above formulas are given primarily with the object of showing explicitly the ways in which the several components of the weights influence the slope. In a later section, matrix methods of solution are described, and these are recommended for use not only for the solution of the general case but for all of the simpler cases, since their use yields additional information as a by-product such as the standard errors of the parameters.

* This particular test is described in Section 8.

The expected value of b as given by (2.15) will, in general, not be equal to β and thus b is only a consistent estimate of β, and is not in the usual sense an unbiased estimate of β. However, b can be considered in another sense, more appropriate for the present problem, to be an unbiased estimate of β. We will demonstrate this in the special case of a straight line passing through the origin so that $\alpha \equiv 0$ and $Y_{io} = \beta X_{io}$. Assume also that the variances may be pooled and that $m_i = m$ so that the weights of all of the points are the same. Then (2.14) becomes:

$$b^2\{s_\epsilon^2[X_iY_i] - r\,s_\epsilon\,s_\eta\,[X_i^2]\} - b\{s_\epsilon^2[Y_i^2] - s_\eta^2[X_i^2]\} - \{s_\eta^2[X_iY_i] - r\,s_\epsilon\,s_\eta\,[Y_i^2]\} = 0 \qquad (2.35)$$

We will show that $b = \beta$ when s_ϵ^2, s_η^2, r, $[X_iY_i]$, $[X_i^2]$ and $[Y_i^2]$ in (2.35) are replaced by their expected values. In this case $X_i = X_{io} + \frac{1}{m}\Sigma\epsilon_{it}$, $Y_i = Y_{io} + \frac{1}{m}\Sigma\eta_{it}$ and thus:

$$E([X_i^2]) = [X_{io}^2] + (n/m)\,\sigma_\epsilon^2 \qquad (2.36)$$

$$E([Y_i^2]) = [Y_{io}^2] + (n/m)\,\sigma_\eta^2 = \beta^2[X_{io}^2] + (n/m)\,\sigma_\eta^2 \qquad (2.37)$$

$$E([X_iY_i]) = \beta[X_{io}^2] + (n/m)\,\rho\,\sigma_\epsilon\,\sigma_\eta \qquad (2.38)$$

When the above expected values are introduced in (2.35) we obtain:

$$\hat{b}^2[X_{io}^2]\{\sigma_\epsilon^2\beta - \rho\,\sigma_\epsilon\,\sigma_\eta\} - \hat{b}[X_{io}^2]\{\sigma_\epsilon^2\beta^2 - \sigma_\eta^2\} - [X_{io}^2]\{\sigma_\eta^2\beta - \rho\,\sigma_\epsilon\,\sigma_\eta\,\beta^2\} = 0 \qquad (2.39)$$

In the above $\hat{b}$ denotes the value of b with the expected values substituted for the sample values. The above quadratic has the unique solution $\hat{b} = \beta$ (corresponding to the positive square root) and thus, in this sense, b is an unbiased estimate of β. Note that this solution for $\hat{b}$ is independent of the sums $[X_{io}^2]$ and that b is an unbiased estimate of β (in the sense described above)

even for n = 1. The proof that $\hat{b} = \beta$ in the general case with $\alpha \neq 0$ and with unequal weights is left as an exercise for the student.

Note that the solution for the above special case by simple regression analysis may be written:

$$b' = [X_i Y_i]/[X_i^2] \tag{2.40}$$

If we introduce the expected values of $[X_i Y_i]$ and of $[X_i^2]$ in (2.40) we obtain:

$$\hat{b}' = \frac{\beta [X_{io}^2] + (n/m) \rho \sigma_\epsilon \sigma_\eta}{[X_{io}^2] + (n/m) \sigma_\epsilon^2} \tag{2.41}$$

Thus we see that regression analysis leads to an estimate of β which is biased in the above-described sense unless $\sigma_\epsilon^2 = 0$.

3. THE ADJUSTED VALUES: X_i', and Y_i'

Since the minimized sum S is invariant to the direction chosen for measuring the deviations of the observations, X_i, Y_i, from the line, we have no clue so far as to the expected locations along the fitted line of the adjusted values, X_i', Y_i'. In fact, we see that the parameters a and b have been determined without reference to such adjusted values which have usually played a prominent role in the derivation of previous least squares solutions. R. J. Adcock[15] and later Karl Pearson[16] defined the "closest fitting" line to be that which minimizes the sum of the squares of the perpendicular distances of the points, X_i, Y_i from the line, but we shall see that this criterion is equivalent to the method of least squares only under very special conditions of weighting.

Since our method of determining the adjusted values involves the use of an invariant statistical form identical to that used by Hotelling[17] [18] in generalizing Student's ratio to the multivariate case, it seems useful to digress somewhat at this point in order to summarize his analysis. Hotelling's generalized T distribution may be used with normally distributed observational data to define an elliptical equi-probability curve with coordinates, X*, Y*, and with its center at the sample mean, X_i, Y_i; these ellipses determine confidence regions for the population means X_{io}, Y_{io}, on the X*, Y* plane characterized by the probability, $1 - p_i(X^*, Y^*)$:

---

[15] R. J. Adcock, The Analyst, (Des Moines), 5, 1878, pp. 53-54.

[16] Karl Pearson, "On Lines and Planes of Closest Fit, to Systems of Points in Space," Phil. Mag., 6 Ser., vol. 2, Nov., 1901, pp. 559-572.

[17] Harold Hotelling, "The Generalization of Student's Ratio," Annals. of Math. Stat., vol. II, no. 3, Aug., 1931, pp. 360-378. The discussion of the degrees of freedom in this original article is not easy to follow. Better discussions are given in Reference 11, Chapter 3, by Harold Hotelling and in Reference 18, pages 407-409.

[18] Harald Cramer, "Mathematical Methods of Statistics," Princeton University Press, 1946.

$$p_i(X^*,\ Y^*) = \left(1 + \frac{T_i^2}{m_i - 1}\right)^{-(m_i - 2)/2} \qquad (3.1)$$

and in the limit as $m_i \to \infty$

$$p_i(X^*,\ Y^*) = e^{-T_i^2/2} \qquad (3.2)$$

where

$$T_i^2 = \frac{1}{(1 - r_i^2)} \left\{ \frac{(Y^* - Y_i)^2}{s_{Yi}^2} - \frac{2r_i(Y^* - Y_i)(X^* - X_i)}{s_{Yi}\, s_{Xi}} + \frac{(X^* - X_i)^2}{s_{Xi}^2} \right\} \qquad (3.3)^{\dagger}$$

If we make a practice of determining the location and size of these ellipses by solving (3.1) and (3.3) for the same probability, p_i, then—for a large number of such ellipses determined from normally distributed data—the population mean of the i^{th} point will lie within a fraction $(1 - p_i)$ of them and will lie outside of a fraction p_i of such ellipses. Although (3.1) and (3.2) are convenient for numerical calculations, it may be noted that $T_i^2(m_i - 2)/2(m_i - 1)$ is distributed as the Fisher-Snedecor variance ratio $F\{2,\ (m_i - 2)\}$; tables and graphs of the significance levels $F(\nu_1,\ \nu_2,\ p)$ are given in a companion paper.[19]

We will now define the adjusted point $(X_i',\ Y_i')$ corresponding to the i^{th} observed point $(X_i,\ Y_i)$ as the particular point $(X,\ Y)$ along the fitted line at which T_i^2 has its minimum value, G_i^2, and the probability, $p_i(X^*,\ Y^*)$ has its maximum value, $p_i' \equiv p_i(X_i',\ Y_i')$, i.e., the location corresponding to the values of X^* and Y^* which simultaneously minimize T_i^2 and satisfy the least squares fitted relation $Y = a + bX$. Using this relation to eliminate either Y^* or X^* from (3.3) and then differentiating the resulting expression with respect to the other variable, we obtain the following equations for the adjustments, i.e., the least squares estimated errors of each of the coordinates of each point.

$$X_i' - X_i = w_i(bs_{Xi}^2 - r_i s_{Xi} s_{Yi})V_{Yi} = w_i(bs_{Xi}^2 - r_i s_{Xi} s_{Yi})(Y_i - a - bX_i) \qquad (3.4)$$

$$Y_i' - Y_i = -w_i(s_{Yi}^2 - r_i bs_{Xi} s_{Yi})V_{Yi} = -w_i(s_{Yi}^2 - r_i bs_{Xi} s_{Yi})(Y_i - a - bX_i) \qquad (3.5)$$

---

† Note that T_i^2 is invariant to a homogeneous strain, translation, or rotation of the coordinate axes. See reference 17.

(19) L. E. Vogler and K. A. Norton, "Graphs and Tables of the Significance Levels $F(\nu_1, \nu_2, p)$ for the Fisher-Snedecor Variance Ratio," N. B. S. Report No. 5069, May, 1957.

Thus the adjusted values (X_i', Y_i') of the i^{th} point have been defined so as to yield consistent estimates of the population mean of this point since they correspond to minimum values of T_i^2, and the corresponding ellipses converge on the population means of the n points as all of the m_i approach infinity.

Upon substitution of the particular values (3.4) and (3.5) in (3.3) we obtain, after some algebraic manipulation:

$$G_i^2 = w_i V_{Yi}^2 = w_i(Y_i - a - bX_i)^2 = w_i\{(Y_i' - Y_i) - b(X_i' - X_i)\}^2 \quad (3.6)$$

The last of the above expressions is easily established by noting that we may subtract $Y_i' - a - bX_i' \equiv 0$ from $Y_i - a - bX_i$ without changing its value. Thus we see that the i^{th} component of the minimized sum S is identically equal to the minimized value, G_i^2, of Hotelling's invariant form T_i^2. Thus, regardless of the form of the statistical distribution of the observational data, our adjusted points are those lying on the least squares fitted relationship which have a minimum squared deviation from the observed points. In the case of data from normally distributed populations, it appears from the above derivation that our adjusted points may also be characterized as lying at their "most likely" locations along the fitted line.

Since G_i^2 is a particular value of T_i^2, it follows that G_i^2 and $S \equiv [G_i^2]$ will have the same invariance properties that Hotelling [17] has established for T_i^2; thus G_i^2 is invariant to an affine transformation, i.e., is invariant to a homogeneous strain, translation, or rotation of the coordinate axes. This proves the statement made in connection with (2.7). This invariance also makes all least squares solutions subject to a very significant limitation from the point of view of the physicist. Let us suppose that all of the observations of the random variables X_{it} and Y_{it} ($t = 1$ to ∞ and $i = 1$ to ∞) are subject to "systematic" errors in the form of constant biases u_o and v_o, respectively. Each of the population mean values X_{io} and Y_{io} will then differ from the "true" physical values X_{iT} and Y_{iT} by the amounts of these constant biases. It will clearly be impossible by least squares to detect such constant bias since the introduction of such a bias is equivalent to a translation of the X coordinate axis by an amount u_o, and the Y coordinate axis by an amount v_o, and our solution is invariant to

such a translation. Thus, throughout our development of least squares solutions it should be remembered that the functional relations estimated are between the population mean values X_{io} and Y_{io} and these may, by virtue of constant biases, differ from the "true" values X_{iT} and Y_{iT}.

If we set (3.3) equal to the values of G_i^2 given by (3.6), we obtain a formula for the particular ellipse with coordinates (X^*, Y^*) and of probability $(1 - p_i')$ of containing the "true" mean; such ellipses are shown on Figs. 2.1, 2.2, and 2.3 for the example defined in Table I and for the cases $r_i = o$, -0.9 and $+0.9$, respectively; note that these ellipses are tangent to the least squares fitted line at the adjusted locations, (X_i', Y_i'), of the n points. Thus we have established the remarkable result that our least squares solution, independently of the form assumed for the statistical distributions of the observational data, may be considered to be exactly equivalent to the solution of the geometrical problem of finding the parameters of the particular linear functional relationship which is tangent to the n ellipses defined by Hotelling's generalized T_i^2 associated with the n observational points and for which the sum $[T_i^2]$ has its minimum value $[G_i^2] \equiv S(a, b)$. † Since these ellipses each converge on the population mean values (X_{io}, Y_{io}) as the m_i approach infinity, and the latter are related by the functional relationship involving the true values α and β of the parameters, it follows that our least squares solution leads to consistent estimates of these parameters as the m_i approach infinity.

It is of some interest to consider under what circumstances the adjusted locations are on the perpendiculars to the line drawn from the points X_i, Y_i. It is evident from Figs. 2.1, 2.2, and 2.3 that this will be the case only when the ellipses degenerate to circles ($s_{Xi} = s_{Yi}$ and $r_i = 0$) or when the minor (or major) axes of the ellipses are parallel to the fitted line; such conditions of weighting would seldom be expected in practice. It should be noted that Karl Pearson did not claim [16] that his line of closest fit was determined

---

† This conclusion depends upon the assumption that unique solutions exist for both the least squares and the geometrical problems; all efforts made to date have failed to develop a set of data for which a unique solution does not exist.

by the method of least squares; he merely gave formulas for fitting the line which minimized the sum of the squares of the perpendicular distances of the points from it. Thus his solution would appear to have more geometrical than statistical significance.

4. THE GENERALIZED NORMAL EQUATIONS FOR DETERMINING THE u UNKNOWN PARAMETERS OF A FUNCTION F INVOLVING k' VARIABLES SOME OF WHICH MAY ENTER F NON-LINEARLY AND ASSUMING NO SYSTEMATIC ERRORS

When more than one unknown parameter is to be determined by least squares, the solution is most easily obtained by matrix methods (see Section 6) for solving the u "normal equations". In this section generalized normal equations will be developed which will be applicable to the problem of determining the u unknown parameters of a function F involving k' variables, only k of which are random.

It is assumed that n sets of observations ($n \geq u$) are available which define n points in the k' dimensional space, i.e., X_{ji} ($j = 1$ to k' and $i = 1$ to n), together with $nk (k \leq k')$ sample mean variances $s^2_{ji} = s^2_{\eta ji}/m_i$ ($j = 1$ to k) and $nk(k-1)/2$ different sample correlation coefficients r_{jhi} ($j \neq h$); actually a total of $k[m_i]$ individual observations of the k random variables are required to define these n sample points and their sample variances and covariances. The k random variables are each assumed to have statistical properties similar to those postulated for the two random variables in Section 2; thus $X_{jit} = X_{jio} + \eta_{jit}$ with the η_{jit} independent with respect to the subscripts i and t, $E(\eta^2_{jit}) = \sigma^2_{\eta ji}$, $E(\eta_{jit}\eta_{hit}) = \rho_{jhi}\sigma_{\eta ji}\sigma_{\eta hi}$, $j = 1$ to k, $h = 1$ to k, and $i = 1$ to n; $\sigma^2_{ji} = \sigma^2_{\eta ji}/m_i$. Furthermore, it is assumed that a function $F(X_{lio}, \ldots X_{jio}, \ldots X_{k'io}, a_1, \ldots a_p, \ldots a_u) = 0$ exists which passes through the population mean values X_{jio} ($j = 1$ to k') of these n points; this function involves u unknown parameters and may be non-linear in one or more of the variables or the parameters. †

---

† If it is known that the function $F(X_{jh})$ must pass through $v (v < u)$ specified points exactly (such points may be considered to be points of infinite weight, i.e., $\sigma^2_{jh} = 0$ ($j = 1$ to k and $h = 1$ to v)), then it will usually be possible to eliminate v of the u unknown parameters from F at the outset, and thus obtain a modified function H involving only $(u - v)$ parameters to be estimated by least squares. For example, if the straight line represented by the function $F \equiv \{Y - \alpha - \beta X\} = 0$ is known to pass through the point $(Y = 1, X = 1)$, we may fit the simpler function

The principle of least squares provides a method for determining consistent estimates, a_p, of the u values of the parameters, α_p, by minimizing the weighted sum of the squares of $F_i \equiv F(X_{1i}, \ldots X_{ji}, \ldots X_{k'i}, a_1, \ldots a_p, \ldots a_u)$ evaluated at the n points:

$$S = [w_i F_i^2] = [(F_i / s_i)^2] = \text{minimum} \qquad (4.1)$$

It will be convenient and will lead to consistent estimates of the parameters and to an invariant form for S if we define the weight w_i of the deviation F_i to be the reciprocal of its variance σ_i^2 as estimated by the following particular approximate formula:

$$(1/w_i) \equiv s_i^2 \equiv \sum_{j=1}^{k} \sum_{h=1}^{k} \left(\frac{\partial F_i}{\partial X_{ji}}\right)\left(\frac{\partial F_i}{\partial X_{hi}}\right) r_{jhi}\, s_{ji}\, s_{hi} \qquad (4.2)$$

In the above $r_{jhi} \equiv 1$ when $j = h$. Part of the approximation in the above expression arises from the fact that only the linear terms in the Taylor's series expansion of F_i were retained in its derivation. In many cases–all linear functional relations, for example–there may be only first derivatives, and then such an approximation is not involved. In other cases, provided we may also assume that the variables X_i are normally distributed, additional terms in the Taylor's series may be retained and improved weights determined; this latter method of obtaining improved weights is explained in a later section in connection with the discussion of a particular problem, but the improvement possible with its use is seldom worth the additional computational work required for its application.

