66–10,440

SHIH, Chang Sheng, 1931–

INTERVAL ESTIMATION FOR THE EXPONENTIAL MODEL AND THE ANALYSIS OF ROTATION EXPERIMENTS.

Iowa State University of Science and Technology
Ph.D., 1966
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

INTERVAL ESTIMATION FOR THE EXPONENTIAL MODEL AND THE ANALYSIS OF ROTATION EXPERIMENTS

by

Chang Sheng Shih

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1966

TABLE OF CONTENTS

I. INTRODUCTION

Modern experimentation is often concerned with investigating the response of a variable y_i to a vector of independent or input variables x. The cause-and-response relationship between x and y_i may be expressed in a general functional form

$$y_i = f(x,\theta) \tag{1.1}$$

where θ is a vector of unknown parameters. Considerable effort has been made to study, under various assumptions, the statistical properties of Equation 1.1. The assumption that the Equation 1.1 is linear in the parameter θ permits simple statistical treatment. Under such assumption, a linear statistical model corresponding to (1.1) may be written in matrix notation as

$$y = X\theta + e \tag{1.2}$$

where y and θ are respectively n- and p-vectors, X is an n x p matrix of rank $p < n$, whose elements are input variables x_{ij}, $i = 1,\ldots,m$, $j = 1,\ldots,p$, and e is an n-vector of errors. Estimation of the parameter presents little difficulty. If the errors are uncorrelated, have all zero mean and the same unknown variance, the least squares procedure yields the unbiased, efficient estimator $\hat{\theta}$ of θ. Further, if the errors are assumed to be normally distributed, exact sample distribution theory of $\hat{\theta}$ is available and exact confidence regions for θ based on $\hat{\theta}$ can be constructed.

The true statistical model relating a response y to the input x is normally unknown to the investigator. In exploring the relation (1.1), therefore, an investigator is constantly faced with the problem of hypothesizing a model which would approximate or describe well the true cause-and-response relationship. Relevant independent variables must be included and functional form of (1.1) must be specified. This is often accomplished with the help of common sense, imagination and the basic knowledge in the domain of the subject of study. Inadequate models may result from poor judgment or fallacious reasoning. While omission of an important independent variable may lead to a poor description of the cause-and-response relationship, error in model specification, even in the presence of all relevant independent variables, may provide little information to the research worker. It is for this latter reason that the linear model (1.2), despite the support of the well-developed statistical theory regarding estimation and hypothesis testing, proves time and again to be unsatisfactory in many fields of natural phenomena and scientific behavior.

An area in which the linear model (1.2) is inadequate is known as "growth processes". By a growth process we refer to a phenomenon in which the response y follows a certain monotonic trend with unknown rate of change and which approaches an unknown asymptote as the independent variable(s)

tends to infinity.[1] Biologically, a growth process may describe the growth of an insect population, of an individual plant or animal approaching maturity, or the yield of a crop as a function of level of nitrogen fertilizer applied.

The curve that describes the correspondence between y and x in a growth process is called a growth curve. Mathematical functions capable of defining growth curves vary in forms but most are non-linear in the parameters. Mitscherlich (21) first introduced the function

$$y = A[1 - 10^{-c(x+b)}] \tag{1.3}$$

in his work concerning fertility experiments. This is now known as Mitscherlich's law, generally used to express the response to a fertilizer. In many areas of study, the equation of a growth curve is more often found in the form

$$y = \alpha + \beta e^{-\gamma x} \qquad \gamma > 0 \tag{1.4}$$

or

$$y = \alpha + \beta \rho^{x} \qquad 0 < \rho < 1 \tag{1.5}$$

which are known as the exponential model, Spillman function or asymptotic regression. The exponential model has α as its asymptote, and $\alpha + \beta$ as its intercept, i.e. the value of y at $x = 0$. The function is monotonic, with the rate of change of y being a linear function of y, as can be seen by differentiating (1.5) with respect to x

[1] In point of fact this model itself is often only true for a range of the independent variable(s).

$$dy/dx = \beta\rho^{x} \log \rho$$
$$= (y - \alpha) \log \rho.$$

In terms of ρ, values of y change in the following manner: for every unit increase of x, y moves from y(x) toward α by a fraction $1-\rho$ of the distance between y(x) and α, where y(x) is the response corresponding to the input x.

In the exponential models (1.4) and (1.5), the parameters γ and ρ, entering the model non-linearly, are often referred to as the non-linear parameters, while the parameters α and β are called linear parameters. In contrast to the linear model (1.2), models with one or more non-linear parameters are said to be non-linear models.

In statistical theory procedures for estimation and hypothesis testing are usually developed without any prior information on the parameters. When prior information on the parameters is available, it is possible, in principle, to obtain statistical procedures which are optimum given the prior information. Unfortunately, the prior information is often vague and fragmentary, making the evaluation of the procedure in application more difficult. There are situations, however, when the knowledge on the parameters is certain. In the exponential model (1.5), for instance, the parameter space of the non-linear parameter ρ is known to be the interval (0, 1). In developing statistical procedures, such information should be fully utilized.

In agriculture a class of cropping system that has been practiced for centuries is crop rotation. Experiments to study the nature of crop rotation date back to the mid nineteen century and are still conducted in many agricultural experiment institutes all over the world. Rotation experiments are long-term experiments that demand great resources in personnel, funds and experimental land. Although valuable information has been obtained during the past, the reward in many cases is not proportional to the investment. One reason for this is the lack of a suitable model that is simple to handle and that provides meaningful explanation of the effects observed in crop rotation. A conventional method of estimating rotation effects by mean crop yield over years and explaining the trend of yield by polynomial regression is inadequate in describing rotation effects. In view of the fact that the yield of a crop in a rotation tends to stabilize in the long run, a growth process for rotation effect expressed by the exponential model offers a solution to the problem of analyzing rotation experiments.

II. ESTIMATION OF THE EXPONENTIAL MODEL

A. A Review of the Principle of Least Squares

We shall review briefly the principle of least squares for the linear model

$$y = X\theta + e \quad . \tag{1.2}$$

Assume that

$$E(e) = 0 \tag{2.1}$$

$$E(ee') = \sigma^2 I \tag{2.2}$$

where E denotes the expectation of the expression in the parentheses, and I is an identity matrix. It is required to determine $\hat{\theta}$, the least squares estimator of θ, for which the sum of squares

$$Q = e'e = (y - X\hat{\theta})'(y - X\hat{\theta}) \tag{2.3}$$

is minimum. Differentiating Q with respect to $\hat{\theta}$ and equating the results to zero give the set of normal equations

$$X'X\hat{\theta} = X'y. \tag{2.4}$$

Because X is of full rank, the unique solution of (2.4) is

$$\hat{\theta} = (X'X)^{-1}X'y.$$

$\hat{\theta}$ is unbiased and efficient, with variance

$$\mathrm{Var}(\hat{\theta}) = \sigma^2(X'X)^{-1}$$

where σ^2 is estimated by

$$\hat{\sigma}^2 = (y - X\hat{\theta})'(y - X\hat{\theta})/(n-p)$$

$$= (y'y - \hat{\theta}'X'y)/(n-p).$$

If in addition normality is assumed for the errors, $\hat{\theta}$ is

also the maximum likelihood estimator of θ, and we have the following well-known properties about the precision of $\hat{\theta}$. Under normal theory, $\hat{\theta}$ is $N(\theta, \sigma^2(X'X)^{-1})$, $\hat{\theta}'X'y$ and $y'y-\hat{\theta}'X'y$ are independently distributed as $\sigma^2\chi_p^2$ and $\sigma^2\chi_{n-p}^2$, respectively. Therefore the ratio

$$\frac{(n - p)(\hat{\theta}'X'y)}{p(y'y - \hat{\theta}'X'y)}$$

follows the F distribution with p and n-p degrees of freedom, and a γ level confidence region for θ based on $\hat{\theta}$ is given by

$$\frac{(n - p)(\hat{\theta}'X'y)}{p(y'y - \hat{\theta}'X'y)} \leq F(\gamma;\ p,\ n-p).$$

If the errors are correlated and

$$E(ee') = V$$

where V is nonsingular, the simple least squares estimator obtained above is unbiased but not efficient. To obtain efficient, unbiased estimator of θ, the following generalized least squares procedure is required.

Since V is symmetric, there exists a nonsingular n x n matrix T such that $TVT' = I$. Let

$$Ty = TX\theta + Te.$$

Since $T'T = V^{-1}$, the set of normal equations

$$(TX)'(TX)\hat{\theta} = (TX)'Ty$$

i.e.

$$X'V^{-1}X\hat{\theta} = X'V^{-1}y$$

leads to the efficient, unbiased estimator

$$\hat{\theta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

with variance

$$\mathrm{Var}(\hat{\theta}) = (\underset{\sim}{X}'V^{-1}X)^{-1}.$$

B. Point Estimation

Returning to the exponential model, we write (1.5) in the following form

$$y_i = \alpha + \beta\rho^{x_i} + e_i \qquad i = 0,1,\ldots,n-1 \tag{2.5}$$

$$E(y_i) = \alpha + \beta\rho^{x_i} \tag{2.6}$$

where $0 < \rho < 1$. For convenience we shall often omit the subscript i. Further, we assume

$$E(e_i) = 0, \quad E(e_i^2) = \sigma^2, \quad E(e_i e_j) = 0, \quad i \neq j. \tag{2.7}$$

Any deviation of error assumption from (2.7), including the assumption of normality, will be stated explicitly.

The principle of least squares when applied to the exponential model (2.5) leads to the set of normal equations

$$\begin{aligned} -na \quad & -b\Sigma r^{x} \quad & +\Sigma y \quad & = 0 \\ -a\Sigma r^{x} \quad & -b\Sigma r^{2x} \quad & +\Sigma r^{x} y \quad & = 0 \\ -ab\Sigma x r^{x-1} \quad & -b^{2}\Sigma x r^{2x-1} \quad & +b\Sigma x r^{x-1} y \quad & = 0 \end{aligned} \tag{2.8}$$

where a, b and r are the least squares estimators of α, β and ρ, respectively.

As explicit closed form solutions for a, b and r as functions of the observations are impossible, one resorts to other methods of estimation. One crude method is to guess a value for ρ and then to regress y on ρ^{x}. A second method also reduces the problem to one of linear regression.

It calls for graphical estimation of the asymptotic value α and the transformation of the equation to

$$\log(y - \alpha) = \log \beta + x \log \rho.$$

The difficulty arising in fitting an exponential model, or more general non-linear models, has led in the past years to rather extensive study of the problem of estimation.

1. Non-linear least squares

Non-linear least squares procedures for estimating the parameters in the exponential model are based on the principle of successive approximation. Initial trial values, or preliminary estimates, of some or all of the parameters are used for the first iteration to yield corrections, which are then added to the initial trial values to form a new set of trial values for further iteration. The process is repeated until the corrections become sufficiently small.

a. Stevens' method Stevens (32) described a least squares method for estimating α, β and ρ, and the errors of these estimates. By differentiating the normal equations (2.8) with respect to a, b and r in turn, the information matrix is found to be

$$D = \begin{pmatrix} n & \Sigma r^x & b\Sigma x r^{x-1} \\ \Sigma r^x & \Sigma r^{2x} & b\Sigma x r^{2x-1} \\ b\Sigma x r^{x-1} & b\Sigma x r^{2x-1} & b^2\Sigma x^2 r^{2x-2} \end{pmatrix}.$$

The asymptotic covariance matrix is then obtained by inverting the information matrix:

$$V = \begin{pmatrix} F_{aa} & F_{ab} & F_{ar}/b \\ F_{ab} & F_{bb} & F_{br}/b \\ F_{ar}/b & F_{br}/b & F_{rr}/b^2 \end{pmatrix} .$$

The F values (F_{aa}, F_{ab}, F_{bb}, F_{ar}, F_{br} and F_{rr}) are functions of x and r only, for the value of b serves only as a coding factor in inverting D.

To find the estimates of α, β and ρ, we first insert trial values of a, b and r, say a_o, b_o and r_o, in (2.8). Instead of zero, the right hand side of the equations will take some values A, B and R respectively. Let M be the matrix obtained from D by deleting all b's and b^2, and

$$M \begin{pmatrix} da_o \\ db_o \\ b_o dr_o \end{pmatrix} = \begin{pmatrix} A \\ B \\ R \end{pmatrix} .$$

If M is non-singular, then

$$\begin{pmatrix} da_o \\ db_o \\ b_o dr_o \end{pmatrix} = M^{-1} \begin{pmatrix} A \\ B \\ R \end{pmatrix}$$

$$= \begin{pmatrix} F_{aa}A + F_{ab}B + F_{ar}R \\ F_{ab}A + F_{bb}B + F_{br}R \\ F_{ar}A + F_{br}B + F_{rr}R \end{pmatrix}$$

$$= \begin{pmatrix} -a_o + F_{aa}\Sigma y + F_{ab}\Sigma r_o^x y + F_{ar}\Sigma x r_o^{x-1} y \\ -b_o + F_{ab}\Sigma y + F_{bb}\Sigma r_o^x y + F_{br}\Sigma x r_o^{x-1} y \\ F_{ar}\Sigma y + F_{br}\Sigma r_o^x y + F_{rr}\Sigma x r_o^{x-1} y \end{pmatrix} .$$

Hence the trial values of a, b and r for second iteration are taken as

$$a_1 = a_o + da_o = F_{aa}\Sigma_y + F_{ab}\Sigma r_o^x y + F_{ar}\Sigma x r_o^{x-1} y$$

$$b_1 = b_o + db_o = F_{ab}\Sigma y + F_{bb}\Sigma r_o^x y + F_{br}\Sigma x r_o^{x-1} y \qquad (2.9)$$

$$r_1 = r_o + dr_o = r_o + (F_{ar}\Sigma y + F_{br}\Sigma r_o^x y + F_{rr}\Sigma x r_o^{x-1} y)/b.$$

Note that in (2.9) a_1 and b_1 do not depend on a_o and b_o. Thus, if we start with a trial value r_o, we can find a_1 and b_1 without assuming the values of a_o and b_o. Letting $b = b_1$ in (2.9), a new trial value r_1 is then calculated.

The Stevens' method requires a good deal of computation. If the values of x are equally spaced, however, much of the computational work can be avoided by coding the x as 0,1,..., n-1 and tabulating the F values. Tables of F values have been provided by Stevens' and extended by others (15,18,19) for various values of r and for n up to 40.

The estimates obtained by Stevens' method have asymptotically the same covariance matrix as the maximum likelihood estimators. However, the procedure depends heavily on the choice of the initial trial value. The estimate r may fail to converge during the process of iterations, and for r values near either zero or one, the matrix M is ill-conditioned.

b. Pimentel-Gomes' method Pimentel-Gomes (28) has shown that, with equally spaced x, efficient estimates of ρ based on (2.8) can be obtained by solving equations of the type

$$y_o J_o(r) + y_1 J_1(r) + \ldots + y_{n-1} J_{n-1}(r) = 0$$

where the $J(r)$ are complicated polynomials in r. Estimates of α and β are expressed as functions of r. To simplify the procedure, he provided tables for the values of $J(r)$ for $n = 4$ and $n = 5$ and for $r = 0.00(0.01)1.00$. In practice r can be obtained by trial and error, with the help of these tables, more rapidly than the Stevens' method. Unfortunately, for values of n not tabulated, the procedure is more prohibitive computationally than the Stevens' method. Also the method relies on Stevens' F values for computing variances and covariances of the estimates.

c. Gauss-Newton method and modifications The method of successive approximation based upon expanding a non-linear model in a Taylor series was originally proposed by Gauss in 1821. Let $f(x_i, \theta)$ be non-linear in the parameters θ, and

$$E(y_i) = f(x_i, \theta) = f, \qquad i = 1,\ldots,n$$

$$\frac{\partial f}{\partial \theta_j} = f_j(x_i, \theta) = f_j, \qquad j = 1,\ldots,p.$$

Expand f in the Taylor series through the first derivatives

$$f(x_i, \theta) \doteq f(x_i, \theta^o) + \sum_j (\theta_j - \theta^o) \frac{\partial f}{\partial \theta_j}\Big|_{\theta^o}$$

or, in matrix notation,

$$E(y) \doteq f^o + X\delta \tag{2.10}$$

where θ^o is a p-vector of trial values of θ, y is an n-vector, f^o is an n-vector of elements $f(x_i, \theta^o)$, δ is a p-vector of elements $\delta_j = \theta_j - \theta_j^o$, and X is an n x p matrix whose ij-th

element is $\frac{\partial f}{\partial \theta_j}\Big|_{\theta^o}$, the partial derivatives being evaluated at θ^o. Note that (2.10) is linear in δ. The least squares process then requires the minimization of

$$\begin{aligned} Q &= Q(\theta^o, \hat{\delta}) \\ &= (y - f)'(y - f) \\ &\doteq (y - f^o - X\hat{\delta})'(y - f^o - X\hat{\delta}). \end{aligned} \tag{2.11}$$

We recognize the similarity between (2.11) and (2.3). Hence the normal equations for (2.11) are

$$X'X\hat{\delta} = X'(y - f^o)$$

which yields the solution

$$\hat{\delta} = (X'X)^{-1}X'(y - f^o).$$

Thus (2.11) is minimized by the vector

$$\theta^1 = \theta^o + \hat{\delta} \tag{2.12}$$

which then takes the place of θ^o for further iteration.

For the exponential model (1.5), for example, we have

$$f(x,\theta) = \alpha + \beta\rho^x$$

$$\partial f/\partial\alpha = 1$$

$$\partial f/\partial\beta = \rho^x$$

$$\partial f/\partial\rho = \beta x\rho^{x-1}.$$

While the least squares method gives rise for the linear model (1.2) to solutions which are unique and which are associated with the absolute minimum of Q in (2.3), the Gauss-Newton method ensures neither. In fact, unless θ^o is sufficiently close to some unknown value, the process may lead to one of the local minima that may exist for Q.

Moreover, the process may fail to give a solution because of divergence of the successive iterates.

To improve the method, Levenberg (17) proposed to limit or "damp" the absolute values of the corrections δ of the parameters by minimizing the weighted sum of Q and δ_j^2. This however requires to a certain extent judgments which one might not be able to make, for the weighting factors are to be determined by relative importance of the quantity Q and the corrections in the minimizing process, and by the relative importance of damping the different corrections. A number of other modifications have been discussed in the literature. The maximum neighborhood method developed by Marquardt (20) performs an optimum interpolation between the Taylor series method and the gradient or steepest descent method.

