ESTIMATING A LINEAR FUNCTIONAL RELATION

Prepared Under Contract No. DA-36-034-ORD-1517 (RD)
(Experimental Designs for Industrial Research)

by

H. F. Smith

Institute of Statistics
Mimeo Series No. 145
January, 1956

1. Introduction and review

How to evaluate from observations, all subject to error, an estimate of a functional relation has been a persistent statistical problem for 80 years. When the relation is linear, when errors of observation are normally distributed, and when either nothing is postulated about distribution of the underlying hypothetical variables or they are assumed to be also normally distributed random variables, a consistent solution is possible only if the ratios of variances and covariances of the errors are known (or alternatively all but one of the second moment parameters of the error distributions). We shall consider here a linear relation between two variates with homogeneous error variances. Usually one assumes the covariance to be zero and the ratio of variances known. With that condition the solution was given by Kummell in 1879 and has been several times rediscovered. Nevertheless the theoretical foundation for Kummell's solution has remained ambiguous; it is known to be consistent, but beyond that its statistical properties such as bias and efficiency have not been investigated. Miss Dent (1935) seems to have been the only writer to attempt to evaluate the sampling variance of the estimated slope and her solution is far from satisfactory. It ignores distinction between parameters and statistics, it is based on a Taylor expansion which is not always convergent, and as it seeks the variance of tan (2β) it degenerates toward infinity in the most important region where the slope, tan β, is near unity.

The purpose of this paper is to show why Kummell's solution is unique; thence to prove that it is efficient and unbiased, with respect to the angle of the line with either coordinate axis; and to obtain its sampling distribution.

Consider the following model. Two variables, η_1 and η_2, are linearly related:

$$\eta_1 = A + B\,\eta_2 \tag{1}$$

$$\text{or} \quad \eta_1 \cos\beta - \eta_2 \sin\beta - \alpha = 0 \tag{2}$$

Experimentation yields paired observations

$$y_{pi} = \eta_{pi} + \delta_{pi}$$

$$p = 1, 2; \quad i = 1 \ldots n$$

The errors of observation, δ_1, δ_2, are assumed to be random variables normally independently distributed with zero means and common variance σ_o^2. Except where otherwise stated nothing is postulated about the distribution of η_1 (or equivalently, owing to the relation (1), of η_2). The model is illustrated in fig. 1 where circles represent equal frequency contours of the distribution of δ_p. In particular we will consider the circles with radius equal to σ_o. Only two sub-populations are shown in the figure although usually n would be substantially greater than two. The relationship (1) is represented by the line AA'.

We shall consider also the more general model where errors of observations are not independently distributed with equal variances. Let the observations then be denoted x_{pi}, and assume them to be normally distributed with variances σ_{11}, σ_{22} and covariance σ_{12} around centers ξ_{pi} which obey the relation

$$\xi_1 = A' + B'\,\xi_2 \tag{3}$$

This model is diagrammed in fig. 2. We consider in particular the contour ellipse

$$\sigma_{22}\varepsilon_1^2 - 2\sigma_{12}\varepsilon_1\varepsilon_2 + \sigma_{11}\varepsilon_2^2 = \sigma_{11}\sigma_{22} - \sigma_{12}^2 \tag{4}$$

where $\varepsilon_p = x_p - \xi_p$. The length of any radius of this ellipse is the standard deviation of any section of the bivariate frequency distribution in the same direction; and the projection of the ellipse on any line is the standard deviation of the marginal distribution of the sub-population projected onto that line.

Provided we know either two of the quantities σ_{pq}, or two ratios among them, this model can be transformed to the previous one by

$$y_1 = a_1x_1 + a_2x_2 \qquad (5)$$

$$y_2 = b_1x_1 + b_2x_2$$

subject to

$$a_1^2\sigma_{11} + 2a_1a_2\sigma_{12} + a_2^2\sigma_{22} = b_1^2\sigma_{11} + 2b_1b_2\sigma_{12} + b_2^2\sigma_{22}$$

$$a_1b_1\sigma_{11} + (a_1b_2 + a_2b_1)\sigma_{12} + a_2b_2\sigma_{22} = 0$$

a transformation which can be made in many ways. All deductions about the "y" model can be transferred to the "x" model by the transformation.

Efforts to find a means of estimating the functional relation (1) or (3), under conditions stated, have followed one of three lines, namely: I by considering criteria of consistency, II least squares, III maximum likelihood. (We exclude from consideration here estimates from moments of higher than second order which become available when ξ is postulated to have a non-normal frequency distribution.) The following review quotes only a few of the papers on the topic to indicate salient features of the literature.