$H \equiv \{Y - 1 - \beta(X - 1)\} = 0$ which now involves only the single parameter β. Such a reduction in the number of unknown parameters should be accomplished whenever this is feasible. Throughout this paper it has been implicitly assumed that F denotes this modified function H with the minimum number of unknown parameters.

Upon substitution of (4.2) in (4.1) we may formally express our solution of the general problem of least squares as the simultaneous solution of the following u equations which involve as unknowns only the u parameters a_p:

$$\frac{1}{2}\ \frac{\partial S}{\partial a_p} = 0 \qquad p = 1 \text{ to } u \qquad (4.3)$$

These u parameters will, in general, enter F_i, w_i, and consequently S in a non-linear way, and this will complicate the simultaneous solution of (4.3). The estimates of the parameters obtained by using (4.3) will always be consistent estimates of the true values of these parameters, although not necessarily statistically unbiased estimates. Furthermore, as was pointed out in the previous section, least squares leads to the best relations between the population mean values $X_{..}$, and these may differ from the "true" values by virtue of systematic biases.

We have seen in Section 2 that a direct solution for the parameters even in the simple linear case involves the evaluation of rather complicated expressions. The solution of the equations (4.3) is often even more complex, and it becomes desirable to develop simpler methods of approach. The following method is more general than that originally developed for this purpose by Gauss, but becomes identical to Gauss' method in those cases for which his solution is adequate. This generalization not only provides consistent estimates for the parameters, but also more accurate expressions for the probable errors of these estimated values of the parameters and for extrapolated values of the function in those cases for which either (1) the expression for the weighting factor is dependent on the parameters or (2) the function F is non-linear in one or more of the variables. The following method is similar, in some respects, to that of Kummell. 2/

4.1 The Generalized Normal Equations

First set $G_i^2 \equiv w_i F_i^2 \equiv (F_i/s_i)^2$. The value of S as given by (4.1) is then expanded in a Taylor's series:

$$S = [G_i^2]_o + \sum_{p=1}^{u} (a_p - a_{po}) \left[\frac{\partial G_i^2}{\partial a_p} \right]_o$$

$$+ \frac{1}{2} \sum_{s=1}^{u} \sum_{t=1}^{u} (a_s - a_{so})(a_t - a_{to}) \left[\frac{\partial^2 G_i^2}{\partial a_s \partial a_t} \right]_o \qquad (4.4)^\dagger$$

The above expression for S is exact to quantities of the second order in $(a_p - a_{po})$. The subscript o indicates that these sums are to be evaluated using approximate values of the estimated parameters, a_{po}. The solution (4.3) may now be written:

$$\frac{1}{2} \frac{\partial S}{\partial a_p} = \left[\frac{1}{2} \frac{\partial G_i^2}{\partial a_p} \right]_o + \sum_{s=1}^{u} (a_s - a_{so}) \left[\frac{1}{2} \frac{\partial^2 G_i^2}{\partial a_p \partial a_s} \right]_o = 0 \qquad (4.5)^\dagger$$

The u equations (4.5) are the generalized normal equations; it may be noted that they are linear in the u unknown quantities $(a_p - a_{po})$. Although they are approximate, this approximation may be made as small as the computer wishes to make it by means of the process of iterative solution, using successively closer approximations for the values a_{po} in evaluating the sums in the square brackets.

---

† In these equations G_i^2 is considered to be a function of the approximate values a_{po} used for its evaluation, but is considered in most of the remainder of this paper to be the particular minimized value obtained with $a_{po} = a_p$.

It will be convenient to write these generalized normal equations in the form:

$$(a_1 - a_{10})A_{11} + \ldots + (a_p - a_{po})A_{1p} + \ldots + (a_u - a_{uo})A_{1u} = A_{10}$$

$$(a_1 - a_{10})A_{p1} + \ldots + (a_p - a_{po})A_{pp} + \ldots + (a_u - a_{uo})A_{pu} = A_{po} \qquad (4.6)$$

$$(a_1 - a_{10})A_{u1} + \ldots + (a_p - a_{po})A_{up} + \ldots + (a_u - a_{uo})A_{uu} = A_{uo}$$

where: $$A_{pq} \equiv \left[\frac{1}{2}\frac{\partial^2 G_i^2}{\partial a_p \partial a_q}\right]_o = \left[\frac{\partial G_i}{\partial a_p}\cdot\frac{\partial G_i}{\partial a_q}\right]_o + \left[G_i \frac{\partial^2 G_i}{\partial a_p \partial a_q}\right]_o \qquad (4.7)$$

$$A_{po} \equiv -\left[\frac{1}{2}\frac{\partial G_i^2}{\partial a_p}\right]_o = -\left[G_i \frac{\partial G_i}{\partial a_p}\right]_o \qquad (4.8)$$

Note that $A_{pq} = A_{qp}$ and that the A_{pq} and A_{po} are specific numerical values which may be calculated from the given values of the observations and weights, together with the assumed values of a_{po}. The reason for dividing A_{pq} into two parts, as in (4.7), will appear in the next section. As successively better approximations to the parameters are determined by solving (4.6), these improved estimates may be used for re-evaluating the quantities A_{pq} and A_{po}. In the limit as a_{po} approaches a_p, A_{po} approaches zero; thus the values of A_{po} provide convenient measures of the degree of convergence attained at various stages of the iterative process.

These generalized normal equations differ from the usual normal equations used by all previous writers, with the exception of Kummell, since they introduce the necessary additional terms which arise from the differentiation of the weights.

As our first illustration (4.5) will be applied to the linear problem of Section 2 in which the observations of both variables are subject to error. Note that $F_i = Y_i - a - bX_i$ and that $(1/w_i)$ is given

by (2.6). Since there are only two parameters, the generalized normal equations for this problem may be expressed:

$$(a - a_o)A_{11} + (b - b_o)A_{12} = A_{10} \tag{4.9}$$

$$(a - a_o)A_{21} + (b - b_o)A_{22} = A_{20} \tag{4.10}$$

$$A_{11} = [w_i]_o \tag{4.11}$$

$$A_{12} = A_{21} = \left[w_i X_i + 2w_i^2 F_i(bs_{Xi}^2 - r_i s_{Xi} s_{Yi})\right]_o \tag{4.12}$$

$$A_{22} = \left[w_i X_i^2 + 4w_i^2 F_i X_i(bs_{Xi}^2 - r_i s_{Xi} s_{Yi})\right. \tag{4.13}$$

$$\left. + 4w_i^3 F_i^2(bs_{Xi}^2 - r_i s_{Xi} s_{Yi})^2 - w_i^2 F_i^2 s_{Xi}^2\right]_o \tag{4.14}$$

$$A_{10} = [w_i F_i]_o \tag{4.15}$$

$$A_{20} = \left[w_i X_i F_i + w_i^2 F_i^2 (bs_{Xi}^2 - r_i s_{Xi} s_{Yi})\right]_o \tag{4.16}$$

Comparing the above generalized normal equations for fitting points to a straight line with those obtained by the usual normal equations as given in most textbooks, we find that the latter yield only the leading terms in (4.12), (4.13), and (4.15) and will thus lead to an erroneous solution, i.e., to a solution which does not minimize S

in this case for which both variables are subject to error.† Note that a value for a_o can be determined for a given value of b_o by setting $A_{10} = 0$. However, if such a step is introduced regularly into the iterative process, it will not converge; it is necessary to use estimates of both a and b, obtained from the preceding step in the process, in calculating the coefficients for the following step.

Although equations in two unknowns such as (4.9) and (4.10) are easily solved by a variety of methods,[9/] [20/] [21/] Deming's systematic solution, as presented in Chapter IX of his book, [7/] is particularly to be recommended since he obtains as a by-product an evaluation of S. This method is described in Section 6 of this paper. If we let M denote the determinant formed from the coefficients of the parameters in the generalized normal equations (4.6)

$$M = |A_{pq}| \tag{4.16}$$

and let R_{pq} denote the corresponding elements of the reciprocal or image determinant, then the solution of the generalized normal equations may be expressed:

$$(a_p - a_{po}) = \sum_{q=1}^{u} R_{pq} A_{qo} \tag{4.17}$$

There will be one such equation for each value of p(p = 1 to u).

---

† For example, although Deming [7/] gives on page 184 the correct expression—equivalent to (2.34) in this paper—for b when $\sigma^2_{\epsilon i} = C^2 \sigma^2_{\eta i}$, this result will not be obtained from the normal equations developed in his book, since he neglected higher powers of the residuals in approximating his (3) on page 50 by (7) on page 53. We see by the above equations that the second order terms in the residuals, F_i, and terms arising from the differentiation of the weights control the solution when both variables are subject to error.

[20/] Paul S. Dwyer, "Linear Computations," John Wiley and Sons, 1951.

[21/] R. L. Anderson and T. A. Bancroft, "Statistical Theory in Research," McGraw-Hill Book Co., Inc., 1952.

It will become evident in the next section that the values, R_{pq}, of the elements of the reciprocal matrix are also required for the evaluation of the standard errors of the parameters, and consequently their determination serves a double purpose.

4.2 The Adjusted Values X'_{ji}

When the least squares estimates of the u parameters, a_p, have been determined, the following formula may be used to determine the k' coordinates of the adjusted values, X'_{ji}, (j = 1 to k') for each of the n points (i = 1 to n):

$$X'_{ji} - X_{ji} = - w_i F_i s_{ji} \sum_{h=1}^{k} r_{jhi} \left(\frac{\partial F_i}{\partial X_{hi}} \right) s_{hi} \qquad (4.18)$$

Since $\sigma^2_{ji} = 0$ for the non-random variables, $X'_{ji} = X_{ji}$ for j = k+1 to k'.

The above general equation for the adjusted values was derived by the same general procedure used in deriving (3.4) and (3.5).

We will now introduce Hotelling's invariant form T_i^2 for the multivariate case and thus establish the exact equivalence of our general least squares solution to the geometrical problem of finding the parameters of the particular functional relationship whose curve in the k' dimensional space is tangent to the n hyper-ellipsoids defined by the T_i^2 associated with the n observational points and for which the sum $[T_i^2]$ has its minimum value $[G_i^2] \equiv S(a_1, \ldots, a_p, \ldots, a_u)$. Let $L_i = |r_{jhi} s_{ji} s_{hi}|$ denote the value of the determinant of the observed moment matrix describing the i^{th} data point, and let R_{jhi} denote the corresponding elements of the image, or reciprocal, determinant. Note that the elements of the determinant L_i are determined by the m_i observations taken under the fixed experimental conditions corresponding to the i^{th} point:

$$r_{jhi} s_{ji} s_{hi} \equiv \frac{1}{m_i(m_i - 1)} \sum_{t=1}^{m_i} (X_{jit} - X_{ji})(X_{hit} - X_{hi}) \qquad (4.19)$$

We may now express Hotelling's invariant form T_i^2 in the following form:

$$T_i^2 = \sum_{j=1}^{k} \sum_{h=1}^{k} R_{jhi} (X^*_{ji} - X_{ji})(X^*_{hi} - X_{hi}) \qquad (4.20)$$

Each value of T_i^2 defines a hyper-ellipsoid in the k dimensional space with its center at the location X_{ji} (j = 1 to k) of the i^{th} observation point. Regardless of the form of the statistical distribution of the observational data, all points X^*_{ji} (j = 1 to k) on the surfaces of these hyper-ellipsoids have the same weighted squared deviation from the mean observed values, X_{ji}. For data from normally distributed populations, the exact sampling distribution of T_i^2 is known, and in this case the surfaces of these hyper-ellipsoids can be characterized as surfaces of equi-probability. Thus the quantity $T_i^2 (m_i - k)/k(m_i - 1)$ is distributed as the Fisher-Snedecor variance ratio $F(k, m_i - k)$. If we set $T_i^2 (m_i - k)/k(m_i - 1) = F(k, m_i - k, p_i)$ we may construct hyper-ellipsoidal confidence regions X^*_{ji} (j = 1 to k) centered on the observed mean X_{ji} (j = 1 to k) which are expected to contain the population mean X_{jio} (j = 1 to k) with a confidence $(1 - p_i)$.

If we substitute the adjusted values X'_{ji} (j = 1 to k) as given by (4.18) for the X^*_{ji} in (4.20) we obtain an expression for the particular value G_i^2 corresponding to the particular hyper-ellipsoid centered on the i^{th} point which is just tangent to the least squares fitted functional relationship at the adjusted location of the i^{th} point:

$$G_i^2 = w_i F_i^2 \qquad (4.21)$$

The reader should make the above substitution, carry out the necessary algebraic manipulations, and thus satisfy himself as to the generality of (4.21). Since these hyper-ellipsoids converge on the true location of the i^{th} point as m_i approaches infinity, we see that our least squares solution yields consistent estimates of the adjusted values and, since the adjusted values lie on the fitted function, consistent estimates of its parameters. The above geometrical argument evidently

cannot be used to show that consistent estimates are obtained when the m_i are finite and n approaches infinity. In this latter case the hyper-ellipsoids will each have different finite magnitudes and, although it seems plausible, it is not intuitively clear that consistent estimates of the parameters of the functional relationship will necessarily be obtained as the number n of these finite hyper-ellipsoids approaches infinity. The proof of this latter consistency property of our solution is not available, but it is intuitively clear that the n points must be more or less uniformly distributed over an adequate range of the variates if consistent estimates of the parameters are to be obtained as n approaches infinity.

5. ESTIMATES OF THE EXPECTED VARIANCES OF THE ESTIMATED PARAMETERS AND OF THE ESTIMATED FUNCTIONAL RELATIONSHIP

One of the major advantages of least squares over some other methods of fitting data to an assumed functional relationship is the possibility of determining standard errors for the estimated parameters as well as standard errors for the functional relationship itself. The theory of all this goes back to Gauss who showed how the standard errors of the parameters could be determined directly from the reciprocal matrix of the coefficients of the parameters in the normal equations. Good discussions of this theory are given by Whittaker and Robinson [22] and by Chauvenet. [23] The modifications to this theory which are necessary when the generalized normal equations are used will be given below. We will see that the basis for the theory is that the normal equations provide a set of linear relations between the errors F_i and the estimates of the parameters. In the particular case of fitting data to a linear relationship in any number of dimensions but with the values of only one of the variables subject to error, the estimates of the parameters and the errors F_i of the function are linearly related, and Neyman and David [24] have presented a proof that Gauss' method of least squares provides the best linear unbiased estimates of the parameters and their standard errors. We have seen above, however, that the estimates of the parameters are not exactly linearly related to the errors F_i in many other cases, and we conclude that the Gauss method will give only approximate results in these other cases. An improvement in accuracy is gained by using the generalized rather than the usual normal equations in accordance with the method outlined below, but there appears to be no simple way to obtain completely unbiased estimates of the standard errors of the parameters in the general case.

---

[22] E. T. Whittaker and G. Robinson, "The Calculus of Observations," Blackie and Son, Limited, London, 1924, pp. 226-259.

[23] William Chauvenet, "Manual of Spherical and Practical Astronomy," J. B. Lippincott Company, Philadelphia, 1863, vol. II, pp. 469-566.

[24] J. Neyman and F. N. David, "Extension of the Markoff Theorem on Least Squares," Statist. Research Mem., 2(1938), p. 105.