A modification, generally known as the modified Gauss-Newton method, was advocated by Hartley (13). Instead of the value θ^1 in (2.12), Hartley proposed the use of

$$\theta^1 = \theta^0 + v\hat{\delta},$$

where v is a scalar minimizing the expression

$$Q(v) = Q(\theta^0 + v\hat{\delta}), \quad 0 < v < 1.$$

Under certain general assumptions, Hartley was able to prove the convergence of the iterative process and the uniqueness of the solution. Such solution, denoted by θ^+, was shown to exist in a convex "neighborhood region" S^+ and yields the absolute minimum of Q. In practice, as noted by Hartley, the difficulty is to find an initial trial value θ^0 which

is known to be in S^+. Therefore a search of the surface Q is often necessary.

A method suggested by Williams (39, pp. 62-64) requires only the expansion of the non-linear parameter of a non-linear model into a Taylor series. For the exponential model (1.5), the expansion takes the form

$$\begin{aligned} y &= \alpha + \beta\rho^x \\ &\doteq \alpha + \beta(\rho_o^x + x\rho_o^x\delta) \\ &= \alpha + \beta\rho_o^x + \beta_1(x\rho_o^x). \end{aligned}$$

The correction δ is obtained as $\delta = \beta_1/\beta$.

d. Hartley and Booker's method More recently Hartley and Booker (14) developed a method of estimation which avoids the search for the absolute minimum of Q and yet yields two estimators, $\tilde{\theta}$ and $\hat{\theta}$, which are asymptotically 100% efficient under fairly general assumptions. The method requires splitting of the n observations into p groups of equal size, where p is the number of parameters in $f(x,\theta)$. Let

$$\bar{y}_h = \frac{1}{m} \sum_i y_{hi} \qquad m = n/p, \quad h = 1,\ldots,p,$$

$$\bar{f}(h,\theta) = \sum_i f(x_{hi},\theta) .$$

Let θ^* be the solution of a system of m non-linear equations

$$\bar{y}_h = \bar{f}(h,\theta^*).$$

The solution θ^* is obtained by the modified Gauss-Newton iteration as the absolute minimum of

$$\bar{Q}(\theta) = \sum_h [\bar{y}_h - \bar{f}(h,\theta)]^2.$$

The estimator θ^* was shown to be consistent under certain general assumptions. Using θ^* as a starting value, $\tilde{\theta}$ is obtained by carrying out one iteration step of the Gauss-Newton method and $\hat{\theta}$ is obtained by carrying out the modified Gauss-Newton iteration to convergence.

2. Ratio estimators of ρ

Consider the model (2.6) when $x_i = 0, 1, 2$. We have

$$E(y_o) = \alpha + \beta$$

$$E(y_1) = \alpha + \beta\rho$$

$$E(y_2) = \alpha + \beta\rho^2 .$$

Since

$$\frac{E(y_2-y_1)}{E(y_1-y_o)} = \frac{E(y_2) - E(y_1)}{E(y_1) - E(y_o)} = \rho,$$

an estimator of ρ is the ratio

$$r_R = (y_2 - y_1)/(y_1 - y_o).$$

More general ratio estimators have been discussed (7, 24, 26, 27). Patterson (26) and Patterson and Lipton (27) presented a class of ratio estimators of the form

$$r_R = \frac{A}{B} = \frac{\Sigma w_i y_i}{\Sigma w_i y_{i-1}} \qquad \Sigma w_i = 0, \quad i = 1,\ldots,n-1. \quad (2.13)$$

The ratio estimator r_R is said to be a linear estimator, denoted by r_L, if w_i are independent of y_i; it is called a quadratic estimator, r_Q, if w_i are linear functions of y_i. Several estimators of ρ that are not given in the form of ratio estimators belong in fact to the class (2.13). Thus for example the Stevens' least squares estimate r is a ratio

estimate in which the w_i are complicated functions of r itself, and the internal least squares estimate of ρ by Hartley (11) is a quadratic estimate.

a. Linear estimators Patterson (26) noted that the main uses of the linear estimators r_L are to provide initial estimates of ρ for the non-linear least squares methods and to enable rapid checks to be made on assumed values of ρ. Linear estimators are easy to compute but are not unbiased. Since (if the errors are normal) the linear estimates are the ratio of normal variates, their theoretical variance is infinite (5). However, they are consistent and the approximate expected value and variance of r_L when the number of observations is large, are

$$E(r_L) \doteq \frac{E(A)}{E(B)}\left[1 + \frac{Var(B)}{(E(B))^2} - \frac{Cov(A,B)}{E(A)E(B)}\right]$$

$$= \rho + \frac{\rho\, Var(B) - Cov(A,B)}{(E(B))^2}, \qquad (2.14)$$

$$Var(r_L) \doteq \left[\frac{E(A)}{E(B)}\right]^2 \left[\frac{Var(A)}{(E(A))^2} + \frac{Var(B)}{(E(B))^2} - \frac{2\,Cov(A,B)}{E(A)E(B)}\right]$$

$$= \frac{1}{(E(B))^2}\,[Var(A) - 2\rho\, Cov(A,B) + \rho^2 Var(B)]. \qquad (2.15)$$

We have

$$Var(A) = Var\left(\sum^{n-1} w_i y_i\right) = \sum^{n-1} w_i^2 \sigma^2$$

$$Var(B) = Var\left(\sum^{n-1} w_i y_{i-1}\right) = \sum^{n-1} w_i^2 \sigma^2$$

$$Cov(A,B) = Cov\left(\sum^{n-1} w_i y_i, \sum^{n-1} w_i y_{i-1}\right) = \sum^{n-2} w_i w_{i+1} \sigma^2$$

$$E(B) = E(\sum^{n-1} w_i y_{i-1}) = E[\sum^{n-1} w_i(\alpha + \beta\rho^{i-1} + e_{i-1})]$$
$$= \beta \sum^{n-1} w_i \rho^{i-1}$$

the lower limit of i being one.

Hence from (2.14) and (2.15)

$$E(r_L) \doteq \rho + \frac{[\rho \sum^{n-1} w_i^2 - \sum^{n-2} w_i w_{i+1}]\sigma^2}{(\beta \sum^{n-1} w_i \rho^{i-1})^2}$$

$$Var(r_L) \doteq \frac{[(1 + \rho^2) \sum^{n-1} w_i^2 - 2\rho \sum^{n-2} w_i w_{i+1}]\sigma^2}{(\beta \sum^{n-1} w_i \rho^{i-1})^2} .$$

Since w_i are functions of ρ, no general expression can be found for w_i for which $Var(r_L)$ is minimum for all ρ. However, it is possible, by suitable choice of w_i, to obtain a linear estimator with minimum variance for some particular value of $\rho = \rho_o$. The variance of such estimator has been shown by Patterson and Lipton (27) to be

$$Var(r_L) = \sigma^2/(\beta^2 \sum_{1}^{n-1} w_i \rho^{i-1})$$

for w_i satisfying

$$(1 + \rho_o^2)w_i - \rho_o (w_{i+1} + w_{i-1}) = k_1(\rho_o^{i-1} + k_2)$$

where $w_o = w_n = 0$, k_1 is any convenient value such that w_i are not all zero and k_2 is such that $\Sigma x_i = 0$.

b. Quadratic estimators Consider the regression of y_{i+1} on $ky_i + hy_{i+1}$. The constants k and h can be so chosen that the regression coefficient is a function of ρ. The estimated regression coefficient then provides an estimate of

ρ. Examples are

$$y_{i+1} = \alpha(1 - \rho) + \rho y_i, \tag{2.16}$$

$$y_{i+1} = \frac{\alpha k(1-\rho)}{k + h\rho} + \frac{\rho}{k + h\rho}\,(ky_i + hy_{i+1}).$$

Quadratic estimates of ρ with minimum variance when $\rho = \rho_o$ are given by

$$r(\rho_o, k/h) = \frac{\Sigma w_i y_i}{\Sigma w_i y_{i-1}}, \qquad i = 1,\ldots,n-1, \tag{2.17}$$

with $w_i = \sum_j d_{ij}(ky_{i-1} + hy_i), \quad j = 1,\ldots,n-1$

where d_{ij} are so chosen that Var(r) is minimum at $\rho = \rho_o$. The ratio k/h specifies the form of a quadratic estimator. In terms of k, h and d_{ij}, an alternative form of the estimator (2.17) is

$$r(\rho_o,\ k/h) = \frac{\underline{y}_1' D(k\underline{y}_o + h\underline{y}_1)}{\underline{y}_o' D(k\underline{y}_o + h\underline{y}_1)} \tag{2.18}$$

where $\underline{y}_o' = (y_o, y_1, \ldots, y_{n-2})$

$\underline{y}_1' = (y_1, y_2, \ldots, y_{n-1})$,

and D is an (n-1) x (n-1) matrix defined as

$$D = V^{-1} = \frac{V^{-1}\underline{1}\underline{1}'V^{-1}}{\underline{1}'V^{-1}\underline{1}}$$

where $\underline{1}$ is an n-vector with all elements equal to unity, V^{-1} is the inverse of the matrix

$$V = \begin{pmatrix} 1+\rho_o^2 & -\rho_o & 0 & 0 & \cdots \\ -\rho_o & 1+\rho_o^2 & -\rho_o & 0 & \cdots \\ 0 & -\rho_o & 1+\rho_o^2 & -\rho_o & \cdots \\ 0 & 0 & -\rho_o & 1+\rho_o^2 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix}.$$

The expectation and variance of $r(\rho_o, k/h)$ are approximately

$$E(r) = \rho + \frac{\sigma^2}{\beta^2} \frac{(h-k\rho)\,tr(D) + (k-h\rho)tr(DU)}{(k + h\rho)F_o}$$

$$+ \frac{\sigma^2}{\beta^2} \frac{(h\rho^2 + 2k\rho - h)F_1 - 2kF_2}{(k + h\rho)F_o^2}$$

$$Var(r) = \frac{\sigma^2}{\beta^2} \frac{(1 + \rho^2)F_1 - 2\rho F_2}{F_o^2}$$

where $F_o = \sum_{o}^{n-1} W_i \rho^{i-1}$, $F_1 = \sum_{o}^{n-1} W_i^2$, $F_2 = \sum_{o}^{n-2} W_i W_{i-1}$, $W_i = \sum_{j=1}^{n-1} d_{ij}\rho^{j-1}$, and U is an (n-1) x (n-1) auxiliary identity matrix

$$U = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} .$$

The notation $r(\rho_o, k/h)$ is a very convenient form for the quadratic estimators with minimum variance at $\rho = \rho_o$. The estimator of ρ obtained from the simple regression (2.16) by Finney (7) is $r(0,1/0) = r(0,\infty)$, Hartley's internal least squares estimate of ρ is $r(1,1)$(26), and the Monroe estimator is $r(1,0)$(29). The class of internal regression estimators of ρ considered by White (36) is $r(1,k/h)$.

The bias and efficiency of some quadratic estimators have been examined to some extent. Patterson (26) noted that, for $n = 4$, $r(0,k/h)$ has very high efficiency over the whole range of ρ, and gives a rather better overall performance than the

linear estimators he proposed in 1956 (24). However, as n increases the quadratic estimators r(0,k/h) become inefficient in the useful range of ρ. The bias of r(0,1/0) is very large but can be considerably reduced by suitable choice of k/h. They are, as pointed out by Patterson, of only limited value since in practice choice of k/h is difficult unless n is small.

The estimators r(1,k/h), on the other hand, are more efficient and for very large n covers a very wide range of values of ρ^{n-1} with high efficiency. Patterson and Lipton (27) suggested the use of the Stevens' method to estimate ρ unless high-speed computing facilities are not available. Of the quadratic estimators, they showed that Hartley's estimator, r(1,1), generally leads to estimates with high efficiency and relatively low bias.

The small sample properties of some estimators of ρ have been reported by Johnston and Grandage (16), who conducted empirical study of the biases and efficiencies of the estimators r(1, 1.5), r(1, 1), r(1,0) and the Stevens' least squares estimator r. Their results showed that Stevens' r has in general a smaller bias than the quadratic estimates but is only slightly better than the r(1, 1.5) estimates. However, the least squares procedure failed to converge in some cases, particularly for large ρ and/or small β. The quadratic estimator r(1, 1.5) was found to have a smaller variance than the other estimators. The estimator r(1, 0) has largest bias, its efficiencies relative to Stevens' r

ranging from 0.1% to 159% with no discernible pattern.

The estimates of α and β can be obtained by the linear regressions of y_i on r_R^i. Patterson and Lipton (27) pointed out that the resulting fit depends on the magnitude of r_R-r, where r is the least squares estimate. If the errors and biases in r_R are small and the true relationship is the exponential model, a good fit can always be obtained. A sampling study by Tang (33) using the estimates of ρ obtained by r(1, 1.5), r(1,1), r(1,0) and the Stevens' r, showed that the four methods of estimating ρ gave similar results. The variances of the estimates of α and β obtained by simple least squares agree quite closely with the actual sample variances when ρ is small. However, when ρ is large, resulting in estimates of ρ close to 1, α and β are sometimes erratically estimated. These generally brought about very large values for sample variances for the estimates of α and β.

C. Interval Estimation

The literature on interval estimation for parameters in non-linear regression models is rather limited. Of the various approaches considered in the past, some lead to approximate confidence regions (1, 37), while others give rise to exact results under various assumptions (10, 12, 38).

Williams (38) considered the construction of interval estimate for the non-linear parameter γ in the non-linear

function

$$y = \alpha + \beta\, f(x,\gamma).$$

Assuming independence and homoscedasticity and denoting least squares estimates of α, β, γ by a, b, c, respectively, Williams expands $f(x,c)$ around c_o to linear term to get the simple linear regression function

$$\begin{aligned} y &\doteq a + b\,[f(x,c_o) + (\delta c)\, f'(x,c_o)] \\ &= a + b\, f(x,c_o) + b_1\, f'(x,c_o) \end{aligned}$$

where $b_1 = b(\delta c)$ and $f'(x,c_o) = df/dc\,|_{c_o}$. He then indicated that to get an interval estimate of γ, one tests the null hypothesis $\gamma = c_o$ by testing the significance of the estimated coefficient b_1 according to usual linear theory. Confidence limits are then defined as those values of c_o for which significance is attained at the level of probability adopted. Williams noted that the method can be readily extended to cases with more than one non-linear parameter, without specifying the class of non-linear function to which the method applies.

A method for determining confidence interval in non-linear models, proposed by Turner, Monroe and Lucas (34), is based on what they call "parabolic" alternative hypotheses. They expand $f(x,\theta)$ in a power series through the quadratic term, and carry out a F-test for the hypothesis that the coefficient of the quadratic term is zero. Confidence limits are then obtained as those trial values of the parameters for

which the F-tests are significant. This method is similar to the Williams' procedure in that confidence limits are based on results of hypotheses testing, although the hypotheses being tested are somewhat different.

Exploiting the fact that the total sum of squares can be partitioned into independent components and relating the partitions to the least squares equations, Halperin (10) showed that Williams' procedure is applicable to the class of non-linear models defined below.

Let

$$y = X\lambda + e$$

where X is n x m, $m < n$, λ is an m vector of unknown parameters, and e is an n-vector and $N(0,\sigma^2 I)$. The elements x_{ij} of X are non-linear functions of $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_r)$, $r + m < n$. Assume that X is of full rank for a non-degenerate region in λ, that each γ_j, $j = 1,2,\ldots,r$, enters one and only one column of X, and that γ_j appears only in the i-th column of X. Consider an n x r matrix D' of which the j-th row is

$$\left(\frac{\partial x_{1i}}{\partial \hat{\gamma}_j}, \frac{\partial x_{2i}}{\partial \hat{\gamma}_j}, \ldots, \frac{\partial x_{ni}}{\partial \hat{\gamma}_j}\right). \tag{2.19}$$

The class of non-linear models in question consists of all models for which the conditions above are satisfied and the matrix (X,D) is of full rank (m + r) for a non-degenerate region in λ and γ.

Writing

$$\theta' = (\lambda', \underline{0}') = (\theta_1', \theta_2')$$

where $\underline{0}'$ is a 1 x r row vector of zeros, we then have

$$y = (X,D)\theta + e.$$

By the principle of least squares the estimate of θ is

$$\hat{\theta} = \begin{pmatrix} X'X & X'D \\ D'X & D'D \end{pmatrix}^{-1} \begin{pmatrix} X' \\ D' \end{pmatrix} y$$

i.e. $\hat{\theta}_1 = (X'X)^{-1}X'y - (X'X)^{-1}(X'D)(U'U)^{-1}U'y$

$$\hat{\theta}_2 = (U'U)^{-1}U'y$$

where $U' = D'[I - X(X'X)^{-1}X']$.

We have the following interesting results.

(1) The statistic

$$F_{r+m,n-r-m} = \frac{(n-r-m)(\hat{\theta}-\theta)'(X,D)'(X,D)(\hat{\theta}-\theta)}{(r+m)[y'y-\hat{\theta}'(X,D)'(X,D)\hat{\theta}]}$$

is distributed as Snedecor's F with R+m and n-r-m degrees of freedom, and gives an exact confidence region for all parameters (λ',γ').

(2) The statistic

$$F_{r,n-r-m} = \frac{(n-r-m)\hat{\theta}_2'U'U\hat{\theta}_2}{r[y'y - \hat{\theta}'(X,D)'(X,D)\hat{\theta}]} \qquad (2.20)$$

is distributed as Snedecor's F with r, and n - r- m degrees of freedom. Since D is not a function of λ, the right hand of (2.20) is free of λ. Hence (2.20) provides a confidence region for γ only. A consequence arising from the construction of D is that, when γ take on the least squares values, $U'y = 0$, and hence the statistic (2.20) is zero. This

seems to be a reasonable and natural requirement for the confidence region for γ, and explains why (2.19) is taken as the j-th row of D'.

(3) Confidence regions for the linear parameters independent of the non-linear parameters do not appear obtainable because the ratio $F_{m,n-r-m}$ cannot possibly be free of γ.