I: Criteria of consistency are exemplified by the proposals of Gini (1921), Seares (1944, 1945) and Hald (1952). Usually this method seeks to apply an adjustment to the regression lines. If the hypothetical values were known with means zero we would have

$$B = \sum \eta_1 \eta_2 / \sum \eta_2^2$$

Consistent estimators for the numerator and denominator are $\sum y_1 y_2$ and $\sum y_2^2 - n\sigma_0^2$ respectively. There are variants. With certain assumptions the Kummell line may be indicated in this way (Lindley, 1947, sec. 7.3).

II. The method of least squares is the commonest approach. The usual idea has been to minimize a sum of squares of deviations of observations from the fitted line. The problem has been to determine in what direction should the deviations be measured.

Adcock (1878-79) and Pearson (1901) minimized sum of squares perpendicular to the line without attention to the ratio of error variances. Roos (1937) however pointed out that this produces a solution which fails to be invariant under change of scale. He considered that the direction in which deviations are measured should depend only on the precisions of the observations and be independent of the slope of the line. He therefore proposed to use deviations at 45° to either coordinate axis when variates are scaled so as to have equal precisions. But Lindley (1947, sec. 8.2) pointed out that even this gives consistent estimates only under rather special conditions.

Kummell (1879) and Deming (1931-43)(assuming $\sigma_{12} = 0$) proposed to minimize

$$S = \sum \left\{ \frac{(x_1 - \xi_1)^2}{\sigma_{11}} + \frac{(x_2 - \xi_2)^2}{\sigma_{22}} \right\} \qquad (6)$$

or, equivalently, a proportionate expression using only the ratio $\lambda = \sigma_{11}/\sigma_{22}$, subject to the restriction (3). Kummell showed that a solution is obtainable only if λ be known and that it is equivalent to minimizing the sum of squares of deviations perpendicular to the line when the variables are scaled so as to have equal error variances, that is when transformed to the form (1). He reached the well known solution of the quadratic equation

$$(S_{11} - \lambda S_{22})B + (\lambda - B^2)S_{12} = 0 \qquad (7)$$

where $S_{11} = \sum (x_1 - \bar{x}_1)^2$, etc.

To fit curves and planes he proposed an approximate method which shows interesting variation on usual procedure. When residuals are not linear in the parameters the classical least squares method begins by expressing the residuals (before squaring) as the linear terms of a Taylor expansion in adjustments to trial parameter values. Kummell first expresses S as a Taylor expansion linear only in the deviations $(x_p - \xi_p)$ and uses this to eliminate the ξ and obtain for the residuals (e.g.: $x_1 - A' - B'x_2$) so-called weights which are usually functions of the parameters. The whole expression S, with ξ thus eliminated, is then expressed as a Taylor expansion in adjustments to trial parameter values, and with the 'weights' remaining as functions of the parameters in subsequent differentiations. (Roos criticizes laxity in his mathematical arguments.)

Deming's approximating procedure is more similar to the usual least squares approximation but differs in that the expansion of the residuals contains simultaneously terms both in $(x_p - \xi_p)$ and in parameter adjustments. It leads to 'weighting' of the observable residuals similar to Kummell's formulation, but with the weights expressed as functions of the trial parameter values and hence entering differentiation as constants. In classical procedure the parameter adjustments tend to zero as iteration proceeds with convergence on the required estimates. Deming's expansion differs in the material respect that the discrepancies $(x_p - \xi_p)$ do not tend to zero and cannot be made to do so without losing contact with the observations which must remain as the anchor for computations. His book repeatedly reiterates that the solutions obtained will differ from the true ones only in squares of the residuals, but it evades enunciating the corollary that since these same squares of residuals constitute the function on whose minimization the solution depends they are not negligible. When the fitted relation is linear the procedure leads to one of the regression lines (albeit with a modified estimate of the error variances). In this case it fails to distinguish (as it appears to purport to do) between regression and the functional relation, and the proposed weighting procedure is wasted effort. When a curved relation is to be estimated it does allow that observations in regions of very steep slope do not get the very high weights which regression would in effect assign to them. But just what may be accomplished seems not to have been threshed out. Recognition of the order of magnitude of neglected terms suggests that bias may still be about as great as by any simpler method; it suggests indeed that the method may be strictly appropriate only when residuals are so small that almost any method of fitting will yield a satisfactory result. (A similar comment was made by J. H. Smith, 1945.)