The solution (4.17) of the generalized normal equations may be written:

$$(a_{po} - a_p) = R_{pl}\left[G_i \frac{\partial G_i}{\partial a_1}\right]_o + \ldots + R_{pu}\left[G_i \frac{\partial G_i}{\partial a_u}\right]_o \qquad (5.1)$$

Define the new variable η_{pi}:

$$\eta_{pi} = R_{pl}\left(\frac{\partial G_i}{\partial a_1}\right)_o + \ldots + R_{pu}\left(\frac{\partial G_i}{\partial a_u}\right)_o \qquad (5.2)$$

Using this variable, (5.1) may now be expressed:

$$a_p = a_{po} - [\eta_{pi} G_i]_o \qquad (5.3)$$

Thus we have expressed a_p as a linear function of the normalized deviations, $G_{\gamma_i} \equiv (F_i/s_i)$. In accordance with the rules for the propagation of variance, we obtain the following approximate expression for the square of the standard error of a_p:

$$s^2_{ap} = [\eta^2_{pi} s^2_{Gi}]_o \qquad (5.4)$$

A pooled estimate of s^2_G, independent of the within-group estimate of variance, s^2_i, may be obtained from the following:

$$s_G^2 = \frac{1}{n-u}\,[G_1^2] = S/(n-u) \qquad (5.5)^{\dagger}$$

The above between-groups estimate of variance was obtained by increasing the mean square value of the n expected deviations, G_i, from the fitted functional relationship by the factor $n/(n-u)$ to allow for the bias resulting from the fact that u parameters were determined in fitting the n points to the function. A discussion of the distribution of s_G^2 is presented in Section 7. Note that G_i has been normalized so that its expected variance is approximately the same for all values of i even in those cases where the observed variances, s_i^2, vary from point to point by more than would be expected for sample variances from the same population. Since s_G^2 is a constant, independent of i, s_{Gi}^2 may be replaced by $S/(n-u)$ in (5.4) and removed from under the summation sign; we then obtain the following estimate of the variance of a_p:

$$s_{ap}^2 = \frac{S}{n-u}\,[\eta_{pi}^2]_o = \frac{S}{n-u}\left[\eta_{pi}\left\{R_{pl}\,\frac{\partial G_i}{\partial a_1} + \ldots + R_{pu}\,\frac{\partial G_i}{\partial a_u}\right\}\right]_o \qquad (5.6)$$

Remembering that MR_{pq} is the cofactor of A_{pq} in the determinant M, (5.6) may be expressed:

---

† In the special case where the number of points equals the number of unknown parameters (n = u) the function F can usually be fitted exactly to these n points so that $F_i = 0$ (i = 1 to u); thus $[G_i^2] = 0$ and the between-groups estimate of variance $s_G^2 = 0$ even though $s_i^2 \neq 0$ for all n points. In this case the confidence intervals for the parameters and the confidence region for the fitted function must be determined from the within-groups estimate of the variance. An example of this use of the within-groups estimate of variance is given in Section 12.

$$Ms^2_{ap}(n-u)/S = M\left[\eta^2_{pi}\right]_o = \left|\begin{array}{ccc} A_{11} & \cdots\cdots & A_{1u} \\ \cdots & \cdots & \cdots \\ \eta_{p1}\dfrac{\partial G_1}{\partial a_1} & \cdots\cdots & \eta_{p1}\dfrac{\partial G_1}{\partial a_u} \\ \cdots & \cdots & \cdots \\ A_{u1} & \cdots\cdots & A_{uu} \end{array}\right|_0 + \ldots.$$

$$+ \left|\begin{array}{ccc} A_{11} & \cdots\cdots & A_{1u} \\ \cdots & \cdots & \cdots \\ \eta_{pn}\dfrac{\partial G_n}{\partial a_1} & \cdots\cdots & \eta_{pn}\dfrac{\partial G_n}{\partial a_u} \\ \cdots & \cdots & \cdots \\ A_{u1} & \cdots\cdots & A_{uu} \end{array}\right|_0$$

$$= \left|\begin{array}{ccc} A_{11} & \cdots\cdots & A_{1u} \\ \cdots & \cdots & \cdots \\ \left[\eta_{pi}\dfrac{\partial G_i}{\partial a_1}\right] & \cdots\cdots & \left[\eta_{pi}\dfrac{\partial G_i}{\partial a_u}\right] \\ \cdots & \cdots & \cdots \\ A_{u1} & \cdots\cdots & A_{uu} \end{array}\right|_0 \qquad (5.7)$$

that is, the determinant $|A_{ij}|$ with the p^{th} row replaced by

$$\left[\eta_{pi}\frac{\partial G_i}{\partial a_1}\right] \cdots\cdots \left[\eta_{pi}\frac{\partial G_i}{\partial a_u}\right].$$

When (5.2) is substituted in (5.7) we obtain:

$$
\begin{aligned}
Ms_{ap}^2\,(n-u)/S = R_{pl}
&\begin{vmatrix}
A_{11} & \cdots & A_{1u} \\
\cdots & \cdots & \cdots \\
\left[\frac{\partial G_i}{\partial a_1}\ \frac{\partial G_i}{\partial a_1}\right] & \cdots & \left[\frac{\partial G_i}{\partial a_1}\ \frac{\partial G_i}{\partial a_u}\right] \\
\cdots & \cdots & \cdots \\
A_{u1} & \cdots & A_{uu}
\end{vmatrix}_0 + \ldots \\
\ldots + R_{pp}
&\begin{vmatrix}
A_{11} & \cdots & A_{1u} \\
\cdots & \cdots & \cdots \\
\left[\frac{\partial G_i}{\partial a_p}\ \frac{\partial G_i}{\partial a_1}\right] & \cdots & \left[\frac{\partial G_i}{\partial a_p}\ \frac{\partial G_i}{\partial a_u}\right] \\
\cdots & \cdots & \cdots \\
A_{u1} & \cdots & A_{uu}
\end{vmatrix}_0 + \ldots \\
\ldots + R_{pu}
&\begin{vmatrix}
A_{11} & \cdots & A_{1u} \\
\cdots & \cdots & \cdots \\
\left[\frac{\partial G_i}{\partial a_u}\ \frac{\partial G_i}{\partial a_1}\right] & \cdots & \left[\frac{\partial G_i}{\partial a_u}\ \frac{\partial G_i}{\partial a_u}\right] \\
\cdots & \cdots & \cdots \\
A_{u1} & \cdots & A_{uu}
\end{vmatrix}_0
\end{aligned}
\qquad (5.8)
$$

Consider first the case where the second term in (4.7) is zero for all combinations of pq. In this case each of the determinants in (5.8) except the one multiplied by R_{pp}, is equal to zero since two rows are then identical. The determinant multiplied by R_{pp} is just equal to M, and we obtain for this special case:

$$s_{ap}^2 = \frac{S}{n - u} R_{pp} \tag{5.9}$$

The above elegant solution for the estimated variance of the estimated parameter a_p (p = 1 to u) was first obtained by Gauss. The corresponding result using the generalized normal equations may be obtained by substituting the following (from 4.7) into (5.8):

$$\left[\frac{\partial G_i}{\partial a_p} \cdot \frac{\partial G_i}{\partial a_q}\right]_o = A_{pq} - \left[G_i \frac{\partial^2 G_i}{\partial a_p \partial a_q}\right]_o$$

$$s_{ap}^2 = \frac{S}{n - u}\left\{R_{pp} - \sum_{v=1}^{u} R_{pv} \sum_{w=1}^{u} R_{pw}\left[G_i \frac{\partial^2 G_i}{\partial a_v \partial a_w}\right]_o\right\} \tag{5.10}$$

In similar fashion we may determine the covariance:

$$r_{pq}\, s_{ap}\, s_{aq} = \frac{S}{n - u}\left\{R_{pq} - \sum_{v=1}^{u} R_{pv} \sum_{w=1}^{u} R_{qw}\left[G_i \frac{\partial^2 G_i}{\partial a_v \partial a_w}\right]_o\right\} \tag{5.11}$$

The above may also be expressed after considerable manipulation as follows:

$$r_{pq}\, s_{ap}\, s_{aq} = \frac{S}{n - u} \sum_{v=1}^{u} \sum_{w=1}^{u} R_{pv} R_{qw}\left[\frac{\partial G_i}{\partial a_v} \cdot \frac{\partial G_i}{\partial a_w}\right]_o \tag{5.12}$$

Equations (5.11) and (5.12) may be used for calculating the variance (5.10) by noting that $r_{pq} \equiv 1$ when p = q.

In numerical work it is desirable to use both (5.11) and (5.12) for evaluating these variances since this provides a very valuable check on the calculations.

Finally a between-groups estimate, s_F^2, of the variance of the function F, corresponding to a given point, X_j(j = 1 to k), on the fitted relation with its parameters determined by least squares from the n observations, may be determined by applying the usual rules for the propagation of variance [7] to the function, $F(X_1, \ldots X_j, \ldots X_k', a_1, \ldots a_p, \ldots a_u)$:

$$s_F^2 = \sum_{p=1}^{u} \sum_{q=1}^{u} r_{pq} s_{ap} s_{aq} \left(\frac{\partial F}{\partial a_p} \right)_o \left(\frac{\partial F}{\partial a_q} \right)_o \qquad (5.13)$$

In the above $r_{pq} \equiv 1$ when p = q. The approximation in the above, aside from the fact that $r_{pq} s_{ap} s_{aq}$ is necessarily the sample rather than the population value, arises from the fact that only the linear terms in the Taylor's series expansion of F were retained in its derivation. In evaluating the derivatives $\left(\frac{\partial F}{\partial a_p} \right)_o$ and $\left(\frac{\partial F}{\partial a_q} \right)_o$ it is important to notice that the same estimated values of the parameters are to be used as were used in calculating R_{pq}. It is then possible to show that the right hand member of (5.13) is inherently greater than or equal to zero. The iterative process of using successively better estimated values of the parameters should be continued until two successive determinations of s_F^2 do not differ appreciably. Comparing (5.9) with (5.10), we see that the correction to Gauss' solution will vanish in the limit as the deviations, G_i, approach zero; thus we see that (5.11) or (5.12) provide the first order correction to allow for the finite sizes of the deviations which, in Gauss' solution were assumed to be negligibly small. The estimate of variance s_F^2 given by (5.13) represents an estimate of the square of the standard error of the function determined by our sample of n sets of observations relative to a hypothetical "true" value for the function which one would expect to determine if all m_i(i = 1 to n) were allowed to approach infinity.

6. SYSTEMATIC METHOD FOR THE NUMERICAL SOLUTION OF THE GENERALIZED NORMAL EQUATIONS

The method described in this section follows closely that proposed by Deming in Chapter IX of reference 7, and the student is referred to his book for proofs of the statements made in this section. The method will be discussed in detail for the case of fitting a straight line involving 2 unknown parameters. An explicit form for determining 3 unknown parameters is then given, and it will be clear from these examples how the method may be extended to more than 3 unknown parameters. Incidental to the determination of the u parameters, this method also provides u + 1 different values of the sum S of the weighted squares of the deviations from the function: thus $S(a_{po})$ is the first estimate of S determined by using the estimated values a_{po}, and then u successively smaller values of S are obtained as the u least squares values of a_p are determined successively and used in the calculations. Furthermore, the method also yields estimates for the variances and covariances of the u estimates of the parameters.

6.1 Fitting a Straight Line

First it is necessary to determine approximate values a_o and b_o for the parameters. This may be done in a variety of ways, but plotting the data and fitting a line by eye is usually the simplest. In some cases these estimates need not be very accurate and, in fact, it will sometimes be convenient simply to let $a_o = b_o = 0$; however, if the relative weights depend on the value of b, it will be desirable to use a good estimate of b_o at the outset. The estimated values a_o and b_o are entered at the top of the tabulation form given on the next page. Next, the weights w_i (i = 1 to n) are calculated and used to calculate the 5 sums A_{11}, A_{12}, A_{22}, A_{10}, and A_{20} defined on page 4.6 together with $S(a_o, b_o) \equiv [w_i F_i^2]_o$. These six values are then entered on the tabulation form. The Arabic numerals in parentheses (1), (2), ... (35) indicate the preferred order of calculation and entry on the tabulation form of the 35 numbers required for a complete solution. Row 4 is now obtained by multiplying the values in Row I by $(-A_{12}/A_{11})$; e.g., item (7) is $(-A_{12}/A_{11})$, item (8) is $(-A_{12}/A_{11}) \cdot A_{12}$

The Iterative Solution of the Generalized Normal Equations for Fitting a Straight Line

a_o = b_o =

Row	$(a - a_o)$	$(b - b_o)$	= C_o	C_1	C_2
I	(1) A_{11}	(2) A_{12}	(3) A_{10}	1	0
2		(4) A_{22}	(5) A_{20}	0	1
3	How Obtained		(6) $[w_i F_i^2]_o \equiv S(a_o, b_o)$	0	0
4	$-\frac{A_{12}}{A_{11}} \cdot$ (I)	(8)	(9)	(7)	0
II	2 + 4	(10) B_{22}	(11) B_{20}	(12)	1
6	$-\frac{A_{10}}{A_{11}} \cdot$ (I)		(14)	(13)	
7	$-\frac{B_{20}}{B_{22}} \cdot$ (II)		(16)		(15)
III	3 + 6 + 7		(17) S(a, b)		
9	I solved for $(a - a_o)$	(19) $(a - a_o)$	(24) $(a - a_o)$	(21) R_{11}	(23) R_{12}
10	II solved for $(b - b_o)$	(18) $(b - b_o)$	(25) $(b - b_o)$	(20) R_{21}	(22) R_{22}
11	$M = A_{11} \cdot B_{22}$	(28) M	(26) $a = a_o + (a - a_o)$	(30) s_a^2	
12	$s_G^2 = \frac{S(a, b)}{n - 2}$	(29) s_G^2	(27) $b = b_o + (b - b_o)$	(31) $r_{ab} s_a s_b$	(32) s_b^2
13				(33) s_a^2	
14				(34) $r_{ab} s_a s_b$	(35) s_b^2

and item (9) is $(-A_{12}/A_{11})\cdot A_{10}$. Next, items (10), (11), and (12) are obtained by adding the numbers immediately above in Rows 2 and 4. Item (13) is $(-A_{10}/A_{11})$ and item (14) is $(-A_{10}/A_{11})\cdot A_{10}$. Item (15) is $(-B_{20}/B_{22})$ and item (16) is $(-B_{20}/B_{22})\cdot B_{20}$. Item (17) is obtained by adding the numbers immediately above it in Rows 3, 6, and 7, i.e., $S(a, b) = S(a_o, b_o) + (14) + (16)$. Note that (14) and (16) are necessarily negative, and (14) represents the reduction of S arising from the change of a_o to a while (16) represents the additional reduction of S arising from the change of b_o to b. Next, item (18) is determined by considering Row II to be the following equation for $(b - b_o)$:

$$B_{22}(b - b_o) = B_{20} \tag{6.1}$$

Thus $(b - b_o)$ = {item (11)/item (10)}, and this is item (18). Next, using this value of $(b - b_o)$, Row I may now be solved for $(a - a_o)$:

$$A_{11}(a - a_o) + A_{12}(b - b_o) = A_{10} \tag{6.2}$$

At this point in the calculations, a decision should be made regarding the adequacy of the original estimates of a_o and b_o; thus we should determine the new estimates $a_o + (a - a_o)$ and $b_o + (b - b_o)$ of a and b, respectively, and repeat the entire procedure outlined above using these new estimates, the revised results being entered on a new tabulation form. In many cases the second set of calculations will lead to values of $(a - a_o)$ and $(b - b_o)$ equal to zero; if not, the above procedure should be repeated until the calculated values of $(a - a_o)$ and $(b - b_o)$ are sufficiently near to zero.

We may now proceed with the calculations of items (20) through (35). Thus R_{21} and R_{11} are determined by considering that they are the values of $(b - b_o)$ and $(a - a_o)$, respectively, which would be obtained by replacing column C_o by column C_1; thus R_{21} = {item 12/B_{22}} and $R_{11} = \{(1 - R_{21}A_{12})/A_{11}\}$. Similarly, R_{22} and R_{12} are determined by replacing column C_o by column C_2; thus $R_{22} = 1/B_{22}$ and and $R_{12} = - R_{22}A_{12}/A_{11}$. Since $R_{12} = R_{21}$ we obtain a check on our

calculations at this point; thus item (20) must be equal to item (23). Furthermore, we may now check $(a - a_o)$ and $(b - b_o)$ and enter the results as (24) and (25) by using the equations (4.17):

$$(a - a_o) = R_{11}A_{10} + R_{12}A_{20} \tag{6.3}$$

$$(b - b_o) = R_{21}A_{10} + R_{22}A_{20} \tag{6.4}$$

We have already determined that $(a - a_o)$ and $(b - b_o)$ are near zero, and the above equations simply indicate that this will be true only to the extent that A_{10} and A_{20} are near zero. Items (26) and (27) represent our final estimates of a and b.

Item (28) is the determinant M defined by (4.16):

$$M = A_{11} \cdot B_{22} \tag{6.5}$$

When M is small it is indicative of a condition of near indeterminacy in the solution for the parameters; this possible difficulty is discussed by Deming [7] who also gives other references.

Item (29) is the estimate s_G^2; if s_G^2 is much larger than unity, it will be desirable to use the tests described in Section 7 for determining the statistical plausibility of the solution. A very large or very small value of s_G^2 may, of course, simply indicate numerical errors in the calculations in some cases.