Hartley (12) introduced a different approach, also based on partitioning the error components, to construct confidence regions. Under certain regularity conditions for the $f(x_i,\theta)$, he asserted, the multivariate normal distribution of the y_i of the model $y = X\theta + e$, will admit a set of m statistics jointly sufficient for θ if and only if $f(x_i,\theta)$ is 'essentially linear', i.e.

$$f(x_i,\theta) = \sum_{j=1}^{m} w_j(\theta)u_{ij}, \qquad (2.21)$$

where the $w_j(\theta)$ are continuous functions of θ_j and the n x m matrix $U = (u_{ij})$ has rank m and does not depend on θ. (Recall that in Halperin's procedure, the matrix (X,D) depends on θ.) Moreover, while in general $f(x_i,\theta)$ will not be representable in the form (2.21) it will usually be possible to represent the $f(x_i,\theta)$ at least approximately as m-term linear forms of parameter functions $w_i(\theta)$. By Cochran's theorem the quadratic forms

$$reg(e) = (U'e)'(U'U)^{-1}(U'e)$$

$$res(e) = e'e - reg(e),$$

the two components of the sum of squares e'e, are independently distributed as $\sigma^2\chi^2$ with m and n-m degrees of freedom. Hence

$$\frac{\text{reg}(y - f(x,\theta))}{\text{res}(y - f(x,\theta))} \leq \frac{m}{n-m} F(\gamma;m,n-m)$$

provides an exact $100\gamma\%$ confidence region for θ.

To illustrate the procedure, consider the exponential model (1.4)

$$f(x_i,\theta) = \alpha + \beta e^{-ax}$$

with $x_i = -\frac{1}{2}(n - 1),\ldots,0,\ldots,\frac{1}{2}(n-1)$ and n odd. By expanding e^{-ax} around zero we obtain

$$\begin{aligned} f(x_i,\theta) &= \alpha + \beta(1 - ax + \tfrac{1}{2}a^2x^2 - \ldots) \\ &\doteq -\beta ax + \beta a^2(\tfrac{1}{2}x^2) + (\alpha + \beta) \\ &= w_1u_{i1} + w_2u_{i2} + w_3u_{i3} \end{aligned}$$

where $w_1 = -\beta a$, $w_2 = \beta a^2$, $w_3 = \alpha + \beta$

$u_{i1} = x_i$, $u_{i2} = \frac{1}{2}x_i^2$, $u_{i3} = 1$.

Hence a joint confidence region for $\theta' = (\alpha, \beta, a)$ is

$$\begin{aligned} 0 \leq \Sigma(y - \alpha - \beta e^{-ax})^2 - (1 + c)[&n(\bar{y} - \alpha - \tfrac{\beta}{n}\Sigma e^{-ax})^2 \\ &+ (\Sigma yx - \beta\Sigma xe^{-ax})^2/\Sigma x^2 \\ &+ (\Sigma y\xi - \beta\Sigma\xi e^{-ax})^2/\Sigma\xi^2] \end{aligned}$$

where $c = (n - 3)/(3F(\gamma; 3,n-3))$

and $\xi_i = x_i^2 - (n^2 - 1)/12$

are the second degree orthogonal polynomials, and all the summations are over $i = 1,\ldots,n$.

If $f(x,\theta)$ is only moderately non-linear, and if an

independent estimate s^2 of σ^2 is available, Beale (1) suggests the approximate confidence region defined by

$$Q(\theta) - Q(\hat{\theta}) \leq ps^2F(\gamma;\ p,v) \qquad (2.22)$$

where s^2 has v degrees of freedom, $\hat{\theta}$ is the least squares estimate of θ, and $Q(\theta) = \Sigma\ (y - f(x,\theta))^2$. He also indicates that, although the estimate

$s^2 = Q(\hat{\theta})/(n-p)$, with $v = n-p$ degrees of freedom

is not an independent estimate of σ^2, its use is justified in essentially the same circumstances as the use of (2.22). The confidence region defined by (2.22) reduces to the standard form when $f(x,\theta)$ is linear. Further, as noted by Beale, the confidence region is still exact if θ is some non-linear function of a transformed set of parameters ϕ such that $f(x,\theta(\phi))$ is well-defined linear function of ϕ for all x. In other situations (2.22) does not define an exact confidence region but will often be an adequate approximation to one.

Wilks and Daly (37) have also suggested a method of constructing confidence regions which in essence is equivalent to Beale's method to the extent that a second-degree approximation is adequate.

We shall examine, for the case when x = 0, 1, 2, confidence intervals for ρ and β based on ratio estimators. From (2.5) let

$$u = y_2 - y_1, \qquad v = y_1 - y_o. \qquad (2.23)$$

When normality is assumed, u and v follow a bivariate normal with means

$$E(u) = U = \beta\rho(\rho-1)$$
$$E(v) = V = \beta(\rho-1) \tag{2.24}$$

and variances and covariance

$$\mathrm{Var}(u) = \mathrm{Var}(v) = 2\sigma^2$$
$$\mathrm{Cov}(u,v) = -\sigma^2. \tag{2.25}$$

Since $U/V = \rho$ and $V^2/(U - V) = \beta$, we see that

$$r_R = u/v$$

and

$$b_R = v^2/(u - v)$$

are ratio estimators for ρ and β.

According to Fieller's theorem (6) an exact confidence interval for r_R is given by

$$\rho^2(v^2-2t^2s^2) - 2\rho(uv + t^2s^2) + (u^2 - 2t^2s^2) \leq 0 \tag{2.26}$$

where s^2 is an independent estimate of σ^2, and t is the appropriate Student t value.

A conservative confidence interval of β based on b_R can be constructed as follows. If an independent estimate s^2 of σ^2 is available with n d.f., a confidence ellipse of (U,V) based on the observations u and v is

$$\begin{pmatrix} v - V \\ u - U \end{pmatrix}' \left(\frac{\mathrm{Var}\begin{pmatrix} u \\ v \end{pmatrix}}{\sigma^2} \right)^{-1} \begin{pmatrix} v - V \\ u - U \end{pmatrix} \leq 2s^2F(\gamma;2,n)$$

where

$$\left(\frac{\mathrm{Var}\binom{u}{v}}{\sigma^2}\right)^{-1} = \frac{1}{3}\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Writing $F = F(\gamma;\ 2,n)$, the boundary of the ellipse is

$$V^2 + UV + U^2 - (2v + u)V - (2u + v)U + (v^2 + uv + u^2 - 3s^2F) = 0. \tag{2.27}$$

The tangent to (2.27) at a point (v_o,u_o) is

$$v_oV + \tfrac{1}{2}(u_oV + v_oU) + u_oU - \tfrac{1}{2}(2v + u)(V + v_o) + \tfrac{1}{2}(2u + v)(U + u_o) + (v^2 + uv + u^2 - 3s^2F) = 0$$

which upon simplification becomes

$$[(v_o - v) + \tfrac{1}{2}(u_o - u)]V + [\tfrac{1}{2}(v_o - v) + (u_o - u)]U + [uv + (u - u_o - \tfrac{1}{2}v)u + (v - v_o - \tfrac{1}{2}u)v - 3s^2F)] = 0. \tag{2.28}$$

Next consider a family of β-curves defined by

$$\beta = V^2/(U - V), \quad \text{or } V^2 + \beta V - \beta U = 0 \tag{2.29}$$

whose tangent at the point (v_o,u_o) is

$$v_oV + \beta(V + v_o)/2 - \beta(U + u_o)/2 = 0$$

i.e. $$(v_o + \beta/2)V - \beta U/2 + \beta(v_o - u_o)/2 = 0. \tag{2.30}$$

Now among the family of β-curves (2.29), there are two to which the ellipse (2.27) is tangent. Let (v_o,u_o) be a point at which the ellipse (2.27) is tangent to a β-curve. It is clear then the two equations of tangents, (2.28) and (2.30), must be identical. Equating the like terms in (2.28) and (2.30) gives

$$v_o - v + \tfrac{1}{2}(u_o - u) = v_o + \beta/2$$

$$\tfrac{1}{2}(v_o - v) + u_o - u = -\beta/2$$

$$vu + (v - v_o - \tfrac{1}{2}u_o)v + (u - u_o - \tfrac{1}{2}v_o)u - 3s^2F = \beta(v_o - u_o)/2. \quad (2.31)$$

Hence $u_o = \beta + 2v + u$

$v_o = -3(\beta + v)$.

Eliminating u_o and v_o in (2.31) we find

$$2\beta^2 + (5v + u)\beta + 3(v^2 - s^2F) = 0. \quad (2.32)$$

If $(5v + u)^2 - 24(v^2 - s^2F) \geq 0$

i.e. $(v + u)^2 + 8vu \geq 24s^2F$

then the two roots β_1 and β_2 of the quadratic equation (2.32) are real and define a conservative confidence interval for β, i.e.

$$P(\beta_1 < \beta < \beta_2) \geq 1-\gamma.$$

D. Estimation of Asymptotic Values and Design Criterion for Known ρ

In this section we shall discuss how a priori knowledge of the non-linear parameter ρ of the model (2.5) may be used in planning an experiment. We begin with examining the least squares estimates of α and β and their variances and covariance. With the error assumption of (2.7), these estimates given ρ are

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{n\Sigma(\rho^x - R)^2} \begin{pmatrix} \Sigma\rho^{2x} & -\Sigma\rho^x \\ -\Sigma\rho^x & n \end{pmatrix} \begin{pmatrix} \Sigma y \\ \Sigma\rho^x y \end{pmatrix},$$

$$\mathrm{Var}\begin{pmatrix} a \\ b \end{pmatrix} = \frac{\sigma^2}{n\Sigma(\rho^x - R)^2}\begin{pmatrix} \Sigma\rho^{2x} & -\Sigma\rho^x \\ -\Sigma\rho^x & n \end{pmatrix},$$

where $R = \Sigma\rho^x/n$. If further $x = 0,1,\ldots,n-1$, then from geometrical progression

$$\Sigma\rho^x = \frac{1-\rho^n}{1-\rho},$$

$$\Sigma\rho^{2x} = \frac{1-\rho^{2n}}{1-\rho^2},$$

$$\Sigma(\rho^x - R)^2 = \Sigma\rho^{2x} - \frac{(\Sigma\rho^x)^2}{n}$$

$$= \frac{1-\rho^{2n}}{1-\rho^2} - \frac{(1-\rho^n)^2}{n(1-\rho)^2}.$$

It follows that

$$\mathrm{Var}(a) = \frac{\sigma^2\Sigma\rho^{2x}}{n\Sigma(\rho^x - R)^2} = \sigma^2\left(\frac{1-\rho^{2n}}{1-\rho^2}\right)\left[\frac{n(1-\rho^{2n})}{1-\rho^2} - \frac{(1-\rho^n)}{(1-\rho)^2}\right]^{-1}$$

$$= \sigma^2\left[n - \frac{(1+\rho)(1-\rho^n)}{(1-\rho)(1+\rho^n)}\right]^{-1}. \qquad (2.33)$$

Let $\rho = e^{2u}$ and recall that

$$\tanh(uv) = \frac{e^{uv} - e^{-uv}}{e^{uv} + e^{-uv}} = \frac{e^{2uv} - 1}{e^{2uv} + 1} = \frac{\rho^v - 1}{\rho^v + 1}$$

$$= \tanh\left(\tfrac{1}{2}v \log \rho\right) \qquad (2.34)$$

and $\coth(uv) = 1/\tanh(uv)$. (2.35)

From (2.33), (2.34) and (2.35) we have

$$\mathrm{Var}(a) = \sigma^2[n - \coth(\tfrac{1}{2}\log\rho)\tanh(\tfrac{n}{2}\log\rho)]^{-1} \qquad (2.36)$$

which can easily be computed, with the help of available mathematical tables, for any values of ρ and n. Similarly,

we have

$$\text{Var}(b) = \frac{\sigma^2}{\Sigma(\rho^x - R)^2}$$

$$= \sigma^2\left[\frac{1-\rho^{2n}}{1-\rho^2} - \frac{(1-\rho^n)^2}{n(1-\rho)^2}\right]^{-1},$$

$$\text{Cov}(a,b) = -\frac{\sigma^2 R}{\Sigma(\rho^x - R)^2}$$

$$= -\frac{\sigma^2(1-\rho^n)}{n(1-\rho)}\left[\frac{1-\rho^{2n}}{1-\rho^2} - \frac{(1-\rho^n)^2}{n(1-\rho)^2}\right]^{-1}$$

$$= \sigma^2\left[\frac{1-\rho^n}{1-\rho} - \frac{n(1+\rho^n)}{1+\rho}\right]^{-1}.$$

We note from (2.33) that Var(a) is a monotone increasing function of ρ for given n and a monotone decreasing function of n for given ρ. It is also seen that, while Var(a) and Cov(a,b) approach zero as $n \to \infty$, Var(b) has a limit of $\sigma^2(1-\rho^2)$.

Although the estimate of β is of certain interest itself, an experimenter working with the exponential model is usually far more interested in the asymptotic response α. Take a nitrogen fertilizer experiment, for instance. Within certain range of nitrogen levels the response of a crop has been hypothesized to follow the exponential model. To obtain high precision for the estimate of α, one may increase the value of n by including a very wide range of nitrogen levels in the experiment. This may not be the best way to conduct the experiment, but it certainly can be done because the

experimenter usually has full control over the levels of nitrogen to be applied.

Next, consider a fertility experiment in which a constant level of nitrogen is applied every year to a crop grown on the same piece of land. The yield of the crop may also tend to an asymptote (9, 25). The input variable x of the exponential model now takes the value 0 for the first year of the experiment, 1 for the second year, etc. To increase n in such an experiment means to extend the experiment over many years. While the experimenter prefers an estimate of the asymptote with high precision, he also wishes to obtain the information with minimum cost and time possible. Obviously, a compromise must be made. The experimenter would plan his experiment so that he may obtain an estimate of α with a minimum acceptable precision in a minimum length of time. Thus, he may calculate the value of Var(a) for any given n, or he may find the value of n that is needed to attain a certain precision in estimating α, by equating Var(a) to a desired value, and solving (2.36) for n. To simplify the procedure of determining n, values of $\mathrm{Var}(a)/\sigma^2$ are tabulated in Table 1 for $\rho = 0.1(0.1)0.9$ and $n = 4(1)10(2)30(5)50$.

The fact that ρ is known leads to a considerable gain in the precision of estimating α, as can be seen by comparing the values of $\mathrm{Var}(a)/\sigma^2$ in Table 1 with corresponding F_{aa} values

Table 1. $\mathrm{Var}(a)/\sigma^2$ for various values of ρ and n

n	ρ 0.1	.2	.3	.4	.5
4	.3600	.3992	.4603	.5608	.7391
5	.2647	.2856	.3173	.3685	.4583
6	.2093	.2222	.2412	.2713	.3234
7	.1731	.1818	.1944	.2139	.2471
8	.1475	.1538	.1628	.1764	.1991
9	.1286	.1333	.1400	.1500	.1663
10	.1139	.1176	.1228	.1304	.1427
12	.0928	.0952	.0986	.1034	.1111
14	.0783	.0800	.0824	.0857	.0909
16	.0677	.0690	.0707	.0732	.0769
18	.0596	.0606	.0619	.0638	.0667
20	.0533	.0541	.0551	.0566	.0588
22	.0481	.0488	.0496	.0508	.0526
24	.0439	.0444	.0452	.0462	.0476
26	.0404	.0408	.0414	.0423	.0435
28	.0373	.0377	.0383	.0390	.0400
30	.0347	.0351	.0355	.0361	.0370
35	.0296	.0299	.0302	.0306	.0313
40	.0258	.0260	.0262	.0265	.0270
45	.0228	.0230	.0232	.0234	.0238
50	.0205	.0206	.0208	.0210	.0213

	.6	.7	.8	.9
4	1.0895	1.8953	4.3399	18.3399
5	.6340	1.0373	2.2598	9.2599
6	.4243	.6552	1.3540	5.3542
7	.3108	.4554	.8925	3.3928
8	.2420	.3389	.6305	2.2975
9	.1969	.2650	.4695	1.6365
10	.1653	.2153	.3643	1.2132
12	.1247	.1541	.2405	.7306
14	.0999	.1189	.1737	.4820
16	.0833	.0964	.1335	.3401
18	.0714	.0810	.1073	.2526
20	.0625	.0697	.0892	.1953
22	.0556	.0612	.0762	.1561
24	.0500	.0545	.0663	.1281
26	.0455	.0492	.0586	.1075
28	.0417	.0448	.0525	.0918

Table 1 (Continued)

n	ρ .6	.7	.8	.9
30	.0385	.0411	.0476	.0797
35	.0323	.0341	.0385	.0591
40	.0278	.0291	.0323	.0464
45	.0244	.0254	.0278	.0380
50	.0217	.0226	.0244	.0321

tabulated By Stevens (32). Since ρ is generally unknown in practice, one has to rely on some "guessed" values of ρ for planning an experiment. It may seem therefore that the Stevens' F_{aa} values should be preferred because they allow for some error of the guesses. However, Box and Lucas (2) point out that in practical problems it will almost invariably be the case that some information of parameter is available. Such information, whether based on accumulated experience or sound theoretical expectation, must be put to good use in the planning and design of an experiment. If one has information as to the range of values in which the non-linear parameter lies, we suggest the use of the larger of the values of n calculated from the hypothesized range of ρ. Since for any fixed n the larger the value of ρ, the larger the magnitude of $\mathrm{Var}(a)/\sigma^2$, the choice of n can be made based only on the upper limit of the hypothesized range of without additional work.

In our discussion above, σ^2 has been the variance of

errors associated with individual observations. Since in real situations replications are practiced and the estimate of α is likely to be based on means rather than individual observations y_i, the variance of the estimate of α is obtained by dividing the quantity in (2.36) by the number of replicates. For some experiments there may be more than one source of random variation so that the error variance consists of several components. For example if the y_i are whole plot observations in a split plot experiment, then

$$\sigma^2 = \frac{1}{s}\sigma_b^2 + \sigma_a^2$$

where σ_a^2 is the whole plot error variance, σ_b^2 the sub-plot error variance and s the number of replicates. Hence we find

$$\text{Var}(a) = (\sigma_b^2/s + \sigma_a^2)[n - \coth(\tfrac{1}{2}\log\rho)\tanh(\tfrac{n}{2}\log\rho)]^{-1}. \tag{2.37}$$

If further for some constant K, the following relation exists

$$\sigma_a^2 = K\sigma_b^2$$

Then (2.37) becomes

$$\text{Var}(a) = (1/s + K)\sigma_b^2[n - \coth(\tfrac{1}{2}\log\rho)\tanh(\tfrac{n}{2}\log\rho)]^{-1}. \tag{2.38}$$

The use of (2.37) and (2.38) is obvious. If estimates of σ_a^2 and σ_b^2 are available, we can find from Equation 2.37 or 2.38 various combinations of n and s that will yield approximately the same value of Var(a). Thus in planning an experiment, one can choose to complete the experiment in a

shorter period with more replicates or in a longer period with less replicates, by varying the combination of n and s.