Lindley (1947) seeks a least squares solution by considering the squares of residuals defined by the form in which the relation happens to be written; in particular using (3) he writes the residuals as

$$(x_1 - A' - B'x_2)$$

With assumptions $\sigma_{11}/\sigma_{22} = \lambda$ and $\sigma_{12} = 0$, he then notes that the variance of these residuals is proportional to $(\lambda + B'^2)$ and states that this must be introduced as a weight so that the function to be minimized is

$$\sum \frac{(x_1 - A' - B'x_2)^2}{\lambda + B'^2} \qquad (8)$$

This produces Kummell's equation, but to imply that it is a weighted least squares solution is, I think, misleading. Weights are introduced into the least squares procedure to take account of variation in precision of the observations. (I omit these here for simplicity. They, Lindley's P_k, Q_k, are easily added if required.) Here all the observations have equal weight, and $(\lambda + B'^2)$ is not a weight in the ordinary sense. Lindley pertinently remarks that his procedure has the advantage "that the redundant ξ are never mentioned", but he does not explain why they can be thus banished.

III. The maximum likelihood solution has been considered by Dent (1935), Lindley (1947) and Kendall (1950, 1954). They begin by writing the likelihood, assuming $\sigma_{12} = 0$, as

$$(2\pi)^{-n}(\sigma_{11}\sigma_{22})^{-n/2} \exp\left\{-\frac{1}{2} \sum\left[\frac{(x_2 - \xi_2)^2}{\sigma_{22}} + \frac{(x_1 - A' - B' \xi_2)^2}{\sigma_{11}}\right]\right\} \quad (9)$$

and treat ξ_{2i} as parameters. With no further postulates Lindley and Kendall conclude that all is not well with the resultant equations because they lead to the ratio of estimates of σ_{22}: σ_{11} being β^2 -- an unacceptable result.

While admitting the result to be unacceptable it is scarcely by itself adequate reason for rejecting the method and they do not elucidate why this result appears. Since the likelihood is formulated for 2n observations indeterminance is not due, as has been suggested, to trying to estimate more parameters than there are observations. We shall see later that the trouble is that, relative to distribution of the only deviations which are observable, these two parameters enter as a single unit.

This formulation, with the ratio σ_{11}/σ_{22} known, is easily seen to be equivalent to Kummell's least squares formulation. The unpleasant feature lies in regarding ξ_{pi} as parameters to be estimated, Neyman and Scott's (1947) "incidental parameters". The word "parameter" as used in statistics has not yet been very specifically defined. Relative to the theory of maximum likelihood it may be defined as a characteristic constant of a probability distribution. To specify a particular parameter we must be able to specify the population of random variables of which it is a characteristic. That being done it is at least theoretically possible to return again and again to re-sample the specified population, thereby increasing the sample size from which its characters may be estimated. But that is just what, in the problem before us, cannot be done. A basic feature of the problem is that we never know and can never state that any two or more pairs of observations are drawn from the same sub-population with a particular ξ_{1i}, ξ_{2i}; they are never definable as the characters of a specifiable population and we can never increase sample sizes for their estimation as is required to demonstrate the optimum properties of maximum likelihood estimators. Furthermore we are not interested in the ξ_{pi} individually, we are concerned only to estimate the line as a whole. The ξ_{pi} are essentially variables, not parameters at all. They may or may not be random variables, that is variables with an associated probability distribution, depending on the procedure by which observations are selected. If they are random variables the parameters of their distribution may become parameters also of the overall distribution of the observable x and appropriate formulation is self evident (appendix 2). If the location of observations is chosen in a way which precludes assigning a probability distribution to ξ, to treat them as parameters may yet be inappropriate, and we should seek some other method to eliminate them from the problem.

2. Least squares formulation

The method of least squares seeks estimates of parameters which minimize a sum of squares of residuals which are usually expressible as

$$x_i - \theta_i \qquad (10)$$

where x_i are observations, and θ_i are functions of the estimandsand known constants. For the Gauss-Markoff theorem to be applicable, with consequent nice statistical properties of the estimators, it is necessary (David and Neyman, 1937) that

(i) $E(x_i) = \theta_i$

(ii) θ_i be a linear function of the estimands

(iii) the relative weights of x_i be known.

Little seems to be known about the precise statistical properties of estimators when these conditions are not met. Conditions (i) and (iii) are invariably assumed. Failure of (ii) creates no difficulty in principle for estimating the parameters, beyond that the solution may have to be approached by iteration; but estimators may no longer be unbiased, and reliability of approximations to their variances and covariances, based on linear approximations, seems uncertain.

When we have to deal with a relation between observations all of which are subject to error we do not obtain a clean separation between observations and estimands as at (10). We must deal with residuals which are mixed functions of the two, for example from (1) or (2);

$$r = y_1 - A - By_2 \text{ or } y_1 \cos\beta - y_2 \sin\beta - \alpha.$$

This additional complication has been slurred over by writers who have endeavored to apply the principle of least squares to such situations.