Items (30), (31), and (32) are obtained by using (5.10):

$$s_a^2 = s_G^2 \left\{ R_{11} - 2R_{11}R_{12} \left[G_i \frac{\partial^2 G_i}{\partial a \partial b} \right]_o - R_{12}^2 \left[G_i \frac{\partial^2 G_i}{\partial b^2} \right]_o \right\} \tag{6.6}$$

$$s_b^2 = s_G^2 \left\{ R_{22} - 2R_{22}R_{12} \left[G_i \frac{\partial^2 G_i}{\partial a \partial b} \right]_o - R_{22}^2 \left[G_i \frac{\partial^2 G_i}{\partial b^2} \right]_o \right\} \tag{6.7}$$

$$r_{ab} s_a s_b = s_G^2 \left\{ R_{12} - (R_{11} R_{22} + R_{12}^2) \left[G_i \frac{\partial^2 G_i}{\partial a \partial b} \right]_o - R_{12} R_{22} \left[G_i \frac{\partial^2 G_i}{\partial b^2} \right]_o \right\} \tag{6.8}$$

We see by (4.7) that the terms involving G_i in the above expression are simply components of A_{12} and A_{22} and, by appropriately arranging the calculations, can be obtained during the course of the evaluation of A_{12} and A_{22}; thus these terms may be expressed as follows:

$$\left[G_i \frac{\partial^2 G_i}{\partial a \partial b} \right]_o = \left[w_i^2 F_i (b s_{Xi}^2 - r_i s_{Xi} s_{Yi}) \right]_o \tag{6.9}$$

$$\left[G_i \frac{\partial^2 G_i}{\partial b^2} \right]_o = \left[2 w_i^2 F_i X_i (b s_{Xi}^2 - r_i s_{Xi} s_{Yi}) + 3 w_i^3 F_i^2 (b s_{Xi}^2 - r_i s_{Xi} s_{Yi})^2 - w_i^2 F_i^2 s_{Xi}^2 \right]_o \tag{6.10}$$

As a valuable numerical check on the above calculations, items (33), (34), and (35) may next be calculated using the following expressions:

$$s_a^2 = s_G^2 \left\{ R_{11}^2 \left[\left(\frac{\partial G_i}{\partial a} \right)^2 \right]_o + 2 R_{11} R_{12} \left[\frac{\partial G_i}{\partial a} \cdot \frac{\partial G_i}{\partial b} \right]_o + R_{12}^2 \left[\left(\frac{\partial G_i}{\partial b} \right)^2 \right]_o \right\} \tag{6.11}$$

$$s_b^2 = s_G^2 \left\{ R_{12}^2 \left[\left(\frac{\partial G_i}{\partial a} \right)^2 \right]_o + 2 R_{22} R_{12} \left[\frac{\partial G_i}{\partial a} \cdot \frac{\partial G_i}{\partial b} \right]_o + R_{22}^2 \left[\left(\frac{\partial G_i}{\partial b} \right)^2 \right]_o \right\} \tag{6.12}$$

$$r_{ab} s_a s_b = s_G^2 \left\{ R_{11} R_{12} \left[\left(\frac{\partial G_i}{\partial a} \right)^2 \right]_o + (R_{11} R_{22} + R_{12}^2) \left[\frac{\partial G_i}{\partial a} \cdot \frac{\partial G_i}{\partial b} \right]_o + R_{12} R_{22} \left[\left(\frac{\partial G_i}{\partial b} \right)^2 \right]_o \right\} \tag{6.13}$$

The terms involving G_i in the above are the remaining components of A_{11}, A_{12}, and A_{22}, thus:

$$\left[\left(\frac{\partial G_i}{\partial a}\right)^2\right]_o = A_{11} = [w_i]_o \qquad (6.14)$$

$$\left[\frac{\partial G_i}{\partial a}\cdot\frac{\partial G_i}{\partial b}\right]_o = [w_i X_i + w_i^2 F_i(b s_{Xi} - r_i s_{Xi} s_{Yi})]_o \qquad (6.15)$$

$$\left[\left(\frac{\partial G_i}{\partial b}\right)^2\right]_o = [w_i X_i^2 + 2w_i^2 F_i X_i(bs_{Xi} - r_i s_{Xi} s_{Yi}) + w_i^3 F_i^2(bs_{Xi} - r_i s_{Xi} s_{Yi})^2]_o \qquad (6.16)$$

This completes the general description of the calculations required for the tabulation forms.

In most cases the procedure up to item (19) described above need not be carried out more than two times. However, the following example indicates how slowly the iterative process may converge in an extreme case. Thus, for the linear example of Table 2.1 with $r_i = 0.9$ and the variances not pooled, six repetitions of the above-described iterative process were required before $(a - a_o)$ and $(b - b_o)$ were considered negligible. Table 6.1 lists the results of calculations on the form using successively better estimates a_o and b_o. The initial value of b_o was arbitrarily set equal to 1 and a corresponding value of a_o determined by setting $A_{10} = 0$; in subsequent steps, however, the previous estimates of both a_o and b_o were used. In the fourth column, the values of a_o and b_o were taken to be approximately the averages of the two previous estimates. Note that only the values in the last column represent the least squares solution, and then only to the extent that they may be considered to be calculated with sufficient accuracy. In the present problem, since the standard errors of the parameters—and thus the function—were found to be so large, it would have been satisfactory to use column 4 as a final solution. The process was carried further to show how it converged. Note that the calculations in the first five columns in Table 6.1 need only have been carried to the point at which the new estimates of a and b became available; the other calculations shown on Table 6.1 are included simply to indicate the behaviors of these items in the iterative process.

Table 6.1

The Iterative Solution of the Normal Equations for the Linear Example of Table 2.1, $r_i = +0.9$, Variances Not Pooled

	1	2	3	4	5	6
a_o	-0.596522	-0.182254	0.911686	0.869	0.883217	0.893603
b_o	1.000000	0.865564	0.582203	0.614	0.629517	0.626848
w_1	2.200627	1.832055	1.166980	1.228491	1.259757	1.254321
w_2	1.307844	1.667553	1.378206	1.471759	1.515409	1.508034
w_3	0.876429	0.824808	0.511546	0.545986	0.563365	0.560352
A_{10}	0	0.177771	0.596571	0.278496	-0.007691	-0.000016
$A_{11} = \left[\left(\frac{\partial G_i}{\partial a}\right)^2\right]$	4.384900	4.324416	3.056732	3.246236	3.338531	3.322707
$\left[\left(\frac{\partial G_i}{\partial a}\right)\left(\frac{\partial G_i}{\partial b}\right)\right]_o$	16.385947	18.167640	14.089969	15.314429	15.996677	15.918974
$\left[G_i \frac{\partial^2 G_i}{\partial a \partial b}\right]_o$	-2.873803	-2.100252	-0.605595	-0.340995	-0.122211	-0.120688
A_{12}	13.512144	16.067388	13.484374	14.973434	15.874466	15.798286
A_{20}	-1.683430	-0.110643	3.049609	1.393608	-0.054371	-0.000033
$\left[\left(\frac{\partial G_i}{\partial b}\right)^2\right]$	87.637878	98.910118	77.540889	86.262304	91.334105	90.853235
$\left[G_i \frac{\partial^2 G_i}{\partial b^2}\right]_o$	-33.477844	-36.490220	-11.477707	-10.170050	-9.182016	-9.024872
A_{22}	54.160034	62.419898	66.063182	76.092254	82.152089	81.828363
$S(a_o, b_o) = [G_i^2]_o$	3.671059	3.714150	3.255054	3.150050	3.125100	3.125042
S(a, b)	3.714150	3.255054	3.150050	3.125100	3.125042	
a	-0.182254	0.911686	0.826615	0.883217	0.893603	0.893567
b	0.865564	0.582203	0.645728	0.629517	0.626848	0.626854
M	54.908298	11.768649	20.109100	22.809688	22.268625	22.305834
R_{11}	0.986372	5.303914	3.285238	3.335962	3.689141	3.668474
R_{12}	-0.246086	-1.365270	-0.670561	-0.656451	-0.712862	-0.708258
R_{22}	0.079859	0.367452	0.152007	0.142318	0.149921	0.148961
s_a^2	5.941998	159.3493	18.807762	19.609167	24.102001	23.651773
s_b^2	0.662294	11.83746	0.956191	0.896478	1.031835	1.011737
$r_{ab} s_a s_b$	-1.848725	-43.319318	-4.120100	-4.087972	-4.788318	-4.793551

The Solution of the Generalized Normal Equations for the Linear Example of Table 2.1, $r_i = 0.9$, Variances Not Pooled

$a_o = 0.893603$ $\qquad$ $b_o = 0.626848$

Row	$(a - a_o)$	$(b - b_o)$	= C_o	C_1	C_2
I	(1) A_{11} 3.322707	(2) A_{12} 15.798286	(3) A_{10} -0.000016	1	0
2		(4) A_{22} 81.828363	(5) A_{20} -0.000033	0	1
3	How Obtained		(6) $[w_i F_i^2]_o \equiv S(a_o, b_o)$ 3.125042	0	0
4	$-\frac{A_{12}}{A_{11}} \cdot$ (I)	(8) 75.115212	(9) +0.000076	(7) -4.754643	0
II	2 + 4	(10) B_{22} 6.713151	(11) B_{20} +0.000043	(12) -4.754643	1
6	$-\frac{A_{10}}{A_{11}} \cdot$ (I)		(14) 0.000000	(13) +0.000005	
7	$-\frac{B_{20}}{B_{22}} \cdot$ (II)		(16) 0.000000		(15) -0.000006
III	3 + 6 + 7		(17) S(a, b) 3.125042		
9	I solved for $(a - a_o)$	(19) $(a - a_o)$ -0.000036	(24) $(a - a_o)$ -0.000036	(21) R_{11} 3.668474	(23) R_{12} -0.708258
10	II solved for $(b - b_o)$	(18) $(b - b_o)$ +0.000006	(25) $(b - b_o)$ +0.000006	(20) R_{21} -0.708258	(22) R_{22} 0.148961
11	$M = A_{11} \cdot B_{22}$	(28) M 22.305834	(26) $a = a_o + (a - a_o)$ 0.893567	(30) s_a^2 23.651773	
12	$s_G^2 = \frac{S(a, b)}{n - 2}$	(29) s_G^2 3.125042	(27) $b = b_o + (b - b_o)$ 0.626854	(31) $r_{ab} s_a s_b$ -4.793551	(32) s_b^2 1.011737
13				(33) s_a^2 23.651774	
14				(34) $r_{ab} s_a s_b$ -4.793551	(35) s_b^2 1.011736

The page following Table 6.1 contains the tabulation form with the values of the 35 items entered for this same problem beginning with the final estimates a_o and b_o at the top of Column 6 in Table 6.1.

The tabulation form for fitting a straight line may also be used for the calculations required in fitting an arbitrary function involving any two unknown parameters provided A_{11}, A_{12}, A_{22}, A_{10} and A_{20}, are defined by (4.7) and (4.8).

6.2 Fitting an Arbitrary Function with 3 or More Unknown Parameters

In the case of 3 unknown parameters, the tabulation form on the following page may be used. The method of calculation of the various entries on this form should be clear from the discussion in the preceding subsection 6.1. The extension of the tabulation forms and methods of calculation to cover the case of more than 3 unknown parameters should be clear from the above-described tabulation forms for 2 and 3 unknown parameters.

The Iterative Solution of the Generalized Normal Equations for Determining u = 3 Parameters and Their Variances

a_{10} = a_{20} = a_{30} =

Row	$(a_1 - a_{10})$	$(a_2 - a_{20})$	$(a_3 - a_{30})$	= C_o	C_1	C_2	C_3
I	(1) A_{11}	(2) A_{12}	(4) A_{13}	(7) A_{10}	1	0	0
2		(3) A_{22}	(5) A_{23}	(8) A_{20}	0	1	0
3			(8) A_{33}	(9) A_{30}	0	0	1
4	How Obtained			(10) $[w_i F_i^2]_o$			
5	$-\frac{A_{12}}{A_{11}} \cdot$ (I)	(12)	(13)	(14)	(11)	0	0
II	2 + 5	(15) B_{22}	(16) B_{23}	(17) B_{20}	(18)	1	0
7	$-\frac{A_{13}}{A_{11}} \cdot$ (I)		(20)	(21)	(19)	0	0
8	$-\frac{B_{23}}{B_{22}} \cdot$ (II)		(23)	(24)	(25)	(22)	0
III	3 + 7 + 8		(26) C_{33}	(27) C_{30}	(28)	(29)	1
10	$-\frac{A_{10}}{A_{11}} \cdot$ (I)			(31)	(30)	0	0
11	$-\frac{B_{20}}{B_{22}} \cdot$ (II)			(33)		(32)	0
12	$-\frac{C_{30}}{C_{33}} \cdot$ (III)			(35)			(34)
IV	4 + 10 + 11 + 12			(36) $S(a_1, a_2, a_3)$			
14	I solved for $(a_1 - a_{10})$		(39) $(a_1 - a_{10})$	(49) $(a_1 - a_{10})$	(42) R_{11}	(45) R_{12}	(48) R_{13}
15	II solved for $(a_2 - a_{20})$		(38) $(a_2 - a_{20})$	(50) $(a_2 - a_{20})$	(41) R_{21}	(44) R_{21}	(47) R_{23}
16	III solved for $(a_3 - a_{30})$		(37) $(a_3 - a_{30})$	(51) $(a_3 - a_{30})$	(40) R_{31}	(43) R_{32}	(46) R_{33}
17	$M = A_{11} - B_{22} - C_{33}$		(55) M	(52) a_1	(57) s^2_{a1}		
18	$s^2_G = \frac{S(a_1, a_2, a_3)}{(n - 3)}$		(56) s^2_G	(53) a_2	(58) $r_{12} s_{a1} s_{a2}$	(60) s^2_{a2}	
19				(54) a_3	(59) $r_{13} s_{a1} s_{a3}$	(61) $r_{23} s_{a2} s_{a3}$	(62) s^2_{a3}
20					(63) s^2_{a1}		
21					(64) $r_{12} s_{a1} s_{a2}$	(66) s^2_{a2}	
22					(65) $r_{13} s_{a1} s_{a3}$	(67) $r_{23} s_{a2} s_{a3}$	(68) s^2_{a3}

7. THE DISTRIBUTION OF S

Throughout this section it will be assumed that each of the observed random variables are from normally distributed populations,[25] and it will be shown, on this assumption, that S/ν_1 (where $S \equiv [w_i F_i^2]$ is the minimized sum defined in (4.1)) is sometimes exactly but is always approximately† distributed as Fisher's variance ratio $F(\nu_1, \nu_2)$ with ν_1 and ν_2 degrees of freedom:

$$\nu_1 = n - u \tag{7.1}$$

$$\nu_2 = \frac{[s_i^2]^2}{[s_i^4/(m_i - 1)]} \qquad \text{(Variances not pooled)} \tag{7.2}$$

$$\nu_2 = [m_i - 1] \qquad \text{(Variances pooled)} \tag{7.3}$$

Here, as before, u denotes the number of unknown parameters estimated in minimizing S, and s_i^2 is defined by (4.2). This distribution may be used to detect statistically significant departures of the observed data from the statistical models assumed in deriving S. Thus if the observed value of (S/ν_1) is larger than $F(\nu_1, \nu_2, p)$ for the probability level p chosen in advance as the minimum value consistent with accepting the model, we may conclude (a) that the observed data contain a statistically significant component of variance arising from the presence of random systematic errors, (b) that the form of the function fitted to the data is incorrect, or (c) that a combination of both of the above factors is responsible for the observed large value of S.

---

† The proof given in subsection 7.3 for $k > 1$ depends on the the assumptions (a) that the ρ_{jhi} have the same values ρ_{jh} independent of i for all n groups of observations i = 1 to n and (b) that the ratios of the population variances $\sigma^2_{\eta ji}/\sigma^2_{\eta hi}$ have the same values C_{jh} independent of i for all n groups of observations i = 1 to n. These assumptions may not be necessary, but they are at least sufficient.

[25] Kenneth A. Norton and Eugene Barrows, "The Kolmogorov Test of the Goodness of Fit of Data Samples to Independently Specified Continuous Distributions, Together with a Test of the Normality of a Small Sample," NBS Report 5070, July, 1957.

When the same population variances cannot be assumed for all n groups, this approximate distribution of S provides the only satisfactory means presently available for testing our models. It should be noted that the statistical tests for heterogeneity of variance will often indicate a common population variance when, in fact, the actual population variances differ. Thus, in those cases where the experimenter has reason to suspect different population variances, it will be better to obtain the solution without pooling the variances even though the statistical tests for heterogeneity of variance indicate no statistically significant differences. This point is emphasized here since there may be a tendency to attempt to justify the simpler method of pooled variances by statistical tests alone, and this will lead occasionally to an incorrect acceptance or rejection of a proposed model.