III. INTERVAL ESTIMATION FOR PARAMETERS IN A RESTRICTED SPACE

A. Interval Estimation for the Mean of a Normal Population

1. General procedure of interval estimation

Consider the construction of level $1-\gamma$ confidence intervals for a parameter θ based on a statistic T. (See for example Cramer (4, pp. 509-514) for more detailed discussion of the topic.) If the sampling distribution of T is $g(T;\theta)$, two functions

$$\underline{T} = L(\theta,\gamma) \quad \text{and } \bar{T} = U(\theta,\gamma) \tag{3.1}$$

can be obtained such that

$$P(L \leq T \leq U;\theta) = \int_L^U g(T;\theta)dt = 1-\gamma. \tag{3.2}$$

For monotone increasing L and U, the inverse functions

$$\underline{\theta} = U^{-1}(T,\gamma) \quad \text{and} \quad \bar{\theta} = L^{-1}(T,\gamma)$$

define a family of confidence intervals for θ at confidence level $1-\gamma$ such that

$$P(\underline{\theta} \leq \theta \leq \bar{\theta};\ \theta) = 1-\gamma.$$

If $L = -\infty$ for all θ, we have

$$P(T \leq U;\ \theta) = \int_{-\infty}^{U} g(T;\theta) = 1-\gamma, \tag{3.3}$$

and hence

$$\underline{\theta} = U^{-1}(T,\gamma) \tag{3.4}$$

defines a family of lower confidence bounds for θ at level $1-\gamma$, i.e.

$$P(\underline{\theta} \leq \theta;\ \theta) = 1-\gamma.$$

Upper confidence bounds can be similarly defined.

The choice of the statistic T and the functions L and U is arbitrary. For a given pair of L and U, the confidence intervals will be shortest if T is efficient. On the other hand, one may obtain confidence intervals with various properties by the choice of L and U. A convenient way to this end is to view confidence intervals as a family of level γ tests, and set L and U in (3.2) and (3.3) equal to the critical values of the corresponding tests. As an example, let A be the acceptance region of a level γ uniformly most powerful (UMP) test for testing H_0: $\theta = \theta_0$ against H_1: $\theta > \theta_0$. If the interval $(-\infty,U)$ of (3.3) coincides with A,U will be the critical value of the UMP test. The lower confidence bounds (3.4) are then level $1-\gamma$ uniformly most accurate confidence bounds, which minimize the probability that $\underline{\theta}$ is less than any $\theta < \theta_0$. In analogy with these bounds, a family of uniformly most powerful unbiased confidence intervals at confidence level $1-\gamma$ can be found by letting the interval (L,U) in (3.2) coincide with the acceptance region of a two-sided level γ uniformly most powerful unbiased (UMPU) test, i.e. by letting L and U be equal to the lower and upper critical values of corresponding UMPU tests.

2. A procedure for interval estimation for parameters in a restricted space

Assume that $a < \theta < b$ and that the least squares estimate $\hat{\theta}$ of θ is $N(\theta,1)$. We may require that $\underline{\theta} \geq a$ and $\bar{\theta} \leq b$.

Obviously the general procedure described above must be modified or it may lead to confidence intervals extending beyond, or lying completely outside, the interval (a,b). We shall examine the problem through the association of confidence interval and hypothesis testing.

Consider the following tests at significance level $\gamma = .05$

$$\text{UMP test for } H_0: \ \theta = \theta_0 = b \qquad H_1: \ \theta < b \tag{3.5}$$

$$\text{UMPU test for } H_0: \ \theta = \theta_0 = b - 1.96 \qquad H_1: \ \theta \neq b - 1.96. \tag{3.6}$$

The acceptance regions consist of all $\hat{\theta}$ on the intervals $(b - 1.645, \infty)$ and $(b - 3.92, b)$, respectively. Note that (3.5) is a one-sided test with a one-sided rejection region, while (3.6) is a two-sided test whose rejection region consists of two parts of equal size. For the null hypotheses $H_0: \ \theta = \theta_0$, $b - 1.96 < \theta_0 < b$, an arbitrary criterion is to use a two-sided test with rejection region divided into two unequal parts. For lack of a more suitable terminology, these tests will be referred to as interpolation tests if their critical values are determined as follows. First, one chooses lower critical values L by linear interpolation between the two points $P_1(b - 1.96, b - 3.92)$ and $P_2(b, b-1.645)$(the lower critical values corresponding to the tests (3.5) and (3.6)), the upper critical values U can be obtained by the relation

$$P(L \leq \theta \leq U; \theta_0) = .95.$$

The same argument applies to the case for $0 < \theta_0 < 1.96$. Next let L(b, .05) be the line $\theta = b$ for $\hat{\theta} > 8.355$, U(a,.05) be the line $\theta = 0$ for $\hat{\theta} < 1.645$, and for $a + 1.96 \leq \theta_0 \leq b-1.96$ let the interval (L,U) coincide with the acceptance region of the two-sided UMPU test for H_0: $\theta = \theta_0$. The functions L and U of (3.1) are now completely specified, and hence confidence intervals for θ are defined. This is in fact the procedure proposed by Wey (35), who has also demonstrated numerically that, for $b-1.96 = 8.05 < \theta_0 < 10 = b$, the interpolation test has greater power than the two-sided UMPU test if $\theta_0 > \theta_1$ and vice versa if $\theta_0 < \theta_1$.

As we have noted, the test criterion is quite arbitrary. For, the interpolation needs not be linear, nor the two limits of interpolation need be fixed. We shall examine some alternatives for defining test criteria and compare the confidence intervals obtained. For convenience we shall call the point P_1 the starting point and the point P_2 the end point of interpolation. Although we shall also use starting points other than (b-1.96, b-3.92), it seems reasonable to retain the end point (b, b-1.645) throughout our discussion as it is the critical value of the UMP test for testing H_0: $\theta = b$. Before we pursue this any further, however, we shall consider the following topic.

3. An approximate sampling distribution

Let the probability density function (p.d.f.) of θ be

$$f(\theta) = 1/K, \qquad 0 < \theta < K \tag{3.7}$$

where K is a known constant. For any given θ let $\hat{\theta}$ be the least squares estimate based on a random sample $(x_1, x_2, \ldots, x_n)$. Assume that the conditional p.d.f. of $\hat{\theta}$ given θ is

$$\begin{aligned} g(\hat{\theta};\ \theta) &= N(\theta,\ 1) \\ &= (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(\hat{\theta}-\theta)^2}, \quad -\infty < \hat{\theta} < \infty. \end{aligned} \tag{3.8}$$

It follows that the joint p.d.f. of θ and $\hat{\theta}$ is

$$f(\theta,\ \hat{\theta}) = (K\sqrt{2\pi})^{-1}\ e^{-\frac{1}{2}(\hat{\theta}-\theta)^2},\ 0 < \theta < K,\ -\infty < \hat{\theta} < \infty.$$

and the unconditional p.d.f. of $\hat{\theta}$ is

$$f(\hat{\theta}) = \int_0^K f(\theta,\ \hat{\theta})d\theta, \qquad -\infty < \hat{\theta} < \infty. \tag{3.9}$$

Because the function (3.9) does not integrate into a simple closed expression, we resort to finding $f(\hat{\theta})$ by numerical method.

Let K be divided into c intervals each of length K/c with midpoint θ_i', and let $p(\theta_i') = 1/c$. Since

$$\int_{\theta_i' - K/2c}^{\theta_i' + K/2c} (1/K)d\theta = 1/c,$$

we represent the p.d.f. (3.7) by the probability function (p.f.)

$$p(\theta_i') = 1/c, \quad \theta_i' = (i-\tfrac{1}{2})K/c, \quad i = 1,2,\ldots,c. \tag{3.10}$$

We also define the following approximation of conditional

p.d.f.

$$g(\hat{\theta};\ \theta_i' - K/2c < \theta < \theta_i' + K/2c) \doteq N(\theta_i',\ 1) \qquad (3.11)$$

Note that the accuracy of the approximation increases as c increases, and the limits of (3.10) and (3.11) as c approaches infinity are the two continuous p.d.f. (3.7) and (3.8), respectively.

To expediate numerical work, we consider the truncated normal p.d.f. $NT(\theta_i', 1)$ defined by

$$P(a' < \hat{\theta} < a'';\ \theta_i') = w \int_{a'}^{a''} N(\theta_i', 1)d\hat{\theta}, \quad a_0 < \hat{\theta} < a_m$$

where $a_0 = \theta_i' - 3.55, \qquad a' \geq a_0$

$a_m = \theta_i' + 3.55, \qquad a'' \leq a_m$

and $$1/w = \int_{a_0}^{a_m} N(\theta_i', 1)d\hat{\theta} \doteq 0.99962.$$

If we divide the interval (a_0, a_m) into m intervals (a_{j-1}, a_j), $j = 1,\ldots,m$, each of length $7.1/m$, the probability that $\hat{\theta}$ given θ_i' falling on the j-th interval is

$$P(a_{j-1} < \hat{\theta} < a_j;\theta_i') = \int_{a_{j-1}}^{a_j} NT(\theta_i',\ 1)d\hat{\theta}.$$

If the integer c is defined such that $K/c = 7.1/m$, the interval $(-3.5,\ K + 3.5)$ can be divided into a total of $n = (Km/7.1 + m-1)$ intervals $(b_{k-1},\ b_k)$, $k = 1,\ldots,n$, of length $7.1/m$. We then define an approximate sampling distribution of $\hat{\theta}$ by the following approximate unconditional

p.d.f. of $\hat{\theta}$

$$P_K(b_{k-1} < \hat{\theta} \leq b_k) = \sum_{i=1}^{c} P(b_{k-1} < \hat{\theta} \leq b_k; \theta_i'), \qquad (3.12)$$

$$k = 1,2,\ldots,n.$$

where the subscript K indicates that θ is uniform over the interval (0, K), and

$$P(b_{k-1} < \hat{\theta} \leq b_k; \theta_i') = (w/n) \int_{b_{k-1}}^{b_k} N(\theta_i', 1) d\hat{\theta}$$

$$\text{if } \begin{cases} b_{k-1} \geq \theta_i' - 3.55 \\ b_k \leq \theta_i' + 3.55 \end{cases}$$

$$= 0 \quad \text{otherwise.}$$

The p.d.f. (3.12) is symmetric about $\hat{\theta} = K/2$. The cumulative distribution based on (3.12) for $\hat{\theta} \geq K/2$, K = 10, is tabulated in Table 2.

It is interesting to note that if K > 7, then $P_K(b_{k-1} < \hat{\theta} \leq b_k)$ is constant for $3.5 \leq b_{k-1} < b_k \leq (K - 3.5)$. This simple fact enables the following short cut for tabulating the approximate p.d.f. of $\hat{\theta}$ for any K.

If we let K = 10, c = 100, m = 71 and $(b_k - b_{k-1}) = 0.1$, then the two tails of the p.d.f. (3.12) are as tabulated in Table 3. For $b_k = 3.6, 3.7, \ldots, 6.5$, it was found that

$$P_{10}(b_{k-1} < \hat{\theta} \leq b_k) = 0.01.$$

Let

$$\delta = 1/[0.7 + (0.01)(10)(K-7)] = 10/K$$

where the constant 0.7 is the total probability of the two

Table 2. Approximate probability of cumulative distribution function of $\hat{\theta}$ for K = 10

θ	P^a
5.0	0.50000
5.5	0.55000
6.0	0.60000
6.5	0.65000
7.0	0.69998
7.5	0.74982
8.0	0.79919
8.5	0.84712
9.0	0.89173
9.5	0.93029
10.0	0.96018
10.5	0.98029
11.0	0.99173
11.5	0.99712
12.0	0.99919
12.5	0.99982
13.0	0.99998
13.5	1.00000

[a] $P = P(\hat{\theta} \leq \theta)$.

Table 3. Approximate p.d.f. of $\hat{\theta}$, K = 10, for $\hat{\theta} < 3.5$ and $\hat{\theta} > 6.6$

b_k	$P_{10}(b_{k-1} < \hat{\theta} \leq b_k)$	b_k
-3.4	.00000090	13.5
-3.3	.00000210	13.4
-3.2	.00000390	13.3
-3.1	.00000630	13.2
-3.0	.00000950	13.1
-2.9	.00001401	13.0
-2.8	.00002001	12.9
-2.7	.00002791	12.8
-2.6	.00003831	12.7
-2.5	.00005202	12.6
-2.4	.00006953	12.5
-2.3	.00009204	12.4
-2.2	.00012035	12.3
-2.1	.00015596	12.2
-2.0	.00019998	12.1
-1.9	.00025410	12.0
-1.8	.00031982	11.9
-1.7	.00039885	11.8
-1.6	.00049299	11.7
-1.5	.00060403	11.6
-1.4	.00073368	11.5
-1.3	.00088354	11.4
-1.2	.00105500	11.3
-1.1	.00124927	11.2
-1.0	.00146726	11.1
-0.9	.00170935	11.0
-0.8	.00197545	10.9
-0.7	.00226526	10.8
-0.6	.00257758	10.7
-0.5	.00291081	10.6
-0.4	.00326294	10.5
-0.3	.00363118	10.4
-0.2	.00401252	10.3
-0.1	.00440357	10.2
0.0	.00480052	10.1
0.1	.00519948	10.0
0.2	.00559643	9.9
0.3	.00598747	9.8
0.4	.00636882	9.7
0.5	.00673706	9.6
0.6	.00708919	9.5
0.7	.00742242	9.4
0.8	.00773474	9.3

Table 3 (Continued)

b_k	$P_{10}(b_{k-1} < \hat{\theta} \leq b_k)$	b_k
0.9	.00802455	9.2
1.0	.00829065	9.1
1.1	.00853274	9.0
1.2	.00875072	8.9
1.3	.00894500	8.8
1.4	.00911646	8.7
1.5	.00926632	8.6
1.6	.00939597	8.5
1.7	.00950701	8.4
1.8	.00960115	8.3
1.9	.00968018	8.2
2.0	.00974590	8.1
2.1	.00980002	8.0
2.2	.00984404	7.9
2.3	.00987965	7.8
2.4	.00990796	7.7
2.5	.00993047	7.6
2.6	.00994798	7.5
2.7	.00996168	7.4
2.8	.00997209	7.3
2.9	.00997999	7.2
3.0	.00998599	7.1
3.1	.00999050	7.0
3.2	.00999370	6.9
3.3	.00999610	6.8
3.4	.00999790	6.7
3.5	.00999910	6.6

tails tabulated in Table 3. We then find for any $K > 7$

$$P_K(b_{k-1} < \hat{\theta} \leq b_k) = \delta P_{10}(b_{k-1} < \hat{\theta} \leq b_k)$$

$$\text{if } b_k \leq 3.5$$

$$= 0.01\delta \text{ if } 3.5 < b_k < K - 3.4$$

and for $b_k \geq K - 3.4$, $P_K(b_{k-1} < \hat{\theta} \leq b_k)$ are readily obtained by symmetry.

4. Expected length of confidence intervals

We shall examine the functions L and U for θ at a restricted space. First let us consider the following criterion for the comparison of confidence intervals associated with various L and U.

Assuming that θ is uniformly distributed over the interval (0, K) and that $\hat{\theta}$ given θ is $N(\theta,1)$. The unconditional p.d.f. of $\hat{\theta}$ is

$$f(\hat{\theta}) = \int_0^K f(\theta, \hat{\theta})d\theta$$

$$= \int_0^K (K\sqrt{2\pi})^{-1} e^{-\frac{1}{2}(\hat{\theta}-\theta)^2} d\theta, \quad -\infty < \hat{\theta} < \infty.$$

Let $(\underline{\theta},\bar{\theta})$ be a family of level $1-\gamma$ confidence intervals for θ based on $\hat{\theta}$, we may write $\lambda = (\bar{\theta}-\underline{\theta})$ so that

$$E(\lambda) = \int_{-\infty}^{\infty} \lambda f(\hat{\theta})d\hat{\theta}$$

defines the expected length of confidence intervals based on $\hat{\theta}$, over all θ. As no simple analytic results are possible, we shall evaluate $E(\lambda)$ numerically.

Recall the approximate sampling distribution $P(b_{k-1} < \hat{\theta} < b_k)$ (3.12). For any given pair of L and U, let

$$\lambda_k = \bar{\theta}_k - \underline{\theta}_k$$

$$= L^{-1}(\hat{\theta}_k,\gamma) - U^{-1}(\hat{\theta}_k,\gamma)$$

where $\hat{\theta}_k$ is the mid-point of the interval (b_{k-1},b_k). We then define two equations for approximate expected length of

confidence intervals for θ to be

$$E'(\lambda) = \sum_{k=1}^{n} \lambda_k P(b_{k-1} < \hat{\theta}_k < b_k),$$

$$E'(\lambda;\theta'') = \Sigma\lambda_k P(b_{k-1} < \hat{\theta}_k < b_k)/\Sigma P(b_{k-1} < \hat{\theta}_k < b_k),$$

where $\hat{\theta}_k > \theta'' - 1.96$. As defined in these equations, $E'(\lambda)$ is the approximate expected length of confidence intervals for all θ and all $\hat{\theta}$, while $E'(\lambda;\theta'')$ is that for θ based on $\hat{\theta} > \theta'' - 1.96$.

We now return to the problem of constructing confidence intervals for θ uniformly distributed over the interval (a,b). Without loss of generality, we let a = 0 and b = K. As the discussion is completely parallel for the choice of L for $\theta > K/2$ and that of U for $\theta < K/2$, it suffices to consider the L functions.