The residuals of classical least squares, and relative to which the principle was developed, are univariate quantities. The residuals with which we have to deal are compounded of two variates, but once compounded in a defined manner the compound becomes again a univariate quantity and should have a univariate probability distribution. The only quantities on which inference must be based are deviations from observed points to the fitted line; there are n such quantities. The hypothetical ξ or η, being irrelevant to the problem, if they can be eliminated, **the deviations must be defined by measurement in some specified** direction.

It seems natural to assume as a first requirement that the residuals should be formulated so that $E(r) = 0$. That condition is satisfied for all formulations which have been proposed when expectation is taken to imply expectation over the bivariate distribution of y_1 and y_2, or of x_1 and x_2, in each sub-population. But after noting that the residuals are essentially univariate quantities it seems reasonable to consider their distribution about the line as conditioned by the direction in which it has been decided to measure them (cf. app. 1). Suppose one may decide to measure the deviations parallel to some arbitrary direction BB', fig. 2. The mean values of conditional distributions formed by sectioning a bivariate distribution in that direction then lie on a diameter CC' which bisects all chords of the contour ellipse which are parallel to BB'. The conditional expectations of r are then quantities QR whose magnitude depends on the distance of BB' from ξ_1 ξ_2. The net effect of minimizing a sum of squares of such deviations is to bias the estimated line in the direction CC'. However such bias can be eliminated if the deviations be measured in one, and only one, particular direction. If the deviations be measured in a direction parallel to tangents to an equi-frequency contour ellipse at its intersection with the functional line AA', that is in the direction DD', fig. 2, the locus of expectations CC' then coincides with AA' and the expectations of deviations (QR) are <u>identically zero independently of</u> $\underline{\xi_1}$, $\underline{\xi_2}$. When deviations are so measured we may regard them as normally distributed <u>about the line AA'</u>, and thus, and only thus, the incidental variables ξ can be eliminated from the problem.

When transformed to our "y" model, figure 1, the condition is seen to be when deviations for independent errors of equal variance are measured perpendicular to the line. Furthermore the variance of such deviations is immediately seen to be σ_o^2. In figure 2 the standard deviation is equal to the length of the radius vector, TD", parallel to DD', for the ellipse (4); but the expression for this length is not simple.

3. Maximum likelihood formulation

For simplicity consider the standardized "y" model, figure 1. The deviations to be considered are those perpendicular to the line. From any text book of analytical geometry (or from elementary trigonometry) the distance of a point from the line is

$$(y_1 - A - By_2)(1 + B^2)^{-1/2} \qquad (11)$$

We have seen that these deviations are normally distributed with mean zero and variance σ_o^2. The likelihood is therefore

$$L = (2\pi\sigma_o^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_o^2}\sum\frac{(y_1 - A - By_2)^2}{1 + B^2}\right] \qquad (12)$$

It is maximized when $\sum(y_1 - A - By_2)^2 / (1 + B^2)$ is minimized. This is Lindley's formulation but we now see that the factor $(1 + B^2)$ (or, more generally, $\lambda + B'^2$) is introduced to evaluate the only deviation from whose distribution the redundant ξ can be eliminated. It is not entered as a weight; that it happens to be proportional to the variance of the marginal distribution when the whole bivariate distribution is projected onto a vertical, in which direction deviations are $(y_1 - A - By_2)$, is an incidental circumstance of the bivariate normal distribution. Since the Kummell solution can now be expressed as a maximum likelihood estimator not involving "incidental parameters", it follows that it is asymptotically efficient.

For further work the equation of the functional line is better expressed in the intercept form (2). In regression analysis the slope of a regression is normally distributed only because values of the independent variable are taken as given constants. Furthermore we cannot evaluate a regression unless these have some spread, that is $\sum(x - \bar{x})^2$ cannot be zero, and the estimated slope cannot approach infinity. In the problem here considered B may approach infinity from either direction while the spread of x_2 is still substantial (depending on σ_{22}). A little consideration shows that the system is circularly symmetric, therefore to obtain a statistic which may be symmetrically distributed and independent of β we should consider the angle of the estimated line, rather than its slope whose distribution may be very skew and dependent on β.

Taking the parameters to be estimated as β and α, the log likelihood is now

$$\ln L = \text{const.} - \frac{n}{2} \ln \sigma_o^2 - \frac{1}{2\sigma_o^2} \sum (y_1 \cos\beta - y_2 \sin\beta - \alpha)^2 \qquad (13)$$

To derive the estimators assume that y_1, y_2 are measured from their means so that $\sum y_p = 0$.