7.1 One Variable Subject to Error (k = 1) and the Variances Pooled

We will consider initially in this subsection the case of one variable (k' = 1), and will assume that n groups of observations are used to estimate the population mean value α which is assumed to be the same for all groups. The least squares estimate for α may be expressed:

$$a = \frac{[w_i Y_i]}{[w_i]} \tag{7.4}$$

$$Y_i = \frac{1}{m_i} \sum_{t=1}^{m_i} Y_{it} \tag{7.5}$$

$$w_i = m_i / s_{\eta i}^2 \tag{7.6}$$

$$S = [w_i (Y_i - a)^2] = [w_i \{(Y_i - \alpha) - (a - \alpha)\}^2] \tag{7.7}$$

$$S = [w_i (Y_i - \alpha)^2 - 2w_i (Y_i - \alpha)(a - \alpha) + w_i (a - \alpha)^2]$$

$$S = [w_i (Y_i - \alpha)^2] - [w_i](a - \alpha)^2 \tag{7.8}$$

Assume now that the variance $\sigma^2_{\eta i}$ is known for each group of observations and let

$$w_i' = m_i/\sigma^2_{\eta i} \tag{7.9}$$

If we write $z_i \equiv \sqrt{w_i'}\,(Y_i - a)$, i = 1 to n, it follows that the z_i are independent random variables normally distributed about zero with unit variance; now, if w_i is replaced by w_i', (7.8) and (7.4) may be written

$$S_\sigma = [z_i^2] - [w_i'](a - \alpha)^2 = \left[\frac{m_i}{\sigma^2_{\eta i}}(Y_i - a)^2\right] \tag{7.10}$$

$$\sqrt{[w_i']}\,(a - \alpha) = \left[\frac{\sqrt{w_i'}\,z_i}{\sqrt{[w_i']}}\right] \tag{7.11}$$

The subscript σ on S is used to indicate that the weights w_i' are considered to be known constants. Note now that the linear form $\sqrt{[w_i']}\,(a - \alpha) \equiv [c_i z_i]$ has coefficients c_i which satisfy the condition $[c_i^2] = 1$. Thus we may apply Fisher's lemma† to (7.10), and conclude (a) that S_σ is distributed exactly as χ^2 with (n - 1) degrees of freedom; or alternatively that $S_\sigma/(n - 1)$ is distributed exactly as Fisher's variance ratio $F(n - 1, \infty)$, and (b) that $(n - 1)[w_i'](a - \alpha)^2/S_\sigma$ is distributed exactly as F(1, n - 1). There are no formal difficulties in extending these conclusions to problems involving u unknown parameters and more than one variable <u>provided</u> the additional variables are not random; for example, this has been done by Cramer in Chapter 37 of reference 18 and the principal change involved is the replacement of (n - 1) by (n - u). These conclusions are not very useful, however, since they depend on an assumed a priori knowledge of the variances $\sigma^2_{\eta i}$. However, if it is <u>reasonable on physical grounds</u> (independently of the observed data) to assume that the n groups have a common population variance σ^2_η, and particularly if, <u>in addition</u>, statistical tests for heterogeneity of the n observed variances indicate

---

† A good discussion of Fisher's lemma is given by Cramer in reference 18, p. 379. Another discussion, involving more elementary mathematics, is given in Section 10.6, p. 262, of the book by A. Hald.[26]

[26] A. Hald, "Statistical Theory with Engineering Applications," John Wiley and Sons, 1952.

that this assumption is not statistically unreasonable,† we may express S as follows:

$$S = \left[\frac{m_i}{s_\eta^2} (Y_i - a)^2 \right] = S_\sigma / (s_\eta^2 / \sigma_\eta^2) \qquad (7.12)$$

where $$[m_i - 1](s_\eta^2/\sigma_\eta^2) = [(m_i - 1)(s_{\eta i}^2/\sigma_\eta^2)] \qquad (7.13)$$

It is well known that each of the terms $(m_i - 1)\, s_{\eta i}^2/\sigma_\eta^2$ on the right hand side of (7.13) is distributed as χ^2 with $(m_i - 1)$ degrees of freedom, and it follows that their sum will be distributed as χ^2 with $[m_i - 1]$ degrees of freedom. Finally, since $S/(n - 1)$ may now be expressed as a ratio of two independent mean squares $(u/\nu_1)/(v/\nu_2)$ where $u \equiv S_\sigma$ is distributed as χ^2 with $\nu_1 = (n - 1)$ degrees of freedom and $v = [m_i - 1](s_\eta^2/\sigma_\eta^2)$ is distributed as χ^2 with $\nu_2 = [m_i - 1]$ degrees of freedom, it follows (a) that $S/(n - 1)$ is distributed exactly as $F(n - 1, [m_i - 1])$ and (b) that $(n - 1)[m_i](a - a)^2/S\, s_\eta^2$ is distributed exactly as $F(1, n - 1)$. These last two conclusions may be extended (See Cramer, chapter 37 in reference 18) to problems involving u parameters and k' variables, only one of which is random, simply by replacing a by a_p, (p = 1 to u) and (n - 1) by (n - u). These are the only problems for which the exact distribution of S is readily determinable.

7.2 One Variable Subject to Error (k = 1) and the Variances not Pooled

We will consider next the one-variate problem for the case where it is not reasonable to assume that the n groups of observations are from populations with the same variance, and will make use of a

---

† Note that statistical tests on a particular observed sample cannot provide a sufficient reason for assuming homogeneity of the variances, although repeated tests on many samples might be considered to provide adequate grounds for such an assumption; also, statistical tests for homogeneity of variances are unnecessary in case valid physical reasoning leads to the assumption of a common population variance.

theorem on quadratic forms derived in a recent paper by Box.[27] Let Q be the weighted sum of n different $\chi^2(\nu_i)$ variates characterized by ν_i degrees of freedom, i = 1 to n:

$$Q = \left[\frac{s^2_{\eta i}}{m_i} \right] = \left[\left(\frac{\sigma^2_{\eta i}}{m_i(m_i - 1)} \right) \left(\frac{(m_i - 1)\, s^2_{\eta i}}{\sigma^2_{\eta i}} \right) \right] = \left[\lambda_i \chi^2(\nu_i) \right] \quad (7.14)$$

$$\lambda_i = \sigma^2_{\eta i} / m_i(m_i - 1) \quad (7.15)$$

$$\chi^2(\nu_i) = (m_i - 1)\, s^2_{\eta i} / \sigma^2_{\eta i} \quad (7.16)$$

Note that the λ_i are simply constants and that $\nu_i = m_i - 1$.

According to Theorem 3.1 in Box's paper, Q/g will be distributed approximately as $\chi^2(h)$ where:

$$h = [\nu_i \lambda_i]^2 / [\nu_i \lambda_i^2] = [\sigma^2_{\eta i} / m_i]^2 / [\sigma^4_{\eta i} / m_i^2 (m_i - 1)] \quad (7.17)$$

$$gh = [\nu_i \lambda_i] = [\sigma^2_{\eta i} / m_i] \quad (7.18)$$

Now consider the ratio R:

$$R \equiv \{S_\sigma / (n - 1)\} / \{Q/gh\} \quad (7.19)$$

Since S_σ is distributed exactly as χ^2 with (n - 1) degrees of freedom, and Q/g is distributed approximately as χ^2 with h degrees of freedom and independently of S_σ, it follows that R is distributed

---

[27] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification," Annals of Mathematical Statistics, Vol. 25, June, 1954, pp. 290-302.

approximately as F(n - 1, h). If we now replace the $\sigma^2_{\eta i}$ in (7.17), (7.18), and (7.19) by the estimated values $s^2_{\eta i}$, we find that gh $\cong$ Q, $S \cong S_\sigma$, $R \cong S/(n-1)$,

$$\nu_2 = [s^2_{\eta i}/m_i]^2/[s^4_{\eta i}/m_i^2(m_i - 1)] \cong h \qquad (7.20)$$

and conclude that S/(n - 1) is distributed approximately as $F(n-1, \nu_2)$.† Furthermore, we may use the estimates w_i for w'_i in (7.10), and thus conclude that $(n-1)[w_\cdot](a - \alpha)^2/S$ is distributed approximately as F(1, n - 1) even when the population variances $\sigma^2_{\eta i}$ vary from group to group. These last two conclusions may be extended to problems involving u parameters and k' variables, only one of which is random, simply by replacing a by a_p and α by α_p (p = 1 to u), and (n - 1) by (n - u). Since $s^2_{\eta i}$ approaches $\sigma^2_{\eta i}$ as m_i approaches infinity, the above two approximate distributions for S and for $(a - \alpha)^2$ become exact as all of the m_i are allowed to increase without limit. It is of interest to compare the above approximate solution for the distribution of S with the approximate solution obtained by Welch 28/ for the special case n = 2. For this case (n - 1) = 1, and we may write:

$$S \equiv \frac{m_1(Y_1 - a)^2}{s^2_{\eta 1}} + \frac{m_2(Y_2 - a)^2}{s^2_{\eta 2}} = (Y_1 - Y_2)^2 \Big/ \left(\frac{s^2_{\eta 1}}{m_1} + \frac{s^2_{\eta 2}}{m_2}\right) \qquad (7.21)$$

The second expression on the right of (7.21) is readily obtained when we substitute in the middle member of (7.21) the following expression for a:

$$a \equiv \left(\frac{m_1 Y_1}{s^2_{\eta 1}} + \frac{m_2 Y_2}{s^2_{\eta 2}}\right) \Big/ \left(\frac{m_1}{s^2_{\eta 1}} + \frac{m_2}{s^2_{\eta 2}}\right) \qquad (7.22)$$

28/ B. L. Welch, "The Generalization of Student's Problem When Several Different Population Variances Are Involved," Biometrika, vol. 34, pp. 28-35, 1947.

† The reader should note that the replacement of $\sigma^2_{\eta i}$ by $s^2_{\eta i}$ increases the variance of the numerator and decreases the variance of the denominator in (7.19); the approximation depends upon the fact that these two effects are approximately compensatory.

Welch concludes,[29] when $\sigma^2_{\eta 1}$ may differ from $\sigma^2_{\eta 2}$, that S is distributed approximately as $F(1, \nu_2)$ with ν_2 defined exactly as in (7.20). A comparison of Aspin's[30] tabulated exact results for this special case indicates that Welch's approximation should have adequate accuracy for most applications. Thus we see that our approximate solution has satisfactory accuracy in this special case, and presume that it will also be satisfactory in the general case.

The numerator in (7.19) represents a between-groups estimate of variance, while the denominator in (7.19) represents a within-groups estimate of the same variance. Box[27] obtained in Section 7 of his paper an expression for the distribution of such a variance ratio without the restriction that the n group population variances must be the same, but his variance ratio was essentially different from (7.19); thus he did not weight either his Y_i or his $(Y_i - a)^2$ in inverse proportion to $\sigma^2_{\eta i}$. Although his solution was appropriate to the simple analysis of variance problem he was considering, our least squares formulation appears to be in a more useful form when no assumption is made about the population variances $\sigma^2_{\eta i}$.

7.3 More Than One Variable Subject to Error (k > 1)

We will show in this subsection that the extension of the above results to the general case in which more than one variable is subject to error involves only the use of (4.1) for defining S and the replacement of $s^2_{\eta i}/m_i$ by s^2_i as defined in (4.2); we obtain in this way a good approximation to the distribution of S. However, since the estimated values, a_p, are not linear functions of the errors of the k random variables, we can obtain only rough approximations to the distributions of $(a_p - \alpha_p)$ in this general case. In the particular case of fitting a straight line to observations on k = 2 random variables, it is shown in Section 12, however, that Wald's method of defining estimates

---

[29] B. L. Welch, "Further Note on Mrs. Aspin's Tables and on Certain Approximations to the Tabled Function," Biometrika, vol. 36, 1949, pp. 293-296.

[30] Alice A. Aspin, "Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated," Biometrika, vol. 36, pp. 290-296, 1949.

a' and b' of the two parameters α and β leads to exact distributions for $(a' - \alpha)$ and for $(b' - \beta)$, but Wald's method may be used only when the population variances $\sigma^2_{\eta i}$ and $\sigma^2_{\epsilon i}$ are independent of i.

For simplicity in presentation the following discussion will be limited to the case of fitting a straight line with both variables subject to error; the extension to the general case discussed in Section 4 is straightforward. We will consider the distribution of S in two artificial limiting cases (both involving implicitly the assumption that m_i is infinite), and we will in this way obtain an approximation to the distribution of S when m_i is large but finite.

We will consider first the somewhat artificial case in which $\sigma^2_{\eta i}$, $\sigma^2_{\epsilon i}$, and ρ_i are assumed to be known for each value of i = 1 to n, and only the parameters a_σ and b_σ and the adjusted values $X'_{i\sigma}$ and $Y'_{i\sigma}$ are estimated from the n groups of data. Such a situation might actually arise in practice if the experimenter had made many-- theoretically an infinite number-- of simultaneous observations of X_{it} and Y_{it} in order to establish the values of $\sigma^2_{\eta i}$, $\sigma^2_{\epsilon i}$, and ρ_i, and then wished later to fit a relatively small sample of data from the same populations to a straight line. We will use the subscript σ to distinguish this case from the solution described in Section 2 involving the values $s^2_{\eta i}$, $s^2_{\epsilon i}$, and r_i obtained from the sample being fitted. In the present case the minimized sum S_σ may be expressed:

$$S_\sigma = \left[\frac{1}{\sigma_i^2} (Y_i - a_\sigma - b_\sigma X_i)^2 \right] \tag{7.23}$$

$$\sigma_i^2 = (\sigma^2_{\eta i} - 2b_\sigma \rho_i \sigma_{\eta i} \sigma_{\epsilon i} + b_\sigma^2 \sigma^2_{\epsilon i})/m_i \tag{7.24}$$

We may also express S_σ in the following form:

$$S_\sigma = \left[\frac{m_i}{1 - \rho_i^2} \left\{ \frac{(Y_i - Y'_{i\sigma})^2}{\sigma^2_{\eta i}} - \frac{2\rho_i (Y_i - Y'_{i\sigma})(X_i - X'_{i\sigma})}{\sigma_{\eta i} \sigma_{\epsilon i}} + \frac{(X_i - X'_{i\sigma})^2}{\sigma^2_{\epsilon i}} \right\} \right] \tag{7.25}$$

The equivalence of (7.23) and (7.25) may be established by substituting the adjusted values $X'_{i\sigma}$ and $Y'_{i\sigma}$ appropriate to this case in (7.25). Hotelling [17] has established the invariance of the magnitude of quadratic forms like those in (7.25) to a rotation of the coordinate axes. Thus,

consider a rotation of the X, Y axes by an angle θ to new axes U, V, respectively, where positive θ corresponds to counterclockwise rotation, and

$$\tan 2\theta = \frac{2\rho_i \sigma_{\eta i} \sigma_{\epsilon i}}{(\sigma^2_{\epsilon i} - \sigma^2_{\eta i})} \tag{7.26}$$

If we assume that $\rho_i = \rho$ and $\sigma^2_{\epsilon i} = C^2 \sigma^2_{\eta i}$, then $\tan 2\theta = 2\rho C/(C^2 - 1)$; in this case, since such a rotation is independent of i, the errors $u_i = \epsilon_i \cos\theta + \eta_i \sin\theta$ and $v_i = -\epsilon_i \sin\theta + \eta_i \cos\theta$ in each group will be uncorrelated in the new coordinate system† and we may write:

$$S_\sigma = \left[\frac{m_i(V_i - V_i')^2}{\sigma^2_v} + \frac{m_i(U_i - U_i')^2}{\sigma^2_u} \right] \tag{7.27}$$

The above is now in exactly the same form as the minimized sum S studied by Deming, 7/ Thus Deming fitted a straight line on the assumption that σ^2_u and σ^2_v were known constants, and that $\rho_{uv} = 0$, and proved for this case that S_σ is distributed like χ^2 with (n - 2) degrees of freedom, 31/ i.e., that $S_\sigma/(n - 2)$ is distributed like $F(n - 2, \infty)$. In view of the above invariant transformation, it appears that Deming's results will also apply when ρ is different from zero provided $\rho_i = \rho$ and $\sigma^2_{\epsilon i} = C^2 \sigma^2_{\eta i}$.

---

† Note that $E(u_i v_i) = (\sigma^2_{\eta i} - \sigma^2_{\epsilon i}) \sin\theta \cos\theta + \rho \sigma_{\epsilon i} \sigma_{\eta i}(\cos^2\theta - \sin^2\theta)$ $= \sigma^2_{\eta i}\{(1 - C^2) \sin\theta \cos\theta + \rho C(\cos^2\theta - \sin^2\theta)\} = 0$ if $\tan 2\theta = 2\rho C/(C^2 - 1)$. Note that this is less restrictive than assuming that the population variances are the same for each group since $\sigma^2_{\eta i}$ may vary with i.

31/ See the discussion and accompanying references in reference 7 on pages 18, 23, 27, 141, and 230. For the case of fitting an arbitrary functional relation to random variables U and V with uncorrelated errors, see W. E. Deming, "On the Application of Least Squares-- III A New Property of Least Squares," Phil. Mag., Ser. 7, vol. XIX, p. 389, Supplement, February, 1935.

We will consider next the even more artificial case in which the population parameters X_{io}, Y_{io}, α, and β are known constants, but in which $s^2_{\eta i}$, $s^2_{\epsilon i}$, and r_i are estimated† from the samples. It will be convenient in this case to introduce the following notation:

$$F_{\beta i} = Y_i - \alpha - \beta X_i \tag{7.28}$$

$$s^2_{\beta i} = s^2_{\eta i} - 2\beta r_i s_{\eta i} s_{\epsilon i} + \beta^2 s^2_{\epsilon i} \tag{7.29}$$

$$\sigma^2_{\beta i} = \sigma^2_{\eta i} - 2\beta \rho_i \sigma_{\eta i} \sigma_{\epsilon i} + \beta^2 \sigma^2_{\epsilon i} \tag{7.30}$$

$$S_o = \left[\frac{m_i (Y_i - \alpha - \beta X_i)^2}{s^2_{\beta i}} \right] = \left[\frac{m_i \{(Y_i - Y_{io}) - \beta (X_i - X_{io})\}^2}{s^2_{\beta i}} \right] \tag{7.31}$$

$$S_{\sigma o} = \left[\frac{m_i \{(Y_i - Y_{io}) - \beta (X_i - X_{io})\}^2}{\sigma^2_{\beta i}} \right] \tag{7.32}$$

The expressions on the right of (7.31) and (7.32) may be obtained by subtracting $(Y_{io} - \alpha - \beta X_{io}) \equiv 0$ from $Y_i - \alpha - \beta X_i$. It is now obvious that $S_{\sigma o}$ is the sum of the squares of n variables each of which is normally and independently distributed about zero with unit variance; thus $S_{\sigma o}$ is distributed as χ^2 with n degrees of freedom. We may now apply Box's theorem to the above expressions in essentially the same way it was applied in the preceding subsection, and thus find that S_o/n is distributed approximately as $F(n, \nu_{2o})$ where:

$$\nu_{2o} = [s^2_{\beta i}/m_i]^2 / [s^4_{\beta i}/m_i^2 (m_i - 1)] \tag{7.33}$$

When all of the m_i are large, b approaches β, $s^2_{\beta i}/m_i$ approaches s^2_i and ν_{2o} approaches ν_2 as defined in (7.2). It should be noted that this last result is valid for completely arbitrary values of $\sigma^2_{\epsilon i}$, $\sigma^2_{\eta i}$, and ρ_i,

---

† These estimates are defined in (2.2) and (2.4).

and this suggests that the requirement that $\rho_i = \rho$ and $\sigma^2_{\epsilon i} = C^2 \sigma^2_{\eta i}$ imposed in deriving the distribution of S_σ may not be necessary when all of the m_i are large. In the special case where we may assume that $\sigma^2_{\beta i}$ has a constant value σ^2_β independent of i, we may set $S_o = S_{\sigma o}/(s^2_\beta/\sigma^2_\beta)$ and use the same arguments as were used in subsection 7.1 to show that S_o/n is distributed in this case exactly as $F(n, [m_i - 1])$.