Let $L_1 = (\theta - 1.645)$ and $L_2 = (\theta - 1.96)$ to two $L(\theta,.05)$ functions corresponding to the one-sided UMP test and the two-sided UMPU test for H_0: $\theta = \theta_0$. In selecting L functions we require L to be bounded by L_1 and L_2. Among all possible L functions consider the family of L defined by

$$\hat{\theta} - (\theta-1.96) = \delta[\hat{\theta} + \theta + 1.96-2(K-1.96)]^q, \quad q > 0 \qquad (3.13)$$

where δ is such that $\hat{\theta} = (K - 1.645)$ when $\theta = K$. The family of curves belonging to (3.13) passes through the starting point $P_1(K-1.96, K-3.92)$ and for K = 10 and q = 1 (3.13) gives Wey's linear interpolation. For $q > 1$, $E'(\lambda)$ increases

as q increases. As a matter of fact, L approaches L_2 rather rapidly as q increases. Values of $E'(\lambda)$ for K = 10 and for several q's, together with the $E'(\lambda)$ based on L_2, are shown in Table 4.

Table 4. Comparison of expected length of confidence intervals for θ, $0 < \theta < 10$, with fixed starting point

L based on	$E'(\lambda)$
Equation 3.13 with q equal to	
1/2	3.3709
1	3.3859
2	3.4027
4	3.4176
10	3.4309
L_2[a]	3.4377

[a] $\lambda_k = 0$ for $\hat{\theta}_k \geq 11.96$.

To examine the relation between $E'(\lambda)$ and the starting point P_1, we let q = 1 and replace the quantity (K - 1.96) in (3.13) by θ'. We then find the following equation for linear interpolation with starting point $P_1(\theta', \theta'-1.96)$ and end point $P_2(K, K-1.645)$

$$\hat{\theta} - \theta + 1.96 = \delta(\hat{\theta} + \theta - 2\theta' + 1.96) \qquad (3.14)$$

where $\theta' \leq K/2$ and δ is such that $\hat{\theta} = (K-1.645)$ when $\theta = K$. Numerical results (Table 5) show that minimum $E'(\lambda)$ is

Table 5. Comparison of expected length of confidence intervals for θ, $0 < \theta < 10$, with various starting points

L based on	$E'(\lambda)$ $0 < \theta < 10$	$E'(\lambda; 5)$ $\theta < 10$
Equation 3.14 with θ' equal to		
5.00	3.3650	3.5213
5.50	3.3633	3.5200
5.60	3.3631	3.5199
5.65	3.3631	3.5199
5.70	3.3630	3.5198
5.72	3.3630	3.5199
5.74	3.3630	3.5198
5.76	3.3630	3.5198
5.78	3.3629	3.5198
5.80	3.3629	3.5198
5.82	3.3629	3.5198
5.84	3.3630	3.5199
5.86	3.3630	3.5198
5.88	3.3630	3.5198
5.90	3.3630	3.5198
5.95	3.3629	3.5198
6.00	3.3632	3.5200
6.30	3.3640	3.5206
7.00	3.3689	3.5241
8.04	3.3859	3.5363
L_2[a]	3.4377	3.5971

[a] $a_{\lambda_k} = 0$ for $\theta_k \geq 11.96$.

obtained approximately by setting θ' equal to the larger of the two quantities K - 4.2 and K/2.

If θ is not bounded below, we may for suitable choice of θ'' compare $E'(\lambda'\theta'')$ for various starting points $P_1(\theta', \theta'-1.96)$. It is clear from the results above that $E'(\lambda;\theta'')$ attains its minimum approximately at $\theta' = (K-4.2)$. $E'(\lambda;5)$ presented in Table 5 are for $\theta < 10$ and for various starting points θ'.

If $\hat{\theta}$ given θ is $N(\theta,\sigma^2)$, the previous discussion still holds if we let $K = (b-a)/\sigma$. When σ^2 is unknown but estimated by $\hat{\sigma}^2$, similar study can be made with t-distribution substituting for normal distribution. Although the numerical work is greatly increased because of the additional consideration of degrees of freedom, it is speculated that results would be similar to those obtained here.

B. Interval Estimation for ρ

1. Behavior of confidence intervals for ρ based on Fieller's theorem and modification

Let $x_1 = v$, $x_2 = u$, $X_1 = V$ and $X_2 = U$, where v, u, V and U are as defined in (2.23) and (2.24), and from (2.25) $\mathrm{Var}(x_1) = \mathrm{Var}(x_2) = 2\sigma^2$ and $\mathrm{Cov}(x_1,x_2) = -\sigma^2$. We have seen from (2.26) that a $1-\gamma$ level confidence interval for ρ based on the ratio estimate $r = x_2/x_1$ with assumption (2.24) consists of all values of ρ that satisfy

$$F(\rho) = (x_1^2 - C^2)\rho^2 - (2x_1x_2 + C^2)\rho + (x_2^2 - C^2) \leq 0 \quad (3.15)$$

where $C^2 = 2t^2s^2$ and if σ^2 is known, s^2 and t are replaced by σ^2 and an appropriate value of normal deviate. The

confidence interval is bounded by the roots of the equation

$$F(\rho) = 0 \tag{3.16}$$

i.e. by

$$\rho = \frac{2x_1x_2 + C^2 \pm \sqrt{(2x_1x_2 + C^2)^2 - 4(x_1^2 - C^2)(x_2^2 - C^2)}}{2(x_1^2 - C^2)}$$

$$= \frac{x_1x_2 + C^2/2 \pm \sqrt{x_1^2 + x_2^2 + x_1x_2 - 3C^2/4}}{x_1^2 - C^2} . \tag{3.17}$$

Fieller's theorem (6) assures that, if x_1 is significantly different from zero, the roots (3.17) are real.

To examine further the nature of the confidence intervals and their limits, we partition the x_1x_2-plane into four parts (see Figure 1):

S_1: $x_1^2 - C^2 > 0$, all x_2

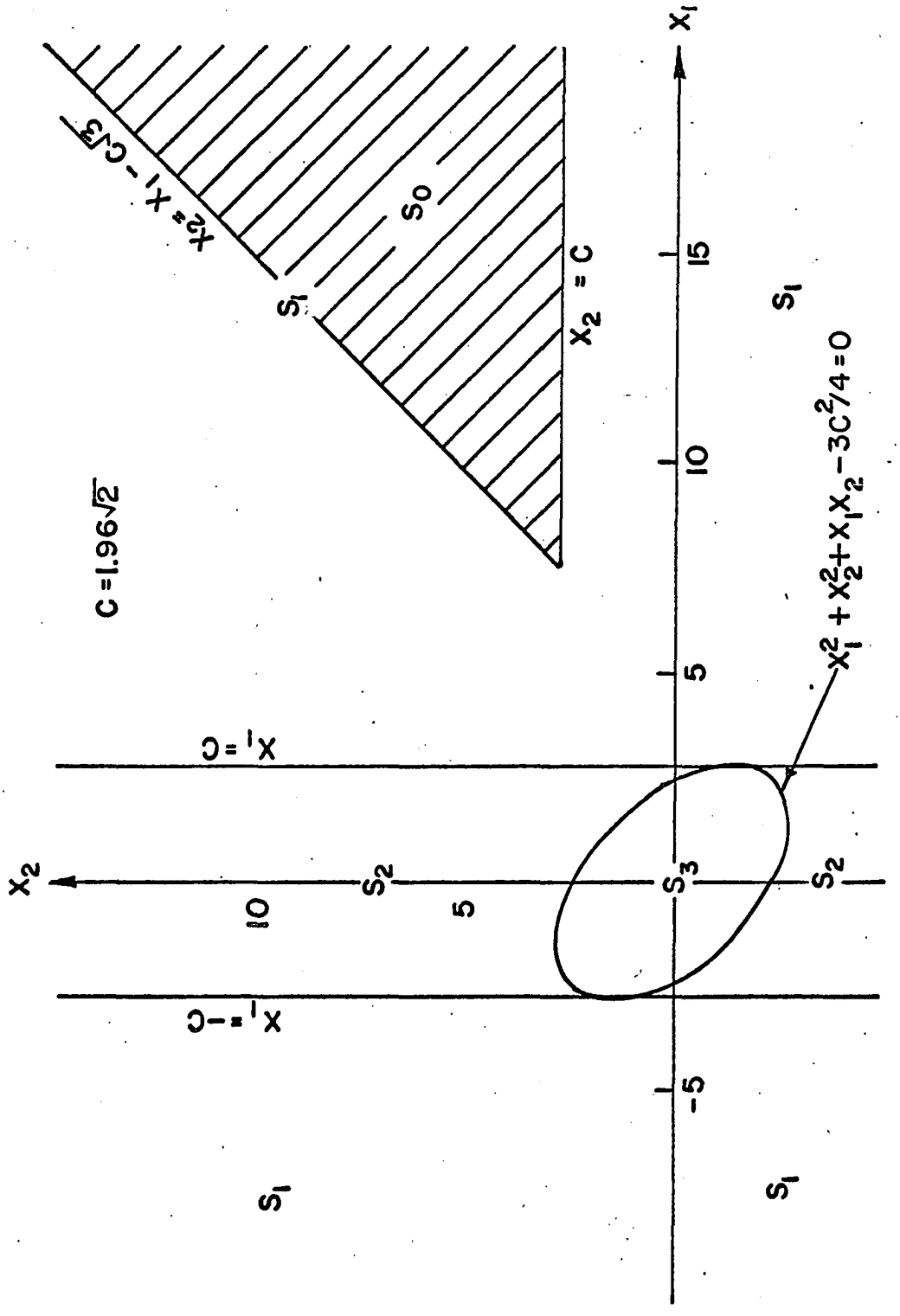
S_2: all x_1 and x_2 such that $x_1^2 - C^2 < 0$ and $x_1^2 + x_2^2 + x_1x_2 - 3C^2/4 \geq 0$

S_3: all sample points such that $x_1^2 + x_2^2 + x_1x_2 - 3C^2/4 < 0$

S_4: $x_1^2 - C^2 = 0$, with probability zero.

It is easily seen that S_1 consists of all sample points for which x_1 is significantly different from zero at γ level, and therefore the existence of real roots for (3.16) is assured by Fieller's theorem. Since the curve $F(\rho)$ for these points is concave upward, a $1-\gamma$ level confidence interval is well-defined and bounded by the two real roots.

Figure 1. Partition of the x_1x_2-plane


x_1
x_2
$x_2 = x_1 - c\sqrt{3}$
$x_2 = c$
S_0
S_1
S_2
S_3
S_1
S_1
S_1
S_2
S_1
$x_1 = c$
$x_1 = -c$
$c = 1.96\sqrt{2}$
$x_1^2 + x_2^2 + x_1 x_2 - 3c^2/4 = 0$
15
10
5
-5
10
5

For all sample points belonging to S_2, on the other hand, real roots exist for (3.16) but the curve $F(\rho)$ is concave downward, suggesting that the confidence intervals as defined by (3.15) consist of all values of ρ either less than the smaller or greater than the larger of the two roots. The curve $F(\rho)$ for sample points in S_3 is concave downward and lies completely below the x_1-axis. Therefore the confidence interval of ρ consists of all real numbers.

The confidence intervals thus constructed may include, or consist exclusively of, values of ρ lying outside the interval (0,1). This portion of the confidence interval provides no information on the parameter ρ since the parameter space of ρ is known to be the interval (0,1). We may, of course, modify the procedure by requiring that, as in the case for the mean from a normal distribution, $\underline{\rho} \geq 0$ and $\bar{\rho} \leq 1$, where $\underline{\rho}$ and $\bar{\rho}$ are, respectively, the lower and the upper confidence limits. With this restriction in mind, we propose the following rules for constructing a $1-\gamma$ level confidence interval based on $F(\rho) \leq 0$.

Let r_1 and r_2, $r_1 < r_2$, be the roots of $F(\rho) = 0$, and ϵ an arbitrary small positive number, then for any sample point $P = P(x_1,x_2)$, the confidence limits are determined as follows.

If P is in S_1

$$\underline{\rho} = \begin{cases} r_1 & \text{if } 0 \leq r_1 < 1 \\ 0 & \text{if } r_1 < 0 \\ 1-\epsilon & \text{if } r_1 \geq 1 \end{cases}$$

$$\bar{\rho} = \begin{cases} r_2 & \text{if } 0 < r_2 \leq 1 \\ 1 & \text{if } r_2 > 1 \\ \epsilon & \text{if } r_2 \leq 0 \end{cases}$$

If P is in S_2

$\underline{\rho} = r_2$ and $\bar{\rho} = 1$, if $0 \leq r_2 < 1$

$\underline{\rho} = 0$ and $\bar{\rho} = 1$, if $r_2 < 0$

$\underline{\rho} = 0$ and $\bar{\rho} = r_1$, if $0 < r_1 \leq 1$

$\underline{\rho} = 0$ and $\bar{\rho} = 1$, if $r_1 > 1$.

The confidence interval consists of two parts:

$(0,r_1)$ and $(r_2,1)$, if $r_1 > 0$ and $r_2 < 1$,

$(0,\epsilon)$ and $(1-\epsilon,1)$, if $r_1 < 0$ and $r_2 \geq 1$.

If P is in S_3, then $\underline{\rho} = 0$ and $\bar{\rho} = 1$.

If we let ϵ be infinitesimally small so that the probability that confidence intervals of the forms $(0,\epsilon)$ and $(1-\epsilon,1)$ will contain ρ, is also infinitesimal, the rules given above define exact confidence intervals at the confidence level chosen.

It can also be shown that, for all sample points (x_1,x_2), $x_2 > C$, bounded by the lines

$$x_2 = x_1 - C\sqrt{3} \qquad \text{and } x_2 = C$$

i.e. sample points that fall in the shaded area S_0 in Figure 1, we always have $\underline{\rho} \geq 0$ and $\bar{\rho} \leq 1$.

2. Graphical methods

We shall next examine the same problem with a different approach which will lead to a convenient graphical method for obtaining confidence intervals. Let

$$z = x_2 - x_1\rho,$$

then z is $N(0,2(1 + \rho + \rho^2)\sigma^2)$. If the acceptance region A for testing the hypothesis

$$H_0: \ Z = X_2 - X_1\rho_0 = 0 \text{ vs. } H_1: \ Z = X_2 - X_1\rho_0 \neq 0 \quad (3.18)$$

is bounded by the critical values $L(\rho_0,\gamma)$ and $U(\rho_0,\gamma)$, then for any given ρ, $0 \leq \rho \leq 1$,

$$P[L(\rho,\gamma) \leq z \leq U(\rho,\gamma);\rho] = 1-\gamma, \quad (3.19)$$

or, equivalently,

$$P[L(\rho,\gamma) + x_1\rho \leq x_2 \leq U(\rho,\gamma) + x_1\rho;\rho] = 1-\gamma. \quad (3.20)$$

On the x_1x_2-plane, A is the band bounded by the two lines

$$x_2 = L(\rho,\gamma) + x_1\rho, \quad (3.21)$$

$$x_2 = U(\rho,\gamma) + x_1\rho, \quad (3.22)$$

and the estimate of ρ can be obtained by substituting r for ρ in the equation

$$x_2 - x_1\rho = 0. \quad (3.23)$$

We observe that, for any sample point (x_{10}, x_{20}), a level $1-\gamma$ confidence interval is defined by

$$P(\underline{\rho} \leq \rho \leq \bar{\rho}) = 1-\gamma \quad (3.24)$$

where $\underline{\rho}$ and $\bar{\rho}$ are the solutions obtained from (3.22) and (3.21), respectively. To show that this is true, note that for any given ρ, if (x_{10}, x_{20}) lies in A, $\underline{\rho}$ and $\bar{\rho}$ defined above will contain ρ. Since from (3.20) probability that a point (x_{10}, x_{20}) given ρ lies in A is $1-\gamma$, the probability that the computed interval $(\underline{\rho}, \bar{\rho})$ will contain ρ is also $1-\gamma$. The argument holds for any L and U as long as the relation (3.19) is satisfied.

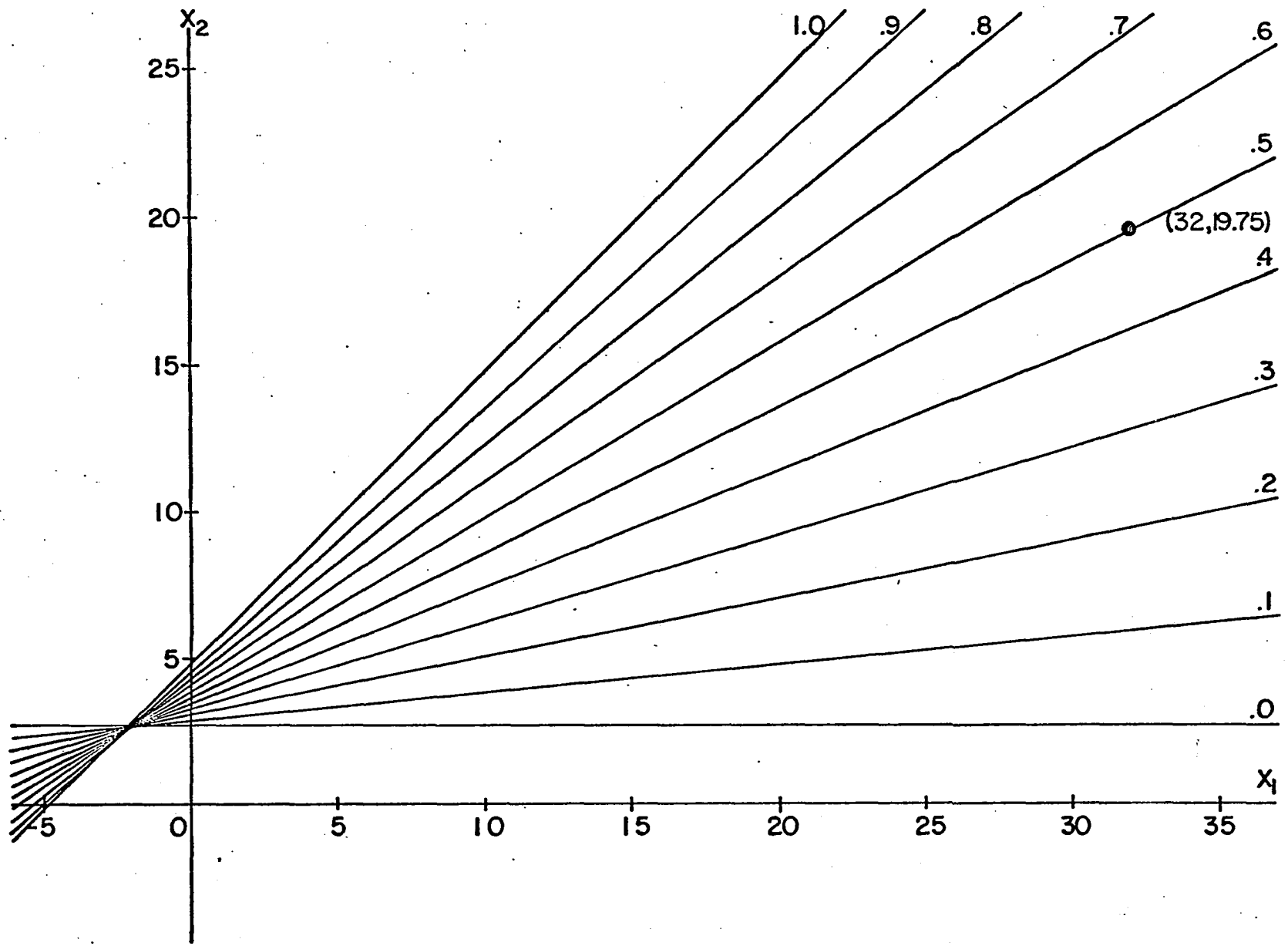
If we choose L and U based on a γ level two-sided UMPU tests, the confidence intervals defined by (3.24) are equivalent to those defined by (3.22). Plot the families of lines (3.21) and (3.22) on the x_1x_2-plane for ρ with L and U defined by the UMPU tests above, there will be one line from each family passing the sample point (x_{10}, x_{20}). The two values of ρ associated with the two lines are the confidence limits.