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma_o^2} \sum (y_1 \cos\beta - y_2 \sin\beta - \alpha) \qquad (14)$$

whence $\hat{\alpha} = 0$.

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma_o^2} \sum (y_1 \cos\beta - y_2 \sin\beta - \alpha)(y_1 \sin\beta + y_2 \cos\beta) \qquad (15)$$

which is zero when

$$\sum \left(\frac{1}{2} (y_1^2 - y_2^2) \sin 2\hat{\beta} + y_1 y_2 \cos 2\hat{\beta}\right) = 0$$

$$\tan 2\hat{\beta} = \frac{2 \sum y_1 y_2}{\sum y_2^2 - \sum y_1^2} \qquad (16)$$

which is equivalent to Kummell's solution (7).

Further deductions are easier in terms of transformed variables obtained by rotating to coordinate axes parallel and perpendicular to the theoretical line. Define

$$u_1 = y_1 \cos\beta - y_2 \sin\beta \qquad (17)$$

$$u_2 = y_1 \sin\beta + y_2 \cos\beta$$

$$E(u_1) = \eta_1 \cos\beta - \eta_2 \sin\beta = \alpha \text{ (Constant independently of i)}$$

$$E(u_{2i}) = \eta_{1i} \sin\beta + \eta_{2i} \cos\beta = U_i \quad \text{say} \qquad (18)$$

U is an alternative variable to η_1 or η_2 and represents the distance of hypothetical points as measured along the line, ST in figure 1. We then have

$$\ln L = \text{const.} - \frac{n}{2}\ln\sigma^2 - \frac{\sum(u_1-\alpha)^2}{2\sigma^2} \tag{19}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum(u_1-\alpha)^2}{2\sigma^4}$$

$$= 0 \text{ when } \hat{\sigma}^2 = \sum(u_1-\alpha)^2/n \tag{20}$$

Since $du_1/d\beta = -u_2$, and u_2 is a constant relative to each conditional distribution along a fixed line DD', the elements of the information matrix are

$$E\left(\frac{\partial \ln L}{\partial \alpha}\right)^2 = -\frac{\partial^2 \ln L}{\partial \alpha^2} = \frac{n}{\sigma^2} \qquad E_0\left(\frac{\partial \ln L}{\partial \alpha}\right)\left(\frac{\partial \ln L}{\partial \beta}\right) = -\frac{\partial^2 \ln L}{\partial\alpha\,\partial\beta} = \frac{\sum u_2}{\sigma^2}$$

$$E_0\left(\frac{\partial \ln L}{\partial \beta}\right)^2 = -\frac{\partial^2 \ln L}{\partial \beta^2} = \frac{\sum u_2^2}{\sigma^2} \qquad E\,\frac{\partial^2 \ln L}{\partial\alpha\,\partial\sigma^2} = E\,\frac{\partial^2 \ln L}{\partial\beta\,\partial\sigma^2} = 0$$

$$E\left(\frac{\partial \ln L}{\partial \sigma^2}\right)^2 = -E\,\frac{\partial^2 \ln L}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} \tag{21}$$

where E_0 implies conditional expectation given u_2. However we have here the peculiarity that as β is varied there is a shift in the directions along which the deviations are measured. On more accurately evaluating var $(\hat{\beta})$ it will turn out, as we might intuitively expect, that a better approximation to the variance-covariance matrix is given by replacing u_2 by U wherever it occurs in (21). A peculiar feature of this formulation is that, allowing u_2 to be a variable, $E(u_{2i}) = U_i$ and $du_2/d\beta = u_1$, we obtain $-E\,\frac{\partial^2 \ln L}{\partial \beta^2} = \frac{\sum U^2}{\sigma^2}$ as required; but the alternative form $E\left(\frac{\partial \ln L}{\partial \beta}\right)^2$ still remains as $\frac{\sum U^2 + n\sigma^2}{\sigma^2}$. The product

term $\frac{\partial \ln L}{\partial \alpha} \cdot \frac{\partial \ln L}{\partial \beta}$ on the other hand has the same expectation as $-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta}$.
Full theoretical implication of that inconsistency eludes me; it is evidently associated with the condition that the direction in which the postulated sampling distribution of observations is defined "wobbles" with errors in estimating its parameters. We shall see later that the condition under which the indicated variances are approached is not $n \to \infty$, but that the ratio (σ^2/ second moment of U) $\to 0$.