If we attempt to apply the above distributions for S_σ and for S_o to the determination of the distribution of S in our more general formulation in which the values of $s^2_{\epsilon i}$, $s^2_{\eta i}$, r_i, X'_i, Y'_i, a, and b must all be estimated from the data, we find that the n separate terms in S are no longer independent as they were in S_o, and the weights are now random variables instead of being known constants as they were in S_σ. Nevertheless, since the estimated values $s^2_{\epsilon i}$, $s^2_{\eta i}$, r_i, X'_i, Y'_i, a and b all approach the constant values $\sigma^2_{\epsilon i}$, $\sigma^2_{\eta i}$, ρ_i, X_{io}, Y_{io}, α, and β, respectively, as m_i is allowed to increase without limit, it appears that S will be approximately equal to both S_o and S_σ when all of the m_i are large. Thus, since all of the above discussion may be extended without formal difficulties to the general case discussed in Section 4, we may expect S/ν_1 to be approximately distributed as $F(\nu_1, \nu_2)$ with ν_1 defined by (7.1), and ν_2 defined by either (7.2) or (7.3), and this approximation should be better the larger the values of all of the m_i. By using $\nu_1 = n - u$ in $F(\nu_1, \nu_2)$, approximate allowance has been made for the fact that the sum of the numerators of the n terms in S, normalized by their respective variances, has only (n - u) degrees of freedom since the estimated values of u parameters were determined in minimizing S; and by using ν_2 as defined in (7.2) or (7.3), approximate allowance has been made for the degrees of freedom in estimating the n variances $\sigma^2_{\beta i}$. Thus, to a first approximation, allowance has been made for the variances of all of the random variables entering the problem. Even for small values of m_i, changes in these random variables will affect our approximate distributions in the same directions as they affect the exact distributions [27] and thus our approximate theory will always provide a dependable, even if not exact, guide to the analyst. Even when the exact distributions become available in a usable form, it seems likely that the above-described approximations will continue to be useful because of their simplicity.

8. THE SIMPLEST LEAST SQUARES PROBLEM: ONE VARIABLE AND ONE PARAMETER

8.1 Without "Systematic" Errors

This is the simplest of all least squares problems, and will be discussed in some detail since it illustrates, under the simplest possible conditions, many of the characteristics and limitations of our least squares solution. We assume in this subsection that $E(Y_{it}) = Y_{io} = \alpha$, and thus, by assumption, exclude any "systematic" errors; also $E\{(Y_{it} - Y_{io})^2\} = \sigma^2_{\eta i}$.

$$F_i = Y_i - a \tag{8.1}$$

$$(1/w_i) = s_i^2 = s_{Yi}^2 = \frac{1}{m_i} s_{\eta i}^2 = \frac{1}{m_i(m_i - 1)} \sum_{t=1}^{m_i} (Y_{it} - Y_i)^2 \tag{8.2}$$

The estimate of variance (8.2) will be called the sample "within group" estimate.

$$(a - a_o) A_{11} = A_{10} \tag{8.3}$$

$$A_{11} = [w_i] \tag{8.4}$$

$$A_{10} = [w_i(Y_i - a_o)] \tag{8.5}$$

$$a = \frac{[w_i Y_i]}{[w_i]} = [m_i Y_i / s_{\eta i}^2]/[m_i / s_{\eta i}^2] \qquad \text{(Variances not pooled)} \tag{8.6}$$

We see by (8.6) that least squares leads in this case to the weighted mean with the weights equal to the reciprocal of the variances of the group means determined from the sample within group estimates of variance. Let us assume that it is reasonable on technical grounds to assume that the variances $s_{\eta i}^2$ for i = 1 to n may be considered to

be samples from the same statistical parent population. In this case we may pool these n estimates and obtain a value s_η^2 which does not depend on i; this common value cancels out in (8.6) which now becomes:

$$a_p = \frac{[m_i Y_i]}{[m_i]} \qquad \text{(Variances pooled)} \qquad (8.7)$$

We see exhibited here the two basic statistical properties of our system of weighting: (8.6) and (8.7) both show that the relative weights to be assigned are directly proportional to the number, m_i, of observations averaged in obtaining the mean values Y_i while (8.6) makes allowance in addition for variations in the observational conditions which might be present in determining these n different mean values. For example, with i = 1 and 2 suppose verniers were available on the measuring instruments and none for i = 3, 4 and 5; this could lead to values of $s_{\eta 1}^2$ and $s_{\eta 2}^2$ which are systematically smaller than the values $s_{\eta 3}^2$, $s_{\eta 4}^2$ and $s_{\eta 5}^2$, even with m_i the same for all 5 groups, and our least squares method has been formulated so as to give an appropriate additional amount of weight to Y_1 and Y_2 as compared to Y_3, Y_4 and Y_5 in this situation.

The following equations apply whether or not the variances are pooled:

$$S(a) = [w_i(Y_i - a)^2] \qquad (8.8)$$

$$R_{11} = \frac{1}{A_{11}} = \frac{1}{[w_i]} \qquad (8.9)$$

$$s_F^2 = s_a^2 = \frac{S}{n-1} R_{11} = \frac{[w_i(Y_i - a)^2]}{(n-1)[w_i]} \qquad (8.10)$$

Note that s_a is an estimate of the standard error of the weighted mean, a, obtained from the n independent groups of observations of Y. Our least squares solution can thus be expressed in the following form:

$$\alpha = a \pm s_a \qquad (8.11)$$

In this particular case, on the assumption that the population mean Y_{io} is equal to the "true" mean α, a is an unbiased estimate of the "true" mean value α, and s_a^2 is an unbiased estimate of the variance of a. We will see as we proceed, however, that the assumption $E(Y_{it}) = \alpha$ will not always be realized in practice because of the presence of systematic errors.

We will now introduce several numerical examples in order to illustrate more clearly the nature of our least squares solution of the one variate problem. In order to ensure that the data are from normal populations, we will construct our observations by using the table of random normal deviates in the Appendix to reference 7. This also has the advantage that we will know in these illustrative examples the population mean $Y_{io} = \alpha$ and population variances $\sigma_{\eta i}^2$; thus, for all of the examples in this section we take $Y_{io} = \alpha = 17$, and for the example in Table 8.1 we take $\sigma_{\eta i}^2 = 1$ for $i = 1$ to 5. Thus the 5 groups of observations in Table 8.1 might correspond to observations made by 5 different observers in 5 different laboratories.

It may be noted in passing that $s_{\eta i}^2$ should normally be calculated by means of the following exactly equivalent formula rather than directly from its definition (2.2):

$$s_{\eta i}^2 = \frac{1}{m_i(m_i - 1)} \left\{ m_i \sum_t Y_{it}^2 - \left(\sum_t Y_{it} \right)^2 \right\} \qquad (8.12)$$

On modern electrical calculators the two sums in (8.12) may be obtained in a single operation; when (8.12) is used it will be necessary to carry more significant figures than would be the case if (2.2) were used, but this is readily done on a modern electrical calculator.

From the data in Table 8.1 and equations (8.6), (8.8) and (8.10) we obtain: $a = 17.223$, $S = 2.939$, and $s_a = 0.124$; thus:

$$\alpha = 17.223 \pm 0.124 \qquad \text{(Variances not pooled)} \qquad (8.13)$$

Table 8.1

$Y_{io} = a = 17;\ \sigma^2_{\eta i} = \sigma^2_{\eta} = 1;\ m_i = m = 10;\ n = 5$

t	i = 1	i = 2	i = 3	i = 4	i = 5
1	18.95	18.87	17.63	14.08	18.72
2	18.57	18.41	18.17	18.53	16.92
3	15.81	16.63	18.25	16.49	18.29
4	15.53	16.75	16.76	15.98	16.04
5	17.35	16.75	16.69	16.22	17.91
6	15.93	17.57	16.66	16.02	15.48
7	17.54	18.27	17.53	17.17	17.70
8	16.54	16.41	15.62	18.31	18.21
9	17.81	16.98	17.19	17.30	17.40
10	18.16	16.62	18.27	15.58	17.35
Y_i	17.219	17.326	17.277	16.568	17.402
$s^2_{\eta i}$	1.45968	0.79054	0.74051	1.74686	1.03440
s^2_{Yi}	0.145968	0.079054	0.074051	0.174686	0.103440
w_i	6.85082	12.64958	13.50421	5.72456	9.66744
$Y_i - Y'_i$	-0.004	0.103	0.054	-0.655	0.179
G^2_i	0.00011	0.13420	0.03938	2.45598	0.30975
p'_i	0.99	0.73	0.85	0.15	0.59
$Y_i - Y_{io}$	0.219	0.326	0.277	-0.432	0.402
G^2_{io}	0.32857	1.34435	1.03616	1.06834	1.56230
p'_{io}	0.58	0.28	0.34	0.33	0.24
$Y_i - Y'_{ip}$	0.061	0.168	0.119	-0.590	0.244
G^2_{ip}	0.03223	0.24449	0.12267	3.01542	0.51573
p'_{ip}	0.86	0.63	0.74	0.12	0.49

Before accepting the above result, it is desirable to make tests to determine whether some of the group means contain systematic errors. All of the tests given in this section of the paper depend upon the assumption that the observations are from normally distributed populations; in the illustrative examples this is insured since they were constructed from a table of random normal deviates. For this problem $S = [G_i^2] = 2.93942$ and it was shown in subsection 7.2 that an approximate test of the hypothesis that all 5 group means are from populations with the same α but with possibly different variances $\sigma_{\eta i}^2$ may be obtained by setting $S/(n - 1) = 0.73485 \cong F(4, \nu_2, p)$ where $\nu_2 = (m - 1)[s_{\eta i}^2]^2/[s_{\eta i}^4] = 40.3803$. From the graphs in reference 19 we find $p \cong 0.58$. Note that p is approximately the probability of observing, in repeated random sampling with samples of this size from populations with arbitrary variances, a value of S larger than the value actually observed. Since $p = 0.58$ is much larger than the level 0.05 arbitrarily adopted † throughout this paper as the minimum permissible value of p for accepting the hypothesis, we conclude that there is no statistical evidence that these 5 groups of data are not from populations with the same mean α.

We may also examine the estimated errors $Y_i - Y_i'$ as given in Table 8.1; for the one variate problem $Y_i' = a$. The probabilities p_i' provide a more detailed, although less accurate, check for the presence of systematic errors. Thus, the probabilities p_i' may be obtained by setting $G_i^2 = F(1, 9, p_i')$ and, in the present illustrative problem, since $\alpha = 17$ is known, we may also determine probabilities p_{io}' by setting $G_{io}^2 = F(1, 9, p_{io}')$. Here G_{io}^2 is the value of T_i^2 obtained with Y^* replaced by α. The probability p_{io}' represents the probability of observing, in repeated sampling from the i^{th} population with samples of this size, a value of G_{io}^2 larger than the value actually observed. Since G_i^2 approaches G_{io}^2 as all of the m_i are allowed to increase without limit, we see that p_i' will also to this degree approximate p_{io}' and thus

---

† Note that the level chosen in practice for rejecting the hypothesis should be adopted in advance of making the test; the probability level actually used should depend on the risk involved, and should be chosen with due regard for the alternative actions to be taken if the hypothesis is either accepted or rejected.

the probabilities p_i' provide rough indications of the separate reliabilities of the n means Y_i. We see in Table 8.1 that none of the probabilities p_i' are less than 0.05, so we have no statistical reason to suspect any of these 5 means. Note, however, that the p_i' differ substantially from the corresponding probabilities p_{io}', and thus they should not be relied on except for a rough check.

In view of the above checks we conclude that (8.13) represents an acceptable solution to our problem.

It was shown in subsection 7.2 that $(n - 1)[w_i](a - \alpha)^2/S$ is distributed approximately as $F(1, n - 1)$. This result permits us to determine a confidence band for α:

$$Y^*\{100(1 - p)\%\} \cong a \pm \{SF(1, n - 1, p)/(n - 1)[w_i]\}^{1/2} = a \pm s_a\{F(1, n - 1), p)\}^{1/2} \tag{8.14}$$

The result (8.14) is to be interpreted as follows. If $n = 5$ samples of $m_i = 10$ individuals are observed repeatedly from these same 5 populations, which are assumed to have the same population mean α but possibly different variances, and if confidence bands as defined by (8.14) are constructed for each such sampling, then as the number of such samplings is increased without limit, it will be found that (approximately) a fraction $(1 - p)$ of the confidence bands so constructed will contain the population mean α. From reference 19 we find $F(1, 4, 0.5) = 0.54863$, $F(1, 4, 0.05) = 7.7086$, and $F(1, 4, 0.005) = 31.333$ so that $Y^*(50\%) \cong 17.131$ to 17.315; $Y^*(95\%) \cong 16.879$ to 17.567 and $Y^*(99.5\%) \cong 16.529$ to 17.917. Note that for this particular sample the 50% confidence band does not contain the population mean $\alpha = 17$, but that the 95% and 99.5% confidence bands do contain the population mean. All we can say is that the population mean will be found in approximately $100(1 - p)\%$ of the confidence bands constructed in this manner.

Suppose now that some physical theory indicates that the "true" value of $\alpha = 16$. Since we can write:

$$(n - 1)[w_i](a - \alpha)^2/S \cong F(1, n - 1, p) \tag{8.15}$$

and we know that $(n - 1)[w_i](a - 16)^2/S = 98 \cong F(1, n - 1, p)$, we find in reference 19 that $p \cong 0.0006$. Thus we may conclude, on the assumption that the theory is correct, either (a) that the observed sample contains "systematic" errors or (b) that the observed sample just happened to have a large deviation from the population mean, in fact in this case, so large that a still larger deviation would be expected to be observed in random sampling with a probability of only 0.0006. Alternatively, if the analyst had adopted 0.05 in advance as his hypothesis rejection level, these data would provide a basis for rejecting the theory.

Since the example in Table 8.1 was constructed by random sampling from 5 groups with the same normal distribution, i.e., $Y_{io} = 17$ and $\sigma^2_{\eta i} = 1$, we know in advance that the variances may be pooled. It will nevertheless be instructive to apply Bartlett's test [32] to these five observed variances $s^2_{\eta i}$ in Table 8.1. Bartlett has shown for n sample variances $s^2_{\eta i}$ from populations with the same population variance σ^2_{η} that $B/(n - 1)$ is distributed approximately as Fisher's variance ratio $F(n - 1, \infty)$, where:

$$B = \frac{2.30259}{C} \{[m_i - 1] \log_{10} s^2_{\eta} - [(m_i - 1) \log_{10} s^2_{\eta i}]\} \qquad (8.16)$$

$$C = 1 + \frac{\left[\frac{1}{m_i - 1}\right] - \frac{1}{[m_i - 1]}}{3(n - 1)} \qquad (8.17)$$

$$s^2_{\eta} = \frac{[(m_i - 1)\, s^2_{\eta i}]}{[m_i - 1]} \qquad (8.18)$$

The above test is useful even for small values of m_i, say 5 or more. If we apply this test to the variances in Table 8.1 we find $s^2_{\eta} = 1.15440$, $C = 1.04444$, $B/(n - 1) = 0.61073 \cong F(4, \infty, p)$; and from reference 19

---

[32] M. S. Bartlett, "Properties of Sufficiency and Statistical Tests," Proc. of the Royal Society of London, vol. 160A (1937), pp. 268-282.

we find p = 0.66. Since this is much larger than the level 0.05 arbitrarily adopted throughout this paper as the minimum permissible value of p for accepting the hypothesis, we conclude that there is no statistical evidence that these 5 sample variances are not from populations with the same variance. Note that B/(n - 1) is only approximately distributed as F(n - 1, ∞); however, Thompson and Merrington[33] have developed tables from which the exact 5% and 1% significance levels for B may be determined, and these will be useful when p is near the rejection level chosen. Reference may also be made to a paper by Hartley[34] in which a test for heterogeneity of variance is given which involves only the ratio of the largest to the smallest variances in the n groups; this test is more convenient but less powerful than Bartlett's test, and is recommended only when m_i has the same value m for all n groups.