Let $\sigma^2 = 1$, then $\mathrm{Var}(x_1) = \mathrm{Var}(x_2) = 2$, $\mathrm{Cov}(x_1, x_2) = -1$ and $\mathrm{Var}(z) = 2(1 + \rho + \rho^2)$. For $\gamma = .05$ and L and U based on the two-sided UMPU tests, we find

$$U = -L = 1.96 \quad 2(1 + \rho + \rho^2).$$

In Figure 2 are plotted several lines $x_2 = U + x_1\rho$ belonging to the family (3.22). The numbers on the top and at the right of the graph indicate the values of ρ for which the U-lines are defined. If we observed that $x_1 = 32$ and $x_2 = 19.75$, then from Figure 2 we see that the U-line

Figure 2. U-lines for the determination of .95 level lower confidence limits


X2
25
20
15
10
5
1.0
.9
.8
.7
.6
.5
(32,19.75)
.4
.3
.2
.1
.0
X1
-5
0
5
10
15
20
25
30
35

defined by $\rho = .5$ passes the point $P_1(x_1,x_2)$ and hence $\underline{\rho} = 0.5$. Similarly we see from Figure 3 that $P_1(x_1,x_2)$ is on the L-line defined by $\rho = 0.75$, and hence $\bar{\rho} = 0.75$.

To present an alternative graphical method, we shall define a mapping from the x_1x_2-plane into the ρz-plane.

Let $z_0 = x_2$ and $z_1 = (x_2 - x_1)$. The line connecting the two points $P_0(0,z_0)$ and $P_1(1,z_1)$ is

$$z = z_0 + (z_1 - z_0)\rho,$$

i.e.

$$z = x_2 - x_1\rho. \tag{3.25}$$

If we plot the L and U-curves for the values of $L(\rho,\gamma)$ and $U(\rho,\gamma)$ obtained in (3.19), we can define a unique mapping from the lines on the x_1x_2-plane into points on the ρz-plane. (For $\gamma = .05$, these curves are shown in Figure 4.)

It is obvious that each line of (3.23) is mapped uniquely into one point on the ρ-axis. Because of the relation in (3.19) and (3.20) it is also clear that the correspondence between the L and U-curves and the families of lines (3.21) and (3.22) is one-to-one. We shall show that the abscissa of the point where (3.25) intersects the L-curve is the upper confidence limit $\bar{\rho}$ defined in (3.24).

Let (x_1,x_2) be a sample point such that

$$x_2 = L(\rho_0,\gamma) + x_1\rho_0 \tag{3.26}$$

for some $\rho = \rho_0$. On the ρz-plane let $P_0 = P_0(0,x_2)$ and $P_L(\rho_0,L(\rho_0,\gamma))$. The line connecting the points P_0 and P_L is

Figure 3. L-lines for the determination of .95 level upper confidence limits

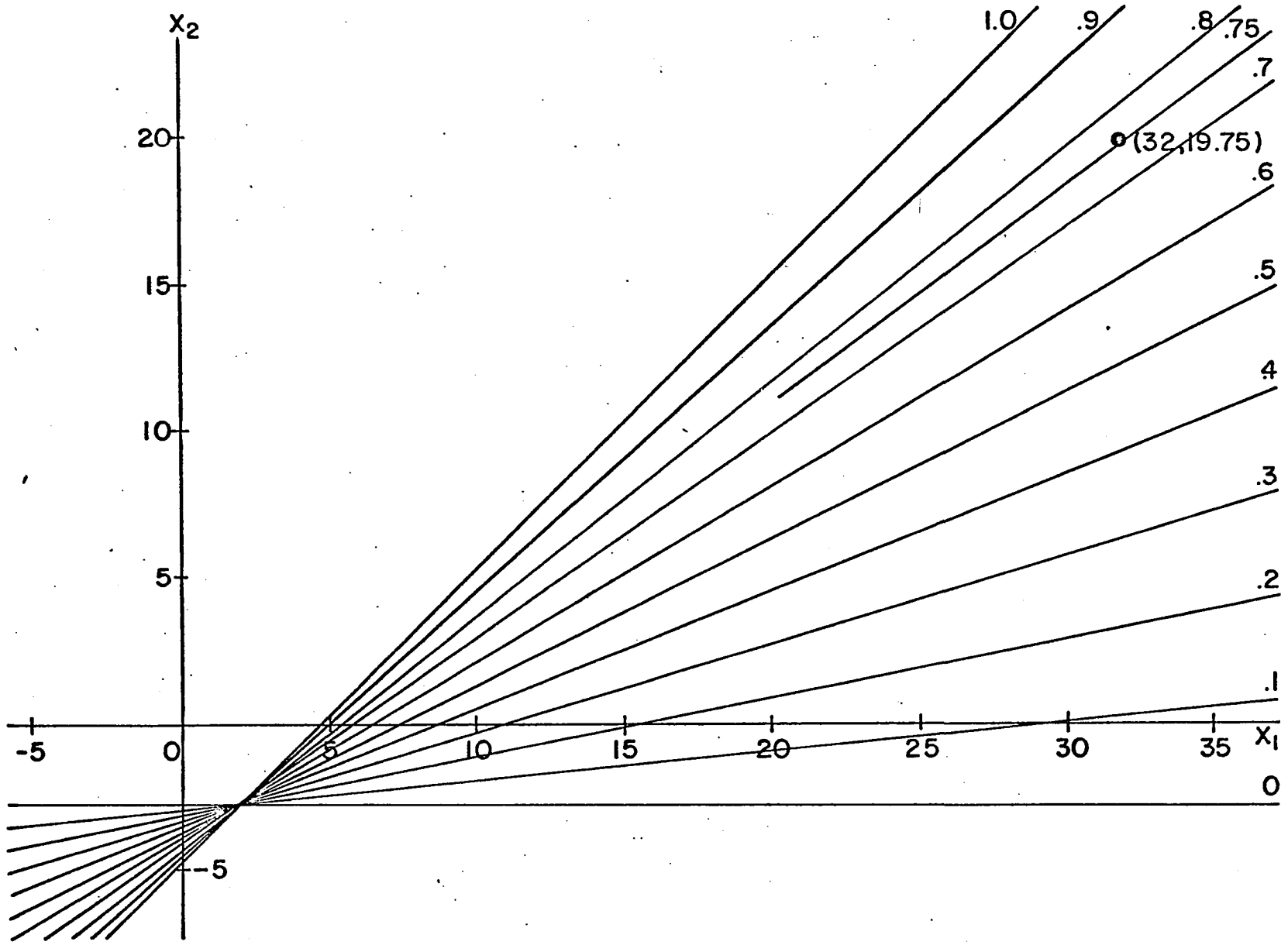
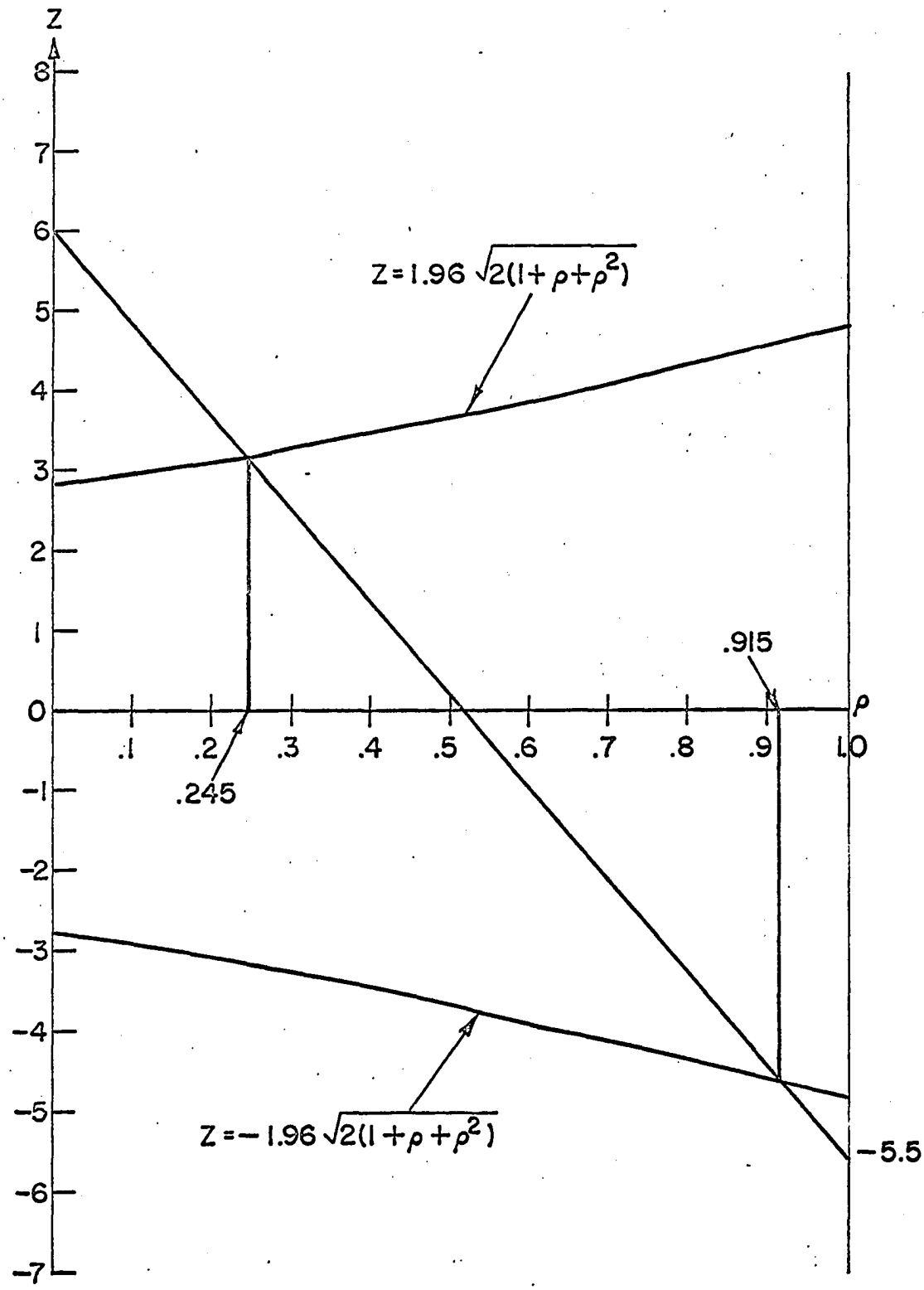

X2
20
15
10
5
-5
0
5
10
15
20
25
30
35
X1
0
-5
1.0
.9
.8
.75
.7
(32,19.75)
.6
.5
.4
.3
.2
.1

Figure 4. Graph for the determination of .95 level confidence intervals based on two-sided UMPU tests


Z
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
Z=1.96 √2(1+ρ+ρ²)
.915
ρ
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
.245
Z=-1.96 √2(1+ρ+ρ²)
-5.5

$$z = x_2 - \frac{\rho}{\rho_0}\,(x_2 - L(\rho_0,\gamma))$$

which intersects the line $\rho = 1$ at a point $P_1(1,z_1) = P_1(1,x_2-x_1')$ for some x_1'. But then we have

$$z = x_2 - \frac{1}{\rho_0}\,[x_2 - L(\rho_0,\gamma)]$$

$$= x_2 - x_1',$$

and hence

$$x_1' = \frac{1}{\rho_0}\,[x_2 - L(\rho_0,\gamma)]. \qquad (3.27)$$

From (3.26) and (3.27) we see that $x_1 = x_1'$.

Since x_2 is arbitrary, the argument holds for all x_1 and x_2 satisfying (3.26), and therefore the line (3.26) is mapped uniquely into one point on the ρz-plane, the point being the intersection of the L-curve and the line $z = x_2 - x_1\rho_0$. With a parallel discussion for the mapping of the line $x_2 = U(\rho_0,\gamma) + x_1\rho_0$, we conclude that, for any sample point (x_{10},x_{20}) on the x_1x_2-plane, the estimate r of ρ obtained from (3.23) is the intercept of the line (3.25) on the ρ-axis, and the confidence limits $\underline{\rho}$ and $\bar{\rho}$ obtained from (3.22) and (3.21) are the same as the projections of the point where the line (3.25) intersects the L and U-curves.

The graphical methods above give rise to confidence limits identical to those obtained directly from Fieller's theorem and therefore the rules described on page 57 for determining confidence limits also apply.

Let $U(0,\gamma)$ be the line $\rho = 0$ for $L(0,\gamma) \leq z_0 \leq U(0,\gamma)$,

and $L(1,\gamma)$ be the line $\rho = 1$ for $L(1,\gamma) \leq z_1 \leq U(1,\gamma)$. Next divide the z-axis into three segments denoted by

II-1: $z > U(0,\gamma)$

II-2: $L(0,\gamma) \leq z \leq U(0,\gamma)$

II-3: $z < L(0,\gamma)$

and similarly for the line $\rho = 1$ we have

I-1: $z > U(1,\gamma)$

I-2: $L(1,\gamma) \leq z \leq U(1,\gamma)$

I-3: $z < L(1,\gamma)$.

The rules on page 57 can now be simplified and are summarized in Table 6. In preparing the table we also adopted an additional rule that we change the signs of x_1 and x_2 whenever x_1 is less than zero. This is justified by the fact that the function $F(\rho)$ is symmetric about the origin of the x_1x_2-plane.

For $\gamma = .05$ and $\sigma^2 = 1$, theL and U-curves are shown in Figure 4. To use the graph we first locate the point $z = x_2$ on the z-axis, then locate the point $z_1 = x_2 - x_1$ on the line $\rho = 1$ (or one may find the point $z = x_2$, move across to the line $\rho = 1$, and measure <u>downward</u> a distance of x_1 to locate the point z_1). As an example, if $x_1 = 11.5$ and $x_2 = 6$, we see from Figure 4 that $\underline{\rho} = .245$ and $\bar{\rho} = .915$.

Table 6. Rules for the determination of confidence limits--based on Fieller's theorem

x_2 is in	x_1 is in	C.I. $(\underline{\rho}, \bar{\rho})$	
II-1	I-1	$(1-\epsilon,1)$	
II-1	I-2	$(r,1)$	
II-1	I-3	(r_1,r_2)	
II-2	I-2	$(0,r_1)$ and $(r_2,1)$	if the line $z = x_2 - x_1\rho$ intersects the L-curve at two points
		$(0,1)$	otherwise
II-2	I-3	$(0,r)$	
II-3	I-2	$(r,1)$	
II-3	I-3	$(0,\epsilon)$	

3. Confidence intervals based on interpolation tests and a criterion for the choice of confidence intervals

As in section A we shall now construct level $1-\gamma$ confidence intervals for ρ by the use of interpolation tests. Let $\gamma = .05$, we first consider the following tests:

$$H_0: \ Z = X_2 - \tfrac{1}{2} X_1 = 0, \qquad H_1: \ Z \neq 0; \tag{3.28}$$

$$H_0: \ Z = X_2 = 0, \qquad H_1: \ Z > 0; \tag{3.29}$$

$$H_0: \ Z = X_2 - X_1 = 0, \qquad H_1: \ Z < 0. \tag{3.30}$$

Note that (3.28) is identical to the test (3.18) for $\rho_0 = \frac{1}{2}$. Its acceptance region is bounded on the x_1x_2-plane by (3.21) and (3.22) for $\rho = \frac{1}{2}$. The test (3.29) is equivalent to $H_0: \ \rho = 0$ vs. $H_1: \ \rho > 0, \ \beta < 0$. A UMP test exists and its acceptance region is bounded above on the x_1x_2-plane by

$$x_2 = 1.645\sqrt{2(1+\rho+\rho^2)} + x_1\rho$$
$$= 1.645\sqrt{2}$$

since $\rho = 0$. Similarly, (3.30) is equivalent to H_0: $\rho = 1$ vs. H_1: $\rho < 1$, $\beta < 0$ with an acceptance region bounded below by $x_2 = -1.645\sqrt{6}$.

We now proceed to construct the interpolation tests. On the ρz-plane, the two upper critical values for the tests (3.29) and (3.28) are the two points $(0, 1.645\sqrt{2})$ and $(.5, 1.96\sqrt{3.5})$, respectively. The straight line connecting these two points is

$$z = 2.326 + 2.682\rho. \tag{3.31}$$

For $0 < \rho_0 < .5$, let the upper critical value for the test H_0: $Z = 0$ vs. H_1: $Z \neq 0$ be the point (z_0, ρ_0), where z_0 is obtained from (3.31) by setting $\rho = \rho_0$. Hence the U-function for $0 < \rho \leq .5$ may be defined by (3.31) and its corresponding L-function can be obtained by integration. Likewise the straight line that connects the two lower critical points $(.5, -1.96\sqrt{3.5})$ and $(1, -1.645\sqrt{6})$ defines for $.5 < \rho < 1$ an L-function

$$z = -3.305 - 0.724\rho$$

with its corresponding U-function obtained by integration. Finally, by letting $U(0, .05)$ be the line $\rho = 0$ for $z_0 = x_2 < 1.645\sqrt{2}$, and $L(1, .05)$ be the line $\rho = 1$ for $z_1 = x_2 - x_1 > -1.645\sqrt{6}$, the L and U-functions for the construction of confidence interval are completely specified. These

functions are plotted in Figure 5 marked with at = ∞.