The asymptotic variance-covariance matrix is therefore indicated to be

$$\begin{array}{c} \\ \sigma^2 \end{array} \begin{array}{c} \begin{matrix} \alpha & \quad\quad\beta\quad\quad & \sigma^2 \end{matrix} \\ \begin{bmatrix} \frac{1}{n} + \frac{\overline{U}^2}{A} & \frac{-\overline{U}}{A} & 0 \\ & \frac{1}{A} & 0 \\ & & \frac{2\sigma^2}{n} \end{bmatrix} \end{array} \qquad (22)$$

where $\overline{U} = \sum U/n$, $A = \sum (U - \overline{U})^2$.

Alternative formulation of the likelihood, and of the associated variance-covariance matrix of the estimators, if estimates of individual U_i may be deemed relevant, is indicated in appendix 1. The likelihood formulation if U is also a normally distributed random variable is stated in appendix 2.

4. Summary of additional results

The distribution of $\hat{\beta}$ is symmetric. By approximating $(\hat{\beta}-\beta)$ by a power series of variables, whose joint moment generating function can be easily obtained, it has been ascertained that for finite samples the variance of β is

$$\frac{\sigma^2}{A}(1+w) + \frac{4\sigma^4}{A^2}(2w+w^2) + \frac{8\sigma^6}{3A^3}(-2+27w+48w^2+16w^3) + O(A^{-4})$$

where $w = (n-1)\sigma^2/A$. (a1)

The parameter of kurtosis is

$$\gamma_2 = \frac{2\sigma^2}{A}\left(2+w-\frac{3w^3}{1+w^2}\right) + O(A^{-2}) \tag{a2}$$

If

$$\hat{u}_1 = (y_1-\bar{y}_1)\cos\hat{\beta} - (y_2-\bar{y}_2)\sin\hat{\beta}$$

then

$$E(\sum\hat{u}_1^2) = \sigma^2\left[n-2-w+\frac{w}{n-1}(1-2w-w^2) - \frac{2w^2}{(n-1)^2}(11+16w+9w^2+3w^3) - O(n^{-3})\right]. \tag{a3}$$

Since (n-2) will usually be large relative to w we may for most practical purposes suppose (n-2) "degrees of freedom". If then we accept as estimator of the error variance

$$\hat{\sigma}^2 = \sum\hat{u}_1^2/(n-2) \tag{a4}$$

its mean square error is

$$E(\hat{\sigma}^2-\sigma^2)^2 = \frac{2\sigma^4}{n-2} - \frac{\sigma^4(4w+w^2)}{(n-2)^2} + \sigma^4.O(n^{-3}) \tag{a5}$$

If

$$\hat{u}_2 = (y_1-\bar{y}_1)\sin\hat{\beta} + (y_2-\bar{y}_2)\cos\hat{\beta}$$

a satisfactory estimator for the variance of β is

$$\hat{var}(\hat{\beta}) = Q(1 + (n-1)Q) \qquad (a6)$$

$$\text{where } Q = \frac{\sum \hat{u}_1^2}{(n-2)\sum \hat{u}_2^2 - n\sum \hat{u}_1^2} \qquad (a7)$$

$$E(\hat{var}(\hat{\beta})) = \frac{\sigma^2}{A}\left(1 + w + \frac{1}{n-1}\right) + \frac{\sigma^4}{A^2}(9 + 22w - 8w^2) + 0\,\frac{\sigma^4}{nA^2} \qquad (a8)$$

which compares reasonably favorably with the value required for an unbiased estimator as indicated by (a1).

An approximate fiducial interval for β is given by

$$\sin^2 2\theta = \frac{4t^2}{n-2} \cdot \frac{S_1 S_2 - S_c^2}{(S_2 - S_1)^2 + 4S_c^2} \qquad (a9)$$

where $\theta = \hat{\beta} - \beta$

$S_p = \sum (y_p - \bar{y}_p)^2$

$S_c = \sum (y_1 - \bar{y}_1)(y_2 - \bar{y}_2)$

t = Student's t for the required fiducial probability and $(n - 2)$ degrees of freedom.

The error in this interval seems likely to be negligible for the kind of data to which such lines are usually fitted, say for fiducial intervals of less than $\pi/4$.

I have been unable to obtain the exact sampling distribution of β under the model so far assumed. But if we add the postulate that U is normally distributed about zero with variance σ_U^2, then, with notation as follows

$$\sigma_1^2 = \sigma^2(u_1) = \sigma_o^2$$

$$\sigma_2^2 = \sigma^2(u_2) = \sigma_U^2 + \sigma_o^2$$

$$Q = \sigma_1^{2\prime} + \sigma_2^2$$

$$C = \sigma_1^2 \cos^2\theta + \sigma_2^2 \sin^2\theta$$

n is even = 2k + 4

the probability density function of $\theta = (\hat{\beta} - \beta)$ is

$$f(\theta) = \frac{2^{n-3}(n-2)}{\pi} \frac{(\sigma_1\sigma_2)^{n-1}}{Q^2 C^{k+2}} \sum_o^k (k+1-i)\binom{k+i}{i}\left(\frac{C}{Q}\right)^i \qquad \text{(a10)}$$

Proof of these statements, and the appendices noted above, will be submitted in a later report.