In some problems it may happen that a test is required of the statistical significance simply of the departure of the largest of a set of variances. For example, it might be reasonable to pool the remaining variances if the largest of the set were eliminated. Cochran has developed such a test,and appropriate tables for the application of this test are given in reference 11.

Using the pooled variance $s_\eta^2 = 1.15440$, we obtain:

$$a_p = [Y_i]/n = 17.158 \tag{8.19}$$

$$S = m[(Y_i - a)^2]/s_\eta^2 = 3.93054 \tag{8.20}$$

$$s_a^2 = [(Y_i - a)^2]/n(n - 1) = 0.022687 \tag{8.21}$$

$$\alpha = 17.158 \pm 0.151 \qquad \text{(Variances pooled)} \tag{8.22}$$

---

[33] Catherine M. Thompson and Maxine Merrington, "Tables for testing the homogeneity of a set of estimated variances," Biometrika, Vol. 33 (1946), pp. 296-304. These tables are also available in reference 9, page 198.

[34] H. O. Hartley, "The maximum F-ratio as a short cut test for heterogeneity of variance," Biometrika, Vol. 37 (1950) pp. 308-312.

In this case the exact results in section 7.1 apply and we may set $S/(n-1) = 0.98263 = F(4, 45, p)$ and find $p = 0.425$; ; $Y^*(50\%) = 17.046$ to 17.270; $Y^*(95\%) = 16.740$ to 17.576 and $Y^*(99.5\%) = 16.315$ to 18.001; furthermore, the probability $p = 0.0016$ is obtained in this case using pooled variances for an assumed "true" value $a = 16$. The probabilities p'_{ip} corresponding to this case are also given in Table 8.1.

Thus we see that only slightly different results: (8.22) or (8.13) are obtained in this example, depending upon whether or not the experimenter assumes that the 5 population variances are the same or assumes that they may be different. In general it is better to assume that the population variances may be different unless there are good physical reasons, independent of the data and independent of statistical tests on it, for expecting that the population variances are the same. For example, it might be expected that the variances would be the same for n groups of measurements made under similar environmental conditions with apparatus of the same manufacture. However, even in this case, the n population variances might be different if n different observers were involved and it was established (possibly by statistical tests for heterogeneity of variance or otherwise) that a significant portion of the variance was contributed directly by the observers.

8.2 Example with Different Population Variances and without "Systematic" Errors

We will illustrate the results of the preceding subsection further by considering a numerical example constructed in such a way that the population variances are different. The numbers in Table 8.2 were obtained from the table of random normal deviates in the Appendix to reference 7 using the assumed population parameters: $Y_{io} = 17$, $\sigma^2_{\eta 1} = \sigma^2_{\eta 2} = 1$, and $\sigma^2_{\eta 3} = \sigma^2_{\eta 4} = \sigma^2_{\eta 5} = 6.25$. Let us assume for our discussion that the observations for the groups i = 1 and i = 2 might have been made using instruments with verniers, while the observations for the groups i = 3, 4 and 5 might have been made with instruments not having verniers. Thus the experimenter would have a good a priori physical reason, independent of an analysis of his data, to doubt whether the variances would be the same for all 5 groups, although he might expect the same variances in groups 1 and 2 and in groups 3, 4, and 5, respectively.

Table 8.2

$Y_{io} = \alpha = 17$; $\sigma^2_{\eta 1} = \sigma^2_{\eta 2} = 1$; $\sigma^2_{\eta 3} = \sigma^2_{\eta 4} = \sigma^2_{\eta 5} = 6.25$

t	i = 1	i = 2	i = 3	i = 4	i = 5
1	16.46	17.42	16.35	22.10	19.08
2	16.79	18.67	14.45	14.30	18.45
3	16.40	17.67	15.98	15.52	16.08
4	15.41	17.06	13.95	17.70	17.65
5	16.40	18.37	14.30	20.25	18.33
6	17.22	16.43	17.00	14.57	15.63
7	18.45	16.31	15.97	17.27	19.85
8	16.16	17.79	14.60	18.70	12.70
9	18.72	16.43	20.95	16.15	15.40
10	17.93	16.00	14.83	17.02	15.98
Y_i	16.994	17.215	15.838	17.358	16.915
$s^2_{\eta i}$	1.14036	0.84249	4.24308	6.09466	4.60936
$s^2_{Y_i}$	0.114036	0.084249	0.424308	0.609466	0.460936
w_i	8.76916	11.86958	2.35677	1.64078	2.16949
$Y_i - Y'_i$	-0.012	0.209	-1.168	0.352	-0.091
G^2_i	0.00129	0.51793	3.21576	0.20317	0.01801
p'_i	0.97	0.49	0.11	0.66	0.90
$Y_i - Y_{io}$	-0.006	0.215	-1.162	0.358	-0.085
G^2_{io}	0.00032	0.54867	3.18223	0.21029	0.01567
p'_{io}	0.99	0.48	0.11	0.66	0.90

If we apply Bartlett's test for homogeneity to the 5 variances in Table 8.2, we find $s^2_\eta = 3.386$, $B/(n-1) = 2.92 \cong F(4, \infty, p)$; and from reference 19 we find $p \cong 0.02$. Since this is less than the probability level 0.05 arbitrarily chosen throughout this paper for rejecting the hypothesis, we conclude (as was to be expected in view of the way the data in Table 8.2 were obtained) that it is statistically unlikely that these 5 variances are samples from populations with the same variance.

Inspection of the observed variances $s^2_{\eta i}$ in Table 8.2 indicates that the use of the vernier evidently reduced the variances for these two groups substantially. To test the statistical significance of this difference, we may pool the variances obtained with and without the vernier: s^2_η(with vernier) = 0.99143 and s^2_η(without vernier) = 4.98237. On the assumption that they are samples from normal populations with the same variance σ^2_η, the ratio: s^2_η(without vernier)/s^2_η(with vernier) would be distributed as $F(\nu_1, \nu_2)$ with $\nu_1 = \sum_{i=3}^{5} (m_i - 1) = 27$ (without vernier) and $\nu_2 = \sum_{i=1}^{2} (m_i - 1) = 18$ (with vernier). If we set $4.98237/0.99143 = 5.025 = F(27, 18, p)$ we find by reference 19 that the probability of observing a ratio as large or larger than this by chance, if the samples were actually from normal populations with the same variance, is $p = 0.0004$. Thus clear physical, and strongly supporting statistical, evidence is available as to the practical importance of the use of the vernier. This example also illustrates the importance of using all of the physical and statistical information available in drawing conclusions from the analysis; in this example our confidence in rejecting the hypothesis that the variances were equal increased from 0.98 (based on Bartlett's test) to 0.9996 when the prior knowledge was added as to the particular groups for which the verniers were used.

Consider now the following three estimates of a: (a) with no pooling of variances $a = 17.006 \pm 0.192$; (b) with partially pooled variances $a = 17.012 \pm 0.181$; and (c) using only the data from groups 1 and 2 for which the vernier was used $a = 17.105 \pm 0.110$. Since the estimate (b) above makes use of all of the physical and statistical evidence available to the analyst, it is to be preferred over the

estimate (c) even though the latter has a smaller estimated standard error. On the basis of the above description of the experiment there is no way of choosing between the estimates (a) and (b); however, the estimate (a) is more conservative and should certainly be adopted if the experimenter has reasons (different observers, for example) to doubt that groups 1 and 2 and groups 3, 4, and 5, respectively, have the same population variances.

8.3 Example with "Systematic" Bias and Random "Systematic" Errors

We assume initially in this subsection that each group of observations (i = 1 to n) is subject to a systematic bias v_o and a random "systematic" error v_i, i.e., $Y_{it} = Y_{io} + \eta_{it} = \alpha + v_o + v_i + \eta_{it}$; as m_i approaches infinity $E(Y_{it}) = Y_{io} + v_i = \alpha + v_o + v_i$ and, as n approaches infinity $E(Y_{io}) = Y_{io} = \alpha + v_o$. Thus v_o is an assumed systematic bias which has the same value for all n groups and for each observation within each group. A constant systematic bias of this kind cannot be detected by least squares. In fact it can be shown more generally that such systematic biases occurring in one or more of the observed variables in the multivariate model of Section 4 cannot be detected by least squares. The proof of this general statement follows from the fact that our least squares solution is invariant to a translation of the coordinate axes and a systematic bias in an observed variable is equivalent to a translation of the corresponding coordinate axis by the amount of this systematic bias. Since such systematic bias cannot be detected by least squares, the analyst should remember that his solution of the above one variate problem can yield only an estimate of the population mean $Y_{io} = \alpha + v_o$ or, more generally, that the population means X_{jio} in the multivariate problem which he can estimate by a least squares analysis may actually be the sum of a "true" value plus an unknown systematic bias v_{jo} present in all of the observations of the j^{th} variable. Thus we conclude that other means than least squares must be used to detect such systematic biases. Throughout the remainder of this paper it will be convenient to eliminate explicit allowance for such systematic biases, and the student should remember that his least squares analysis leads only to population mean values and not necessarily to the "true" values.

Our one variate model with random systematic errors may now be described as follows: $Y_{it} = Y_{io} + v_{it} = Y_{io} + v_i + \eta_{it}$ approaches infinity $E(Y_{it}) = Y_{io} + v_i$, $E(\eta_{it}^2) = \sigma_{\eta i}^2$, $E(\eta_{it}\,\eta_{jt}) = 0$ for $i \neq j$ and $E(\eta_{it}\,\eta_{iu}) = 0$ for $t \neq u$; and as n approaches infinity $E(v_i^2) = \sigma_v^2$, $E(v_i\,v_j) = 0$ for $i \neq j$. The values in Table 8.3 were obtained from a table of random normal deviates using the population parameters $\alpha = 17$, $\sigma_{\eta i}^2 = \sigma_\eta^2 = 1$, $m_i = m = 10$ and $\sigma_v^2 = 9$. We will assume that the variances may be pooled and the solution in this case is:

$$\alpha = 17.238 \pm 1.026. \tag{8.23}$$

For this example $S = [G_i^2] = 467.62$, and we may test the hypothesis $\sigma_v^2 = 0$ by setting $S/(n - 1) = 116.91 = F(4, 45, p)$. In this case we find that $p < 0.0001$; this is the probability of observing a value of S as large or larger than the value actually observed in random sampling from populations with the same population means and variances and with $\sigma_v^2 = 0$. Since p is less than the level 0.05 chosen for rejecting the hypothesis, we conclude that $\sigma_v^2 \neq 0$.

It is of interest now to examine the estimated errors $Y_i - Y_i'$ as given in Table 8.3; note that one of these has a large negative value: -3.9262. If the experimenter has some actual physical reason for believing that the third group of measurements might be biased in this particular direction, he might be led to reject this group entirely. However, the statistical analysis can give him still further assistance in arriving at a correct solution. With the third group eliminated, $n = 4$ and we obtain $\alpha = 18.130 \pm 0.454$; $S = 58.176$, and if we set $S/(n - 1) = 19.392 = F(3, 36, p)$ we find $p < 0.0001$; thus we still have statistical evidence of random systematic errors in the remaining 4 group means and conclude that the elimination of the third group of measurements did not improve matters appreciably.

Thus we see that the within group variances $s_{\eta i}^2$, although consistent among themselves, do not measure all of the variance of the data. Other random errors v_i evidently also occur from one group to the next, and the experimenter will wish to understand these; for example, if the n groups of measurements were made on n different days, an explanation would naturally be sought in terms of possibly

Table 8.3

$Y_{io} = 17$; $\sigma^2_{\eta i} = 1$ (i = 1 to 5); $\sigma^2_v = 9$

t	i = 1	i = 2	i = 3	i = 4	i = 5
1	17.21	17.43	13.48	19.29	19.29
2	16.56	16.66	12.04	20.16	16.67
3	18.19	18.01	13.79	19.70	18.01
4	17.50	18.43	12.48	18.90	17.81
5	18.87	17.64	13.45	19.61	17.20
6	17.96	16.68	13.19	18.76	17.40
7	17.70	16.65	12.52	20.05	18.01
8	18.84	17.83	13.81	19.45	18.47
9	16.83	17.66	14.24	19.45	17.62
10	17.43	17.84	14.12	19.34	18.08
Y_i	17.709	17.483	13.312	19.471	17.856
$Y_{io} + v_i$	17.69	17.27	13.31	19.67	17.84
$s^2_{\eta i}$	0.59832	0.38858	0.55447	0.19601	0.51485
s^2_{η}	0.45044	0.45044	0.45044	0.45044	0.45044
w_i	22.20031	22.20031	22.20031	22.20031	22.20031
$Y_i - Y'_i$	0.4708	0.2448	-3.9262	2.2328	0.6178
G^2_i	4.92076	1.33040	342.21881	110.67733	8.47334
p'_i	0.053	0.28	<0.0001	<0.0001	0.017
$Y_i - \alpha$	0.709	0.483	-3.688	2.471	0.856
G^2_{io}	11.1597	5.1791	301.9540	135.5516	16.2670
p'_{io}	0.0089	0.050	<0.0001	<0.0001	0.0027
or w_i	23.5606	23.5606	0	23.5606	23.5606
$Y_i - Y'_i$	-0.421	-0.647		1.341	-0.274
G^2_i	4.1759	9.8627		42.3686	1.7688
p'_i	0.073	0.013		<0.0001	0.22

different experimental conditions on these n days. In particular, the analyst will wish to obtain an estimate $\hat{\sigma}_v^2$ of the variances of these random systematic errors. The following derivation of an estimate of $\hat{\sigma}_v^2$ will be applicable to the general case in which both m_i and $\sigma_{\eta i}^2$ may differ from group to group. Consider the minimized sum S_σ for our present model which now involves random "systematic" errors v_i:

$$S_\sigma = [w_i'(Y_i - \overline{Y})^2] \tag{8.24}$$

$$w_i' = m_i/\sigma_{\eta i}^2 \tag{8.25}$$

$$Y_i - \overline{Y} = v_i + \overline{\eta}_i - \frac{\left[\frac{m_i(v_i + \overline{\eta}_i)}{\sigma_{\eta i}^2}\right]}{\left[\frac{m_i}{\sigma_{\eta i}^2}\right]} \tag{8.26}$$

$$\overline{\eta}_i = \frac{1}{m_i}\sum_{t=1}^{m_i} \eta_{it} \tag{8.27}$$

Using these relations it may be shown that the expected value of S_σ may be expressed:

$$E(S_\sigma) = (n - 1) + \sigma_v^2 \left\{ \frac{[w_i']^2 - [w_i'^2]}{[w_i']} \right\} \tag{8.28}$$

If we replace $\sigma_{\eta i}^2$ by $s_{\eta i}^2$ on both sides of (8.28) we obtain the following estimate for $\hat{\sigma}_v^2$:

$$\hat{\sigma}_v^2 = \left\{ \frac{S}{(n - 1)} - 1 \right\} \left\{ \frac{(n - 1)[w_i]}{[w_i]^2 - [w_i^2]} \right\} \tag{8.29}$$

Note that this reduces, when the variances $\sigma^2_{\eta i}$ are the same so that the $s^2_{\eta i}$ may be pooled, to the following unbiased estimate:

$$\hat{\sigma}^2_v = \frac{s^2_\eta}{\overline{m}} \left\{ \frac{S}{(n-1)} - 1 \right\} \qquad (\sigma^2_{\eta i} = \sigma^2_\eta \text{; variances pooled}) \qquad (8.30)$$ †

where $\overline{m} = \dfrac{[m_i]^2 - [m_i^2]}{(n-1)[m_i]}$.

Using the result obtained in Section 7 that $S/(n-1)$ is distributed approximately as $F(\nu_1, \nu_2)$ with ν_1 and ν_2 defined by (7.1) and (7.2), respectively, we may use the method of Bross [35] to obtain the following approximate fiducial distribution for σ^2_v:

$$\sigma^2_v(1-p) = \hat{\sigma}^2_v \left\{ \frac{S}{n-1} - F(\nu_1, \nu_2, p) \right\} \Big/ \left\{ \frac{SF(\nu_1, \infty, p)}{n-1} - F(\nu_1, \nu_2, p) \right\} \qquad (8.31)$$

In the above $\sigma^2_v(1-p)$ denotes the value which the population value σ^2_v will exceed in repeated random sampling from the same normal population with an approximate fiducial probability ‡ of $(1-p)$. For values of p so small that $S < (n-1)F(\nu_1, \nu_2, p)$, $\sigma^2_v(1-p) = 0$.

---

[35] I. Bross, "Fiducial intervals for variance components," Biometrics, Vol. 6, page 136, 1950.

† This is the same as the unbiased estimate of σ^2_v determined by analysis of variance; see page 328 in reference 9 or page 312 in reference 11.

‡ The subtle distinction between confidence and fiducial intervals is well described by M. G. Kendall in the book "The Advanced Theory of Statistics," Vol. II, Chapter 20, Charles Griffin and Company, London, 1946. For most applications the distinction between confidence and fiducial intervals is of little importance.