Since our earlier discussion of graphical methods for constructing confidence intervals for ρ does not depend on the form of the L and U-functions, we may apply the same method to Figure 5. However, we observe that the L and U-curves in Figure 5 are not symmetric. The asymmetry results in requiring an additional assumption on the parameter β, which, fortunately, is a very mild one. While the graphical method based on Figure 4 is valid for all finite values of β, the method applied to Figure 5 requires that $\beta < 0$, or equivalently, $X_1 > 0$ and $X_2 > 0$. For β greater than zero, the method can be used if we <u>always</u> change the signs of x_1 and x_2. With the assumption $\beta < 0$, the rules for determining confidence limits for ρ are summarized in the following.

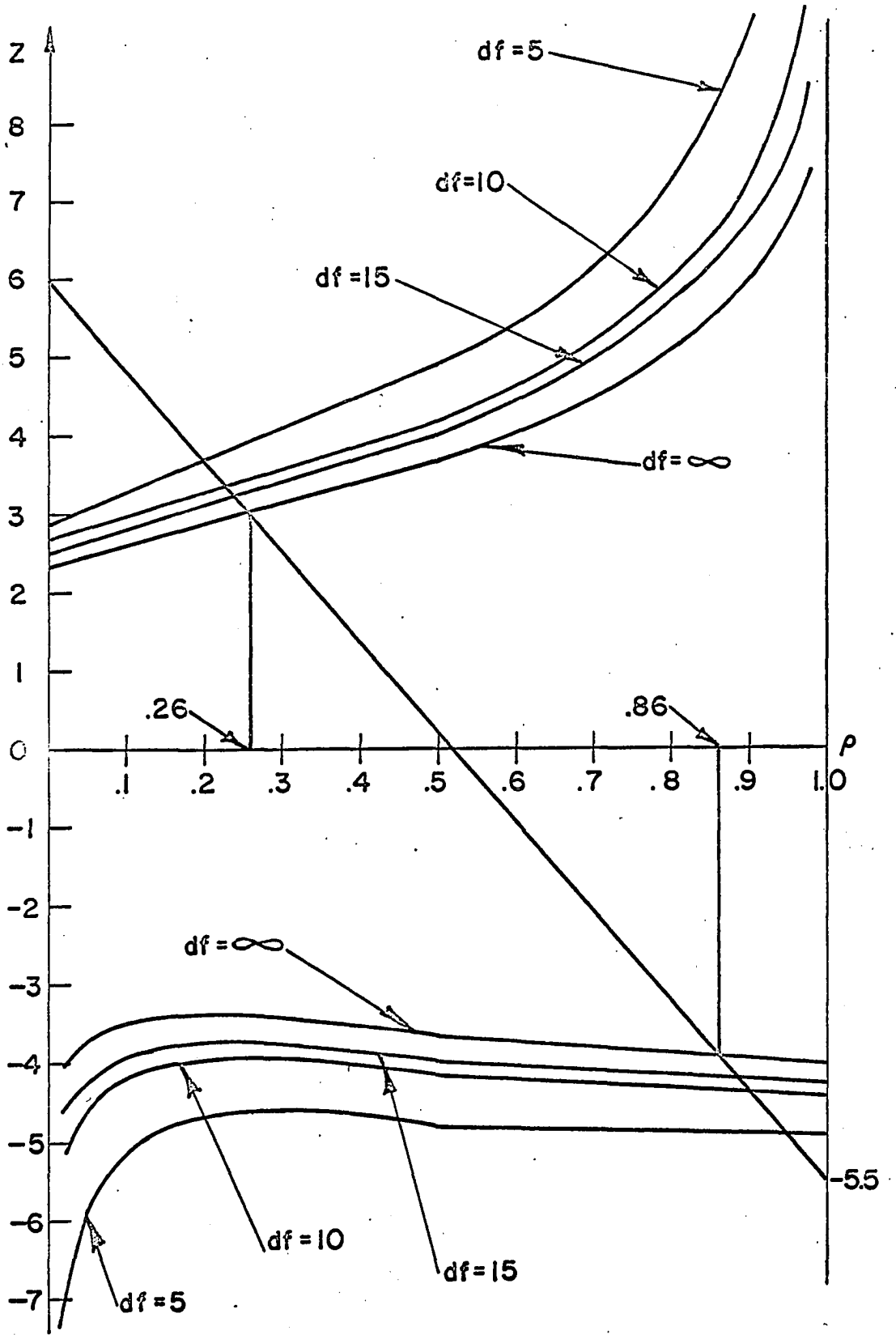
Let r be the abscissa of the point where the straight line

$$z = x_2 - x_1\rho \tag{3.25}$$

intersects either L or U-curve, and r_1 and r_2, $r_1 < r_2$, be the abscissas of the two points where the line (3.25) intersects L at two points, U at two points, or L and U each at one point, then the .95 confidence interval for ρ determined from Figure 5 is

1. (r_1,r_2) if (3.25) intersects L and U each at one point;
2. $(0,r)$ if (3.25) intersects L at one point;

Figure 5. Graph for the determination of .95 level confidence intervals based on interpolation tests


Z
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
df = 5
df = 10
df = 15
df = ∞
.26
.86
ρ
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
df = ∞
df = 10
df = 15
df = 5
-5.5

3. $(r,1)$ if (3.25) intersects U at one point;
4. $(0,r_1)$ and $(r_2, 1)$ if (3.25)intersects either L at two points or U at two points;
5. $(0,1)$ if (3.25) intersects L and U each at two points or no intersections exist.

For the example that $(x_1,x_2) = (11.5, 6.0)$, we see from Figure 5 that $\underline{\rho} = .26$ and $\bar{\rho} = .86$.

It is interesting to note that if $\rho \leq .5 \leq \bar{\rho}$, a confidence interval obtained based on the interpolation tests is always shorter than one based on two-sided UMPU tests. However, the same statement may or may not be true if either $\bar{\rho} < .5$ or $\underline{\rho} > .5$.

To compare the results obtained with Figures 4 and 5, we compute the unweighted average length of confidence intervals for a total of 5545 sample points (x_1,x_2) in the shaded area S_0 of Figure 1. These sample points consist of $x_1 = 8.0(0.5)60.0$ and $x_2 = 3.0(0.5)x_2'$, where x_2' is the largest value of x_2 that satisfies $x_2 < x_1 - C\sqrt{3}$ for any given x_1. The confidence limits were computed by numerical method, giving a much better precision than one would have by obtaining the limits directly from the Figures. The results (Table 7) show that for x_1 not too large, the gain by the interpolation tests is considerable.

4. Generalization of the procedure

We shall consider the use of the graphical method in some more general situations.

Table 7. Average length of confidence intervals for $8 \leq x_1 \leq 60$, $C < x_2 < x_1 - C\sqrt{3}$, where $C = 1.96\sqrt{2}$

x_1	Number of Sample Points	Average length based on (1) Interpolation tests	Average length based on (2) Two-sided UMPU tests	$\frac{(1)}{(2)} \times 100$
$8 \leq x_1 \leq 10$	15	.7010	.7979	87.86
$10 < x_1 \leq 15$	105	.5148	.5615	91.68
$15 < x_1 \leq 20$	205	.3813	.4073	93.62
$20 < x_1 \leq 25$	305	.3034	.3189	95.14
$25 < x_1 \leq 30$	405	.2517	.2621	96.02
$30 < x_1 \leq 35$	505	.2152	.2225	96.71
$35 < x_1 \leq 40$	605	.1881	.1934	97.27
$40 < x_1 \leq 45$	705	.1671	.1710	97.74
$45 < x_1 \leq 50$	805	.1504	.1532	98.14
$50 < x_1 \leq 55$	905	.1367	.1389	98.48
$55 < x_1 \leq 60$	1005	.1254	.1269	98.79

If $\sigma^2 \neq 1$, the same graphs in Figures 4 and 5 can be used if the observed values of x_1 and x_2 are divided by the known standard deviation σ. If σ^2 is unknown but an independent estimate s^2 of σ^2 with df degrees of freedom is available, a set of L and U-function can be obtained by setting $\sigma^2 = 1$ and replacing the normal deviates, 1.96 and 1.645 for example, by appropriate Student t values. In determining confidence limits x_1 and x_2 are first divided by s.

For $n > 3$, let a linear ratio estimator of ρ be $r = x_2/x_1$, where

$$x_1 = \sum_{1}^{n-1} w_i y_{i-1} \qquad \text{and} \qquad x_2 = \sum_{1}^{n-1} w_i y_i$$

and w_i are independent of y_i such that $\sum_{1}^{n-1} w_i = 0$. With error assumption defined in (2.7) and normality, x_1 is normal with mean $X_1 = \beta \sum_{1}^{n-1} w_i \rho^{i-1}$ and variance $\mathrm{Var}(x_1) = \sigma^2 \sum_{1}^{n-1} w_i^2$, x_2 is normal with mean $X_2 = \beta \sum_{1}^{n-1} w_i \rho^i$ and variance $\mathrm{Var}(x_2) = \sigma^2 \sum_{1}^{n-1} w_i^2$. We also have

$$\mathrm{Cov}(x_1, x_2) = \sigma^2 \sum_{1}^{n-2} w_i w_{i+1} .$$

It follows that $z = x_2 - x_1\rho$ is normal with mean

$$Z = X_2 - X_1\rho = 0$$

and variance

$$\mathrm{Var}(z) = \sigma^2[W_1(1+\rho^2) - 2\rho W_2]$$

where $W_1 = \sum_{1}^{n-1} w_i^2$ and $W_2 = \sum_{1}^{n-2} w_i w_{i+1}$. It is obvious that all previous discussion for the construction of confidence intervals holds for the more general case. No further elaboration is therefore necessary.

Finally, assume that the parameter space of ρ is the interval $(0,K)$, $K \neq 1$. Our procedure still applies if we make the following modifications.

(1) In Figures 4 and 5, replace the line $\rho = 1$ by the line $\rho = K$ and let $z_1 = x_2 - Kx_1$. The two points $P_0(0,z_0) = P_0(0,x_2)$ and $P_1(K,z_1) = P_1(K,x_2-Kx_1)$ will determine the

required straight line (3.42)

$$z = x_2 - x_1\rho$$

which is then used to determine the confidence limits.

(2) In all rules concerning the determination of confidence limits, replace the value 1 by K.

Although the parameter space (0,K) for K > 1 is irrelevant for the parameter ρ of the exponential model, it is of great interest for some other ratio estimators. Its application is seen in areas such as bioassay.

IV. ANALYSIS OF ROTATION EXPERIMENTS

A. Principles of Rotation Experiments

In agricultural practice a class of cropping systems known as crop rotation is characterized by growing in cyclic sequence several crops on the same land. Experiments which are conducted to investigate the characteristics of various rotations are referred to as rotation experiments. Agronomically, the main objective of a rotation is to provide natural control over one or more of the factors such as weeds, pests, diseases and soil fertility. Rotation experiments have therefore been designed to evaluate, in terms of crop yields, the effectiveness of various rotation systems.

In the literature, rotation experiments are often divided into two types:

(a) experiments comparing the effects of different treatments on the crops of a rotation, and

(b) experiments comparing the effects of different rotations. An experiment of type (a) is also known as a fixed-rotation experiment because only one rotation is involved. The problem thus is to study the treatment effects of certain fertilization or cultivation practices. Patterson (23) has reported a fixed-rotation experiment in which four manurial treatments were applied to a barley-sugar beet-potatoes rotation over a period of 19 years. In experiments of type (b), several rotations are usually

included in one experiment, and one major objective is to compare the yield of the same crop in different rotations. A well-known example is the rice-pasture experiment discussed by Yates (40). Many rotation experiments are, however, a combination of types (a) and (b), namely, the treatments consist of different fertilization or cultivation practices applied to more than one rotation. Most of the rotation experiments conducted by Iowa Agricultural Experiment Station, including the one reported by Fuller and Cady (9), belong to this category.

Short-term, relatively simple rotation experiments are possible if information is needed only for some restricted area. Rotation experiments of more general nature are mostly complex long-term experiments. One factor affecting the duration of a rotation experiment is the length of rotation or cropping cycle. Assuming one crop per year, the length of the rotation corn-oats (C-O) is two years and that of the rotation corn-corn-oats-meadow-meadow-meadow, six years. Continuous corn can be regarded as a special case of rotation with length one. A rotation experiment may be defined by the cyclic order of the crops included. For convenience, one of the crops is usually assigned to head the cycle and hence a basic rotation is specified. For any rotation, phase denotes the position of a particular crop in the basic rotation. In the rotation C-O, for example, corn

is of phase one, and oats of phase two.

Basically there are two ways to start a rotation experiment. One may start a rotation experiment by including all phases of all rotations in the first year of experiment and follow the cyclic sequences in the subsequent years. Alternatively one may start with a set of all basic rotations in the first year, and add a new basic rotation for each rotation system in each of the subsequent years till all phases of the rotation are represented. When the latter is adopted, an assumption is that the soil fertility can be maintained at a fixed level by appropriate handling of the experimental field so that the basic rotations started in different years are subject to the same initial soil fertility.

As an example, let a rotation experiment consist of two rotations, C_1-C_2-O and C_3-C_4-O-M. If all phases are included in the first year of the experiment, the crops that appear on the plots in the first four years are

	Plot						
Year	1	2	3	4	5	6	7
1	C_1	C_2	O	C_3	C_4	O	M
2	C_2	O	C_1	C_4	O	M	C_3
3	O	C_1	C_2	O	M	C_3	C_4
4	C_1	C_2	O	M	C_3	C_4	O

If the experiment is initiated with a set of all basic rotations, we then have

Year	Plot 1	2	3	4	5	6	7
1	C_1			C_3			
2	C_2	C_1		C_4	C_3		
3	O	C_2	C_1	O	C_4	C_3	
4	C_1	O	C_2	M	O	C_4	C_3

The concept of cycle is a useful one in describing rotation effects. Consider any rotation system with a certain crop. Regardless of the condition of the initial soil fertility, if the same cultivation is practiced, the rotation effect will tend to stabilize in the long run. In terms of yield this means that the crop yield will approach a limiting value as the cycles advance. The change in yield from cycle to cycle has long been used to measure rotation effects. Cochran (3) has demonstrated the use of linear and quadratic terms to represent the trend of yield changes. The same approach is still being used (25). However, we prefer to look upon the change of crop yield from cycle to cycle as the realization of a growth process, with the rate of change depending upon the rate at which a rotation system improves or deteriorates the soil fertility till a theoretical equilibrium is finally reached.

As our measurement of rotation effects will be based on the observed change in yields from cycle to cycle, it is desirable to know when a rotation starts its first cycle. Different definitions of the first cycle can be obtained

depending on the way an experiment is started, the nature of rotations and the interpretation of cycles. Cycle 1 of a test crop can be defined as the year following the completion of a full term of the non-test crops. Thus in a corn-soybean-corn-oats basic rotation, cycle 1 of corn after oats occurs in the fifth year and cycle 1 of corn after soybean in the seventh year of the experiment. The corn in the first and third years of the experiment are of cycle zero. If an experiment is initiated with a set of all basic rotations, cycle 1 can equally well be defined as the first time a test appears.

A fundamental principle in designing a rotation experiment is that all phases of all rotations should occur in each year. Following this principle the number of plots required for a complete replicate is then equal to the sum of the number of phases of all rotations. If complete blocks are used, the experiment is said to have a basic design. Unfortunately, a basic design results in blocks of large size if the rotations are lengthy, or if the number of rotations is relatively large. Therefore, the use of incomplete block designs is sometimes necessary. Incomplete block designs used in rotation experiments are reduced designs and phase-confounded designs. For a discussion of these designs see Patterson (25).

B. Conventional Statistical Analysis

Statistical analysis of experimental data consists of the formulation of a statistical model and the estimation of parameters and hypothesis testing. Strictly speaking the formulation of a model is a basis for the design of experiment and will not enter the picture of analysis unless a unique model is not specified. The estimation of parameters in the model therefore constitutes the major part of analysis, although in some areas of study it is customarily followed by hypothesis testing. For the linear model (1.2) with some necessary assumptions these can be conviently carried out with minimum efforts. There exists a number of problems in the analysis of rotation experiments, however. One problem is the construction of a suitable model for the estimation of treatment effects. But the major difficulty lies in the error structure. That the error structure poses a problem is understandable. For it is only logical to reject the usual assumption of independent errors in view of the particular nature of experimentation. We shall examine these problems in the following paragraphs as we review Patterson's paper (25), which is the most comprehensive article on the subject of the analysis of rotation experiments.

As pointed out by Patterson (25), the conventional method of the analysis of rotation experiments consists of estimating mean effects over the years and fitting polynomial curves to determine the trend of rotation effects. This

is the method demonstrated in the pioneer paper by Cochran (3), and is also the general approach adopted by Patterson in his paper.

1. The statistical model

The analysis presented by Patterson has been developed for rotation experiments of type (b) in terms of what he calls model I. Only a slight modification is needed for the analysis of rotation experiments of type (b) with other cultivation treatments.

Let a test crop, or test, be a crop which is included in different basic rotations and on which the rotation effects are to be compared. Let y_{ij} and T_{ij} be the observed and true mean yield for test j in year i, T_j be the true mean yield for test j, and

$$x'_{im} = x_{im} - \bar{x}_m, \qquad m = 1,2,\ldots,q$$

where x_{im} is the m-th seasonal or trend variate in year i and $\bar{x}_m$ is the mean of x_{im} over the years. Then model I is represented by

$$\begin{aligned} y_{ij} &= T_{ij} + v_{ij} \\ &= T_j + \sum_m \beta_{jm} x'_{im} + u_{ij} + v_{ij}, \qquad \begin{matrix} i = 1,2,\ldots,N, \\ j = 1,2,\ldots,t, \end{matrix} \end{aligned} \tag{4.1}$$

where for example x'_{im} might be polynomials in time, and

$$\begin{aligned} &E(u_{ij}) = E(v_{ij}) = 0 \\ &E(u_{ij}u_{i'j'}) = \begin{cases} \gamma_{jj'} & \text{if } i=i' \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4.2}$$

$$E(v_{ij}v_{i'j'}) = \begin{cases} \dfrac{\sigma_p^2 + \sigma_w^2}{r} & \text{if } i=i',\ j=j' \\ \dfrac{\sigma_p^2}{r} & \text{if } i\neq i', \text{ but } y_{ij} \text{ and } y_{i'j'} \text{ occur on the same plots.} \\ 0 & \text{otherwise} \end{cases} \qquad (4.3)$$

where r is the number of replicates. The experimental errors as defined in (4.3) is the sum of two components: One is the plot error constant over the years with variance σ_p^2, and the other the plot x year error with variance σ_w^2. Note also that u_{ij} represent residual seasonal variations which are correlated for observations appearing in the same year. These also might be called "year" or weather effects. Notice that these are visualized as a vector of (correlated) random variables.