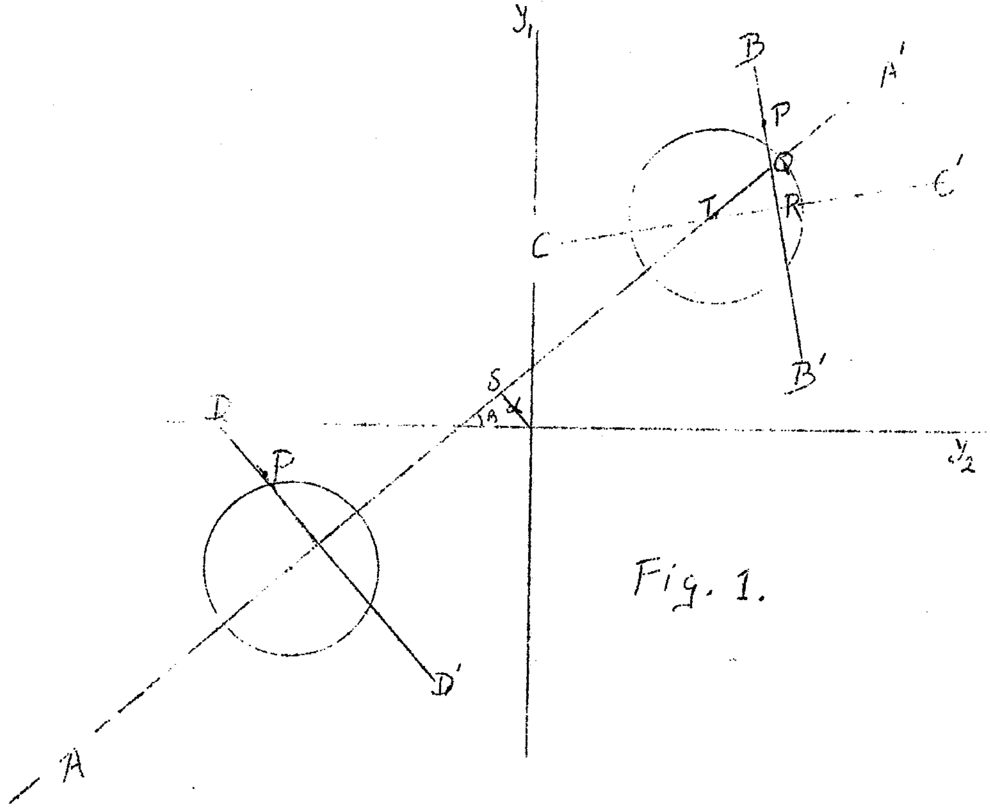

Fig. 1.