To illustrate the use of the above formula, we will determine for the example in Table 8.3 the 90% fiducial interval for σ_v^2: $\sigma_v^2(0.95) < \sigma_v^2 < \sigma_v^2(0.05)$ with probability 0.9:

$$\sigma_v^2(0.95) = 5.22086\{116.91 - 2.5790\}/\{116.91 \times 2.3719 - 2.5790\} = 2.173$$

$$\sigma_v^2(0.05) = 5.22086\{116.91 - 0.17519\}/\{116.91 \times 17768 - 0.17519\} = 29.59$$

For this particular problem the population value $\sigma_v^2 = 9$ happens to lie within this 90% fiducial interval; however, the population value would be expected to lie within only approximately 90% of a large number of such intervals so constructed from random samples.

The derivation of the above-described approximate fiducial distribution for σ_v^2 depended on the assumption that both the η_{it} and the v_i are samples from normal populations with mean zero. On this assumption we may also obtain confidence intervals for a from the data of Table 8.3 in exactly the same way as was done in subsections 8.1 and 8.2; but now, because of the additional variance between groups, these intervals are naturally much larger: $Y^*(50\%) = 16.478$ to 17.998; $Y^*(95\%) = 14.389$ to 20.087 and $Y^*(99.5\%) = 11.495$ to 22.981.

Since we have only a confidence $p < 0.0001$ that the samples in Table 8.3 are free from random "systematic" errors, we may conclude that such errors are, in fact, present and obtain a solution to our problem on this assumption. The only modification to the analysis which is required is the replacement of the weights $w_i = m/s_\eta^2$ by $w_i'' = 1/(\hat{\sigma}_v^2 + s_Y^2)$ with $s_Y^2 = s_\eta^2/m$. In the present case, since the 5 weights were the same in determining the solution (8.23), the revised solution will still be (8.23) since the w_i'' still do not vary from group to group. However, S" will now be smaller by the factor $s_\eta^2/(s_\eta^2 + m\hat{\sigma}_v^2) = 0.0085536$ and $S''/(n-1) = 1$. Note that S" is not a random variable; consequently, S" cannot be used for testing the physical hypothesis $a = 16$. However, this hypothesis can still be tested by setting $(a - 16)^2/s_a^2 = 1.456 = F(1, n-1, p)$ and we find $p = 0.30$. This is the probability of observing a sample departing from $a = 16$ more than the sample in Table 8.3 by random sampling from populations

with both η_{it} and v_i normally distributed. Since this value of p is larger than 0.05, we conclude that the sample in Table 8.3 does not necessarily provide evidence for rejecting the theory. However, in this case, since we have statistical evidence for the presence of random "systematic" errors, these errors (estimated by $Y_i - 16$) should certainly be thoroughly investigated before the theory is accepted.

8.4 Example with Random "Systematic" Errors and Unequal Population Variances

The model and the analysis in this case is essentially the same as in the preceding subsection, although differing slightly in detail.

Table 8.4 is representative of data which might be obtained in practice with the population parameters: $a = 17$, $\sigma^2_{\eta 1} = \sigma^2_{\eta 2} = 1$, $m_1 = m_2 = 10$; $\sigma^2_{\eta 3} = \sigma^2_{\eta 4} = \sigma^2_{\eta 5} = 4$, $m_3 = m_4 = m_5 = 5$; and $\sigma^2_v = 9$. We find $a = 18.501$, $S = 22.3875$ and $\hat{\sigma}^2_v = 0.83881$. In this case the revised weights will vary from group to group:

$$w''_i = 1/(\hat{\sigma}^2_v + s^2_{Yi}) \tag{8.32}$$

Using these revised weights we find $a'' = [w''_i Y_i]/[w''_i] = 19.097$, and $S = [w_i(Y_i - a'')^2] = 34.4875$.† Using this revised estimate of S in (8.29) we obtain $\hat{\sigma}^2_v = 1.3908$ as a revised estimate of σ^2_v. This revised value may now be used to obtain a second revision of w''_i; this second revision is given in Table 8.4 and leads to $a'' = 19.166$ and $S = 37.4528$ so that finally:

---

† Note that the estimate a" involves the weights w''_i since we wish to average out both within-group and between-group errors; on the other hand, S is a ratio of between-group and within-group variances and is therefore defined in terms of the original w_i.

Table 8.4

$Y_{io} = 17;\ \sigma^2_{\eta 1} = \sigma^2_{\eta 2} = 1;\ \sigma^2_{\eta 3} = \sigma^2_{\eta 4} = \sigma^2_{\eta 5} = 4;\ \sigma^2_v = 9$

t	i = 1	i = 2	i = 3	i = 4	i = 5
1	18.23	18.14	21.25	22.19	20.23
2	18.30	17.01	20.03	17.67	17.25
3	17.59	18.41	19.23	22.51	19.39
4	18.53	17.22	17.99	21.05	18.89
5	17.97	16.89	20.79	20.95	22.09
6	17.64	19.41			
7	19.38	18.51			
8	18.38	16.59			
9	19.55	17.95			
10	19.88	17.18			
Y_i	18.545	17.731	19.858	20.874	19.570
$Y_{io} + v_i$	17.90	17.78	20.03	20.57	18.65
$s^2_{\eta i}$	0.63754	0.79954	1.67992	3.67768	3.16580
s^2_{Yi}	0.063754	0.079954	0.33598	0.735536	0.63316
w_i	15.685	12.507	2.9763	1.3596	1.5794
$Y_i - a$	+0.044	-0.770	1.357	2.373	1.069
G^2_i	0.0304	7.4154	5.4807	7.6561	1.8049
p'_i	0.86	0.025	0.080	0.051	0.25
$Y_i - \alpha$	1.545	0.731	2.858	3.874	2.570
G^2_{io}	37.440	6.6833	24.311	20.399	10.432
p'_{io}	<0.0001	0.031	0.0079	0.011	0.032
w''_i	0.68750	0.67992	0.57911	0.47029	0.49408
$Y_i - a''$	-0.621	-1.435	-0.692	1.708	0.404
G^2_i	6.0488	25.7547	1.4252	3.9663	0.2578
p'_i	0.037	0.0047	0.30	0.12	0.64

$$\alpha = 19.17 \pm 0.52 \qquad (8.33)$$

Using this second revision of S leads to $\hat{\sigma}_v^2 = 1.53$. In this case $S/(n-1)$ is distributed approximately as $F(\nu_1, \nu_2)$ with $\nu_1 = (n-1) = 4$ as before, but with $\nu_2 = [s_{Yi}^2]^2/[s_{Yi}^4/(m_i - 1)] = 6.9787$, and we obtain by (8.31) the following 90% fiducial interval for σ_v^2:

$$0.44 < \sigma_v^2 < 9.4 \qquad (8.34)$$

Again the population value $\sigma_v^2 = 9$ happens to lie within the 90% fiducial interval, but in 10% of the cases of random sampling it would not be expected to do so.

We may also obtain the following approximate confidence intervals for α: $Y^*(50\%) = 18.78$ to 19.56; $Y^*(95\%) = 17.73$ to 20.61 and $Y^*(99.5\%) = 16.26$ to 22.08.

It is of interest to note that the estimate $a = 18.501$ obtained with the weights w_i happens in this case to be nearer the population value $\alpha = 17$ than the more properly weighted estimate $a'' = 19.166$. This will occur occasionally because of sampling fluctuations; nevertheless, the use of the procedure described above is recommended since it will yield better results on the average for a large number of samples.

This completes the discussion of the one variable problem. All of the above methods of analysis have their counterparts in least squares analyses involving several random variables and, in some respects, the methods of handling these more general problems are the same as for one variable. Thus the random systematic errors in the multivariate problem are estimated in much the same way; unfortunately, however, the allocation of the several components of these random systematic errors to the corresponding random variables by this method is necessarily somewhat arbitrary, and this is one of the principal difficulties of extending the method of least squares as formulated in Sections 2, 3 and 4 to include the effects of random systematic errors. In the particular case of fitting a straight line to two random variables, Wald's method, as generalized in Section 11, eliminates some of this arbitrariness of including the effects of random systematic errors, but no method is presently available which leads to a completely unambiguous solution in all cases.

CONCLUSION

It is unfortunate that the unique solution to the problem of fitting a series of points to a specified functional relation is so complex in the general case where more than one of the observed variates determining these points is subject to error. However, no short cuts have been found to a correct statistical understanding of experimental data obtained under these rather typical conditions. It is hoped that the methods presented herein are in sufficiently usable form that they will be employed by experimenters wishing to obtain consistent conclusions from their analyses of experimental data.

In those cases where the experiments are still in the planning stage, use may often be made of the methods presented in this paper to design the experiments in such a way that repeated observations of the coordinates of each point become available; in this way more nearly optimum use may be made of statistical theory in the analysis of the resulting experimental data.

ACKNOWLEDGEMENTS

First I would like to mention the extensive assistance of Dr. E. L. Crow of the Boulder Laboratories, whose careful reading and constructive criticism led to several revisions of this paper. Since I did not always take his advice, he should not be held responsible for any remaining errors or obscurities in the presentation.

In addition to his many ideas which I have acknowledged throughout the text, I wish to acknowledge also the early training in least squares which I received from Dr. W. E. Deming in his course at the Department of Agriculture.

The assistance I received from Dr. C. E. Eisenhart came principally from my study of his paper "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," Annals of Mathematical Statistics, vol. X, pp. 162-186, June, 1939.

Mention should also be made of several helpful suggestions, particularly regarding the method of presentation, made by Dr. Harold Staras, Dr. George Hufford and Dr. George Birnbaum.

APPENDIX I

Maximum Likelihood Estimation of the 3n + 2 Parameters X_{io}, Y_{io}, $\sigma^2_{\eta i}$, ρ, and C^2 for k = 2 Normally Distributed Variables

A statistical model often encountered in least squares fitting when only two of the variables are random, involves n bivariate distributions with 2n different means and 2n different variances, but with the same correlation coefficient ρ and the same ratio C^2 between the 2n variances: $\sigma^2_{\epsilon i} = C^2 \sigma^2_{\eta i}$. If the two random variables are both normally distributed, we may use the method of maximum likelihood to obtain consistent and asymptotically efficient estimates of the 3n + 2 parameters defining these n bivariate distributions.

$$L = [\ell_i] \tag{I-1}$$

$$\ell_i = \sum_{t=1}^{m_i} \ln \left\{ \frac{1}{2\pi C \sigma^2_{\eta i} \sqrt{1 - \rho^2}} \exp \left(- \frac{f_{it}}{2\sigma^2_{\eta i}(1 - \rho^2)} \right) \right\} \tag{I-2}$$

$$f_{it} = \left\{ \frac{(X_{it} - X_{io})^2}{C^2} - \frac{2\rho(X_{it} - X_{io})(Y_{it} - Y_{io})}{C} + (Y_{it} - Y_{io})^2 \right\} \tag{I-3}$$

The maximum likelihood estimates $\hat{X}_{io}$, $\hat{Y}_{io}$, of the 2n mean values are determined by solving simultaneously the following 2n equations:

$$\frac{\partial L}{\partial X_{io}} = \frac{\partial \ell_i}{\partial X_{io}} = \frac{1}{(1 - \rho^2)\sigma^2_{\eta i}} \sum_{t=1}^{m_i} \left\{ \frac{(X_{it} - X_{io})}{C^2} - \frac{\rho(Y_{it} - Y_{io})}{C} \right\} = 0 \tag{I-4}$$

$$\frac{\partial L}{\partial Y_{io}} = \frac{\partial \ell_i}{\partial Y_{io}} = \frac{1}{(1 - \rho^2)\sigma^2_{\eta i}} \sum_{t=1}^{m_i} \left\{ (Y_{it} - Y_{io}) - \frac{\rho(X_{it} - X_{io})}{C} \right\} = 0 \tag{I-5}$$

$$\hat{X}_{io} = \frac{1}{m_i} \sum_{t=1}^{m_i} X_{it} \equiv X_i \qquad \text{(I-6)}$$

$$\hat{Y}_{io} = \frac{1}{m_i} \sum_{t=1}^{m_i} Y_{it} \equiv Y_i \qquad \text{(I-7)}$$

The maximum likelihood estimates $\hat{\sigma}^2_{\eta i}$, $\hat{\rho}$, and $\hat{C}$ of the remaining $n + 2$ parameters are determined by solving simultaneously the following $n + 2$ equations:

$$\frac{\partial L}{\partial \sigma^2_{\eta i}} = \frac{\partial \ell_i}{\partial \sigma^2_{\eta i}} = \sum_{t=1}^{m_i} \left\{ -\frac{1}{\hat{\sigma}^2_{\eta i}} + \frac{\hat{f}_{it}}{2(1 - \hat{\rho}^2)\hat{\sigma}^4_{\eta i}} \right\} = 0 \qquad \text{(I-8)}$$

$$\frac{\partial L}{\partial \rho} = \frac{1}{(1 - \hat{\rho}^2)} \left[\sum_{t=1}^{m_i} \left\{ \hat{\rho} - \frac{\hat{\rho}\hat{f}_{it}}{(1 - \hat{\rho}^2)\hat{\sigma}^2_{\eta i}} + \frac{(X_{it} - X_i)(Y_{it} - Y_i)}{\hat{C}\hat{\sigma}^2_{\eta i}} \right\} \right] = 0 \qquad \text{(I-9)}$$

$$\frac{\partial L}{\partial C} = \left[\sum_{t=1}^{m_i} \left\{ -\frac{1}{\hat{C}} + \frac{(X_{it} - X_i)^2}{(1 - \hat{\rho}^2)\hat{\sigma}^2_{\eta i}\hat{C}^3} - \frac{\hat{\rho}(X_{it} - X_i)(Y_{it} - Y_i)}{(1 - \hat{\rho}^2)\hat{\sigma}^2_{\eta i}\hat{C}^2} \right\} \right] = 0 \qquad \text{(I-10)}$$

In the above equations $\hat{f}_{it}$ denotes the value of f_{it} with the parameters replaced by their maximum likelihood estimates. Using the sample values defined by (2.2) and (2.4), the above $n + 2$ equations may be expressed:

$$\hat{\sigma}^2_{\eta i} = \frac{(m_i - 1)}{2m_i(1 - \hat{\rho}^2)} \left\{ \frac{s^2_{\epsilon i}}{\hat{C}^2} - \frac{2\hat{\rho} r_i s_{\epsilon i} s_{\eta i}}{\hat{C}} + s^2_{\eta i} \right\} \qquad \text{(I-11)}$$

$$\hat{\rho} = \frac{1}{[m_i]\hat{C}} \left[\frac{(m_i - 1) r_i s_{\epsilon i} s_{\eta i}}{\hat{\sigma}^2_{\eta i}} \right] \qquad \text{(I-12)}$$

$$\hat{C}^2 = \frac{1}{[m_i]} \left[\frac{(m_i - 1) s_{\epsilon i}^2}{\hat{\sigma}_{\eta i}^2} \right] \qquad \text{(I-13)}$$

It does not appear to be useful to further separate the variables in the above equations since they may be solved directly in their present form by numerical methods involving an iterative process. For example, estimates of $\hat{\rho}$ and $\hat{C}$ may be obtained by setting $\hat{\sigma}_{\eta i}^2 \cong s_{\eta i}^2 (m_i - 1)/m_i$ so that:

$$\hat{C}_{est}^2 \cong \frac{1}{[m_i]} \left[m_i s_{\epsilon i}^2 / s_{\eta i}^2 \right] \qquad \text{(I-14)}$$

$$\hat{\rho}_{est} \cong \frac{1}{[m_i] \hat{C}_{est}} \left[\frac{m_i r_i s_{\epsilon i}}{s_{\eta i}} \right] \qquad \text{(I-15)}$$

Using these estimated values in (I-11), revised estimates of $\hat{\sigma}_{\eta i}^2$ may be obtained; these may then be substituted in (I-12) and (I-13) to obtain better estimates of $\hat{\rho}$ and $\hat{C}$. The above process could be repeated, if better accuracy were required, but this would presumably not often be the case in practice since the values $\hat{\sigma}_{\eta i}^2$, $\hat{\rho}$ and $\hat{C}$ are themselves only estimates of the population values of these parameters, and these estimated values are required in our least squares solution only in the determination of the relative weights.

Fortunately, a case often encountered in our least squares application is the one in which the magnitudes of ρ and C are known a priori and thus need not be estimated by the method of maximum likelihood from the data. In this special case the n maximum likelihood estimates of $\hat{\sigma}_{\eta i}^2$ may be expressed:

$$\hat{\sigma}_{\eta i}^2 = \frac{(m_i - 1)}{2m_i(1 - \rho^2)} \left\{ \frac{s_{\epsilon i}^2}{C^2} - \frac{2\rho r_i s_{\epsilon i} s_{\eta i}}{C} + s_{\eta i}^2 \right\} \qquad \text{(I-16)}$$

If we determine the expected value of (I-16) we find that $E(\hat{\sigma}^2_{\eta i}) = \frac{(m_i - 1)}{m_i} \sigma^2_{\eta i}$. Thus the bias of the maximum likelihood estimate $\hat{\sigma}^2_{\eta i}$ as given by (I-16) may be removed by multiplying by the factor $m_i/(m_i - 1)$. When ρ and C are unknown, the bias of $\hat{\sigma}^2_{\eta i}$ (as given by (I-11)) is not as readily determined but, in the absence of better information, the same factor may be used:

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards: the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.


NBS