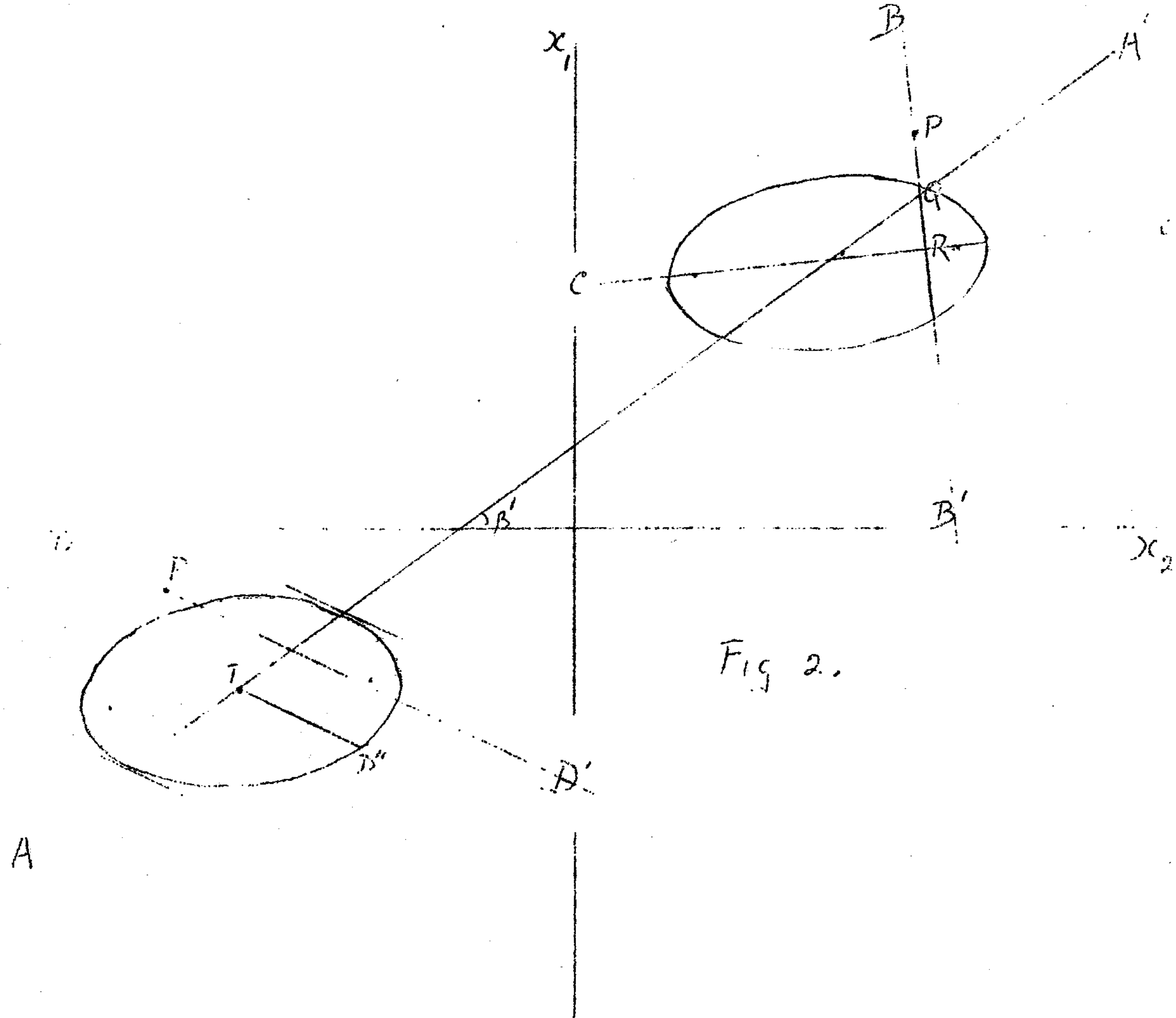

Fig 2.

Literature cited

Adcock, R. J. (1877-78) Least Squares. The Analyst (Des Moines, Iowa) 4:183-184 and 5:21-22.

__________ A problem in least squares. ibid. 5:53-54 and 149-150.

David, F. N. and Neyman, J. (1937) Extension of the Markoff Theorem on least squares. Stat. Res. Mem. 2:105-116.

Deming, W. Edwards (1931-35) The application of least squares. Phil. Mag. Ser. 7, 11:146-158; 17:804-829; 19:389-402.

__________ (1943) Statistical adjustment of data. Wiley, New York.

Dent, Beryl M. (1935) On observations of points connected by a linear relation. Proc. Roy. Phys. Soc. 47:92-106.

Durbin, J. (1955) Errors in variables. Rev. Inst. Intern. Stat. 22:23-32.

Gini, C. (1921) Sull' interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali. Metron 1:63-82.

Hald, A. (1952) Statistical theory with engineering applications. Wiley, New York (sec.19, 14).

Kendall, M. G. (1951) Regression, structure and functional relationship I. Biometrika 38:11-25.

__________ (1954) Notes on multivariate analysis. Inst. Stat. Mimeo Ser. No. 95.

Kummell, C. H. (1879) Reduction of observation equations which contain more than one observed quantity. The Analyst (Des Moines, Iowa) 6:97-105.

Lindley, D. V. (1947) Regression lines and the linear functional relationship. J. Roy. Stat. Soc. Suppl. 9:218-244.

Neyman, J. and Scott, E. L. (1948) Consistent estimates based on partially consistent observations. Econometrica 16:1-32.

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. Phil. Mag. Ser. 6, 2:559.

Roos, C. F. (1937) A general invariant criterion of fit for lines and planes where all variates are subject to error. Metron 13:3-20.

Seares, F. H. (1944-45) Regression lines and the functional relation. Astrophys. J. 100:255-263; 102:366-376.

Smith, J. H. (1945) (Book review.) J. Am. Stat. Assoc. 40:380.

Villars, D. S. and Anderson, T. W. (1943) Some significance tests for normal bivariate distributions. Ann. Math. Stat. 14:141-148.