Let the Kronecker product of two matrices A and B be denoted by $A \otimes B$. Let I_t be the t x t identity matrix. The model (4.1) can then be expressed in matrix notation as follows

$$y = K_t T + K_b \beta + u \qquad (4.4)$$

where $y = (y_1', y_2', \ldots, y_t')'$

y_j is the N x 1 matrix of the mean yield $y_{ij}, j=i,2,\ldots,t$

$T = (T_1, T_2, \ldots, T_t)'$

$\beta = (\beta_1', \beta_2', \ldots, \beta_t')'$

$\beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jq})', \ j=1,2,\ldots,t$

$$u = (u_{11} + v_{11},\ u_{21} + v_{21}, \ldots, u_{Nt} + v_{Nt})$$

$$K_t = I_t \otimes 1_N$$

$$K_b = I_t \otimes X$$

X is the N x q matrix of elements x'_{im}.

To determine the covariance matrix V of y, usual procedure of finding E(uu') suffices except that the evaluation of elements involving σ_p^2/r when $i \neq i'$ will tend to be tedious. Patterson accomplishes this by introducing the N x p incidence matrix H_j. If test j is on plot k in year i then the k-th element in row i of H_j is equal to 1 and all other elements in row i are zero. Let

$$K_p = (H'_1, H'_2, \ldots, H'_t)',$$

then the covariance matrix of y is

$$V = \frac{K_p K'_p \sigma_p^2}{r} + \gamma \otimes I_N + \frac{\sigma_w^2}{r} I_t \otimes I_N \qquad (4.5)$$

where γ is the t x t matrix of the $\gamma_{jj'}$.

The analysis is carried out in two steps: the estimation of σ_p^2 and σ_w^2 and the estimation of rotation effects and regression coefficients.

2. Estimation of variance components

Patterson recommends the use of the analysis of variance method for the estimation of σ_p^2 and σ_w^2, because for some ratios of σ_p^2/σ_w^2 the method gives high efficiencies for the linear functions of σ_p^2 and σ_w^2 he investigates. Following the usual procedure of estimating variance components from an analysis of variance table, this method consists of

obtaining plot and plot x year error mean squares, equating the mean squares to their expected values, and solving for the estimates. Several procedures are available for obtaining the plot mean square, we shall review the method of solving normal equations.

We wish to find the sum of squares due to plot, eliminating year effects. Let A_i be the estimated year effect for year i and p_{kh} the estimate of the effect of plot k, k = 1,2, ...,p, in replicate h. Then for the h-th replicate, h = 1,2, ...,r, we have the set of N+p normal equations:

$$\begin{aligned} A_i: &\quad \sum_k n_{ik}A_i + \sum_k n_{ik}p_{kh} = \sum_{j,k} n_{ik}y_{ijkh} \\ p_{kh}: &\quad \sum_i n_{ik}A_i + \sum_i n_{ik}p_{kh} = \sum_{i,j} n_{ik}y_{ijkh} \end{aligned} \tag{4.6}$$

where y_{ijkh} = the yield of test j on plot k in replicate h in year i

n_{ik} = 1 if the test crop appears on plot k in year i

= 0 otherwise.

Note that if $n_{ik} = 0$, the observation y_{ijkh} does not exist. Since $\sum_k n_{ik} = t$, eliminating A_i in (4.6) leads to the set of normal equations for p_{kh} in matrix notation

$$\emptyset p_h = P_h \tag{4.7}$$

where $p_h = (p_{1h}, p_{2h}, \ldots, p_{ph})'$

$\emptyset = (\emptyset_{kk'})$ is a p x p matrix such that

$$\phi_{kk'} = \begin{cases} \sum_i n_{ik} - \frac{1}{t} \sum_i n_{ik}^2 & \text{if } k=k' \\ -\frac{1}{t} \sum_i n_{ik} n_{ik'} & \text{if } k \neq k' \end{cases}$$

$P_h = (P_{1h}, P_{2h}, \ldots, P_{ph})'$ where

$$P_{kh} = \sum_{i,j} n_{ik} y_{ijkh} - \frac{1}{t} \sum_{i,j,k'} n_{ik} n_{ik'} y_{ijk'h}.$$

It is clear that ϕ is not of full rank. Let v be the rank of ϕ. If $v = (p-1)$, a unique solution of (4.7) can be obtained by imposing the side condition $\sum_k p_{kh} = 0$, i.e.

$$p_h = (\phi + gJ_p)^{-1} P_h$$

where J_p is a p x p matrix of unit elements and g is any scalar. This is in fact a special case for solving the normal equations $X'X\beta = X'y$ subject to the condition $H\beta = 0$, where $X'X$ is a p x p matrix of rank $r < p$, and for $t \geq p-r$, H is a p x t matrix satisfying certain general conditions (see for example Scheffé (30,pp. 15-19)). For $v < p-1$, Equations 4.7 fall into p-v independent groups. Patterson suggests that each group be solved separately and the results pooled. The plot error sum of squares is then given by

$$\sum_h p_h' P_h - \frac{1}{r} \sum_h p_h' \sum_h P_h.$$

If we calculate the total sum of squares for error over all years, the plot x year error sum of squares can be obtained by substraction. The expected values of the error mean squares are as follows:

	d.f.	E(MS)
Plot	$(r-1)v$	$\sigma_w^2 + \frac{N(t-1)}{v}\sigma_p^2$
Plot x year	$(r-1)[N(t-1)-v]$	σ_w^2
Total	$N(r-1)(t-1)$	$\sigma_w^2 + \sigma_p^2$

Estimates $c_{jj'}$ of $\gamma_{jj'}$ can be obtained as

$$c_{jj'} = \frac{S_{jj'} - S_{jj'0}}{(N-q-1)}$$

where $S_{jj'} = (y_j - 1_N t_j - Xb_j)'(y_{j'} - 1_N t_{j'} - Xb_{j'})$

$S_{jj'0}$ is the estimate of

$$\frac{(N-q-1)\delta_{jj'}\sigma_w^2 + tr(H_j H_{j'}' F)\sigma_p^2}{r}$$

and $F = I_N = \frac{J_N}{N} - X(X'X)^{-1}X'$.

3. Estimation of rotation effects and regression coefficients

The parameters T and β in (4.4) may be estimated by simple least squares procedure. The normal equations are

$$\begin{pmatrix} K_t'K_t & K_t'K_b \\ K_b'K_t & K_b'K_b \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} K_t' \\ K_b' \end{pmatrix} y$$

where t and b are the simple least squares estimates of T and β. We find that

$$\begin{pmatrix} K_t'K_t & K_t'K_b \\ K_b'K_t & K_b'K_b \end{pmatrix}^{-1} = \begin{pmatrix} N^{-1}I_t & 0 \\ 0 & I_t \otimes (X'X)^{-1} \end{pmatrix}. \qquad (4.8)$$

Hence the solution of (4.8) is

$$\begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} N^{-1}(I_t \otimes 1_N')y \\ I_t \otimes [(X'X)^{-1}X'y] \end{pmatrix} . \qquad (4.9)$$

It is seen from (4.9) that the same results would have been obtained if one estimates T_j and β_j separately for each test j.

Because of the error assumptions in (4.2) and (4.3), the expression in (4.8) does not describe a covariance matrix for the estimates. Recall from (4.5) that

$$\text{Cov}(y) = V = \frac{K_p K_p' \sigma_p^2}{r} + \gamma \otimes I_N + \frac{\sigma_w^2}{r} I_t \otimes I_N,$$

the covariance of $(t'b')'$ is

$$\text{Cov} \begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} \frac{I_t}{N} & 0 \\ 0 & I_t \otimes (X'X)^{-1} \end{pmatrix} \begin{pmatrix} K_t' \\ K_b' \end{pmatrix} V (K_t \; K_b) \begin{pmatrix} \frac{I_t}{N} & 0 \\ 0 & I_t \otimes (X'X)^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{pmatrix},$$

where

$$M_{11} = \frac{K_t' V K_t}{N^2}$$

$$= \frac{1}{N^2} (I_t \otimes 1_N')K_p K_p' (I_t \otimes 1_N) \frac{\sigma_p^2}{r} + \frac{1}{N}\left(\frac{\sigma_w^2}{r} I_t + \gamma\right),$$

$$M_{22} = [I_t \otimes (X'X)^{-1}]K_b' V K_b [I_t \otimes (X'X)^{-1}]$$

$$= \frac{\sigma_p^2}{r} [I_t \otimes (X'X)^{-1}X']K_p K_p' [I_t \otimes X(X'X)^{-1}]$$

$$+ \left(\frac{\sigma_w^2}{r} I_t + \gamma\right) \otimes (X'X)^{-1},$$

$$M_{12} = \frac{1}{N} K_t' V K_b \left[I_t \otimes (X'X)^{-1}\right]$$
$$= \frac{1}{N} (I_t \otimes 1_N') K_p K_p' \left[I_t \otimes X(X'X)^{-1}\right].$$

Note that M_{11} is the covariance matrix of t, M_{22} the covariance matrix of b, and M_{12} represents the covariances between t and b. It follows therefore that

$$\mathrm{Cov}(t_j, t_{j'}) = \frac{1}{N^2} 1_N' H_j H_{j'}' 1_N \frac{\sigma_p^2}{r} + \frac{1}{N} (\delta_{jj'} \frac{\sigma_w^2}{r} + \gamma_{jj'}),$$

$$\mathrm{Cov}(b_j, b_{j'}) = (X'X)^{-1} H_j H_{j'}' X(X'X)^{-1} \frac{\sigma_p^2}{r}$$
$$+ (\frac{\delta_{jj'}\sigma_w^2}{r} + \gamma_{jj'})(X'X)^{-1},$$

$$\mathrm{Cov}(t_j, b_j) = \frac{1}{N} 1_N' H_j H_{j'}' X(X'X)^{-1} \frac{\sigma_p^2}{r},$$

where $\delta_{jj'} = 1$ if $j=j'$

$= 0$ if $j \neq j'$

and 1_N is an N-vector with unit elements.

Patterson assumes that the estimate t_j is efficient if the cycles of the rotation that includes test j are complete, namely, N is a multiple of the length of the test. The estimates of regression coefficients are, however, inefficient whether the cycles are complete or not. Fully efficient t and b can be obtained by generalized least squares procedure by solving the normal equations

$$\begin{pmatrix} K_t' \\ K_b' \end{pmatrix} V^{-1} (K_t \; K_b) \begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} K_t' \\ K_b' \end{pmatrix} V^{-1} y,$$

i.e.

$$\begin{pmatrix} K_t'V^{-1}K_t & K_t'V^{-1}K_b \\ K_b'V^{-1}K_t & K_b'V^{-1}K_b \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} K_t'V^{-1}y \\ K_b'V^{-1}y \end{pmatrix} .$$

Since the covariance matrix V is usually unknown, the variance components σ_w^2, σ_p^2 and $\gamma_{jj'}$ are in practice replaced by thin respective estimates.

4. An alternative approach

We shall let the vector T be absorbed by the vector β. Without loss of generality, we let $x_{i1}' = 1$ and $\beta_{j1} = T_j$ for all i and all j. Further, let $v_{ij} = P_k + e_{ijk}$, k = 1,2, ...,p, and

$$E(P_k P_{k'}) = \sigma_p^2/r \qquad \text{if } k = k'$$
$$= 0 \qquad \text{if } k \neq k'$$

$$E(e_{ijk}e_{i'j'k'}) = \sigma_w^2/r \quad \text{if } i = i',\ j = j',\ k = k'$$
$$= 0 \qquad \text{otherwise,}$$

then model (4.1) can be written

$$y_{ijk} = \sum_m \beta_{jm}x_{im}' + u_{ij} + P_k + e_{ijk}. \tag{4.10}$$

Next let β_1 be a b-vector of estimable functions of β_{jm}, δ a d-vector of estimable functions of u_{ij}, and P a p_1-vector of estimable functions of P_k. We then have the reparameterized model

$$y = X_1\beta_1 + X_2\delta + X_3P + e$$
$$= X\theta + e$$

where $\theta = (\beta_1' \ \delta' \ P')'$ and $X = (X_1X_2X_3)$ is the coefficient

matrix resulted from the reparametrization of model (4.10). The simple least squares estimate of θ is given by

$$\hat{\theta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\delta} \\ \hat{P} \end{pmatrix} = (X'X)^{-1}X'y$$

$$= \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} X'y$$

where T_{11} is that block of $(X'X)^{-1}$ associated with β_1, and so on, and the sum of squares due to fitting P, eliminating β_1 and δ, is $\hat{P}'T_{33}^{-1}\hat{P}$.

In constructing the vector P, we may assume $1_{p_1}P' = 0$. We also assume that

$$\hat{P} = (P - \bar{P}) + e_p$$

where $E(P - \bar{P})(P - \bar{P})' = C\sigma_p^2/r$ and $E(e_p e_p') = T_{33}\sigma_w^2/r$ so that

$$Cov(\hat{P}) = V_1 = C\sigma_p^2/r + T_{33}\sigma_w^2/r.$$

Hence

$$\begin{aligned} E(\hat{P}'T_{33}^{-1}\hat{P}) &= tr(T_{33}^{-1} V_1) \\ &= tr[T_{33}^{-1} (C\sigma_p^2/r + T_{33}\sigma_w^2/r)] \\ &= tr(T_{33}^{-1} C\sigma_p^2/r + I\sigma_w^2/r) \\ &= tr(T_{33}^{-1} C\sigma_p^2/r) + p_1\sigma_w^2/r \ . \end{aligned}$$

A parallel discussion for δ leads to the result that sum of squares due to δ eliminating others is $\hat{\delta}'T_{22}^{-1}\hat{\delta}$ and

$$E(\hat{\delta}'T_{22}^{-1}\hat{\delta}) = tr(T_{22}^{-1}V_2)$$

$$= \mathrm{tr}(T_{22}^{-1}D_1) + \mathrm{tr}(T_{22}^{-1}D_2)$$

where D_1 has elements of σ_p^2 and σ_w^2 and D_2 has elements of γ_{ij}. The exact form of D_1 and D_2 will depend on the rotations included in the experiment. Equating $\hat{P}'T_{33}^{-1}\hat{P}$ and $\hat{\delta}'T_{22}^{-1}\hat{\delta}$ to their expected values, one can obtain the estimates of σ_w^2, σ_p^2 and γ_{ij}. It follows that the generalized least squares estimate of β_1, where V is replaced by its estimate $\hat{V}$, is

$$\hat{\beta}_1 = (X_1'\hat{V}^{-1}X_1)^{-1}X_1'\hat{V}^{-1}y,$$

and

$$\mathrm{Cov}(\hat{\beta}_1) = (X_1'\hat{V}^{-1}X_1)^{-1} .$$

5. A particular type of transformation

We shall digress to consider one type of transformations that we shall use in a later stage. Recall that if $y = X\beta + e$ and $E(ee') = V$, the transformation

$$z = Ty = TX\beta + Te \tag{4.11}$$

where T satisfies $E(Tee'T') = TVT' = \sigma_I^2$, leads to the fully efficient generalized least squares estimate of β

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{4.12}$$

If the matrix V in (4.12) is replaced by its estimate $\hat{V}$, the $\hat{\beta}$ is consistent and asymptotically efficient. When the number of observations is large, the operation of inverting the covariance matrix V may be tedious. For some special forms of V the use of the following simple transformations is equivalent to the generalized least squares procedure, but

eliminates the need for the inversion of V.

Assume that

$$\begin{aligned} \mathrm{Cov}(y_{ij},y_{i'j'}) &= \sigma^2 = \sigma_a^2 + \sigma_b^2 && \text{if } i=i',\ j=j' \\ &= \rho\sigma^2 = \sigma_a^2 && \text{if } i\neq i',\ j=j' \\ &= 0 && \text{otherwise} \end{aligned}$$

where $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$, then for

$$K = \sqrt{\frac{1-\rho}{1+(n-1)\rho}} - 1, \qquad y_{.j} = \sum_i y_{ij}/n$$

the transformed variables

$$z_{ij} = y_{ij} + Ky_{.j} \tag{4.13}$$

are such that

$$\begin{aligned} \mathrm{Var}(z_{ij}) &= \mathrm{Var}(y_{ij} + Ky_{.j}) \\ &= (1-\rho)\sigma^2 \\ &= \sigma_a^2 \end{aligned}$$

$$\mathrm{Cov}(z_{ij},z_{i'j}) = \mathrm{Cov}(z_{ij},z_{i'j'}) = 0,\ i\neq i',\ j\neq j'.$$

If

$$\begin{aligned} \mathrm{Cov}(y_{ij},y_{i'j'}) &= \sigma^2 && i = i',\ j = j' \\ &= \rho_1\sigma^2 && i = i',\ j \neq j' \\ &= \rho_2\sigma^2 && i \neq i',\ j = j' \\ &= 0 && i \neq i',\ j \neq j' \end{aligned}$$

and

$$T_1 = \sqrt{\frac{1-\rho_1-\rho_2}{1+(m-1)\rho_1-\rho_2}} - 1,$$

$$T_2 = \sqrt{\frac{1-\rho_1-\rho_2}{1-\rho_1+(n-1)\rho_2}} - 1,$$

$$T_3 = \sqrt{\frac{1-\rho_1 - \rho_2}{1 + (m-1)\rho_1 + (n-1)\rho_2}} - (1 + T_1 + T_2).$$

Then for $\rho_1 \geq 0$, $\rho_2 \geq 0$ and $1 - \rho_1 - \rho_2 > 0$, the variables

$$z_{ij} = y_{ij} + T_1 y_{i.} + T_2 y_{.j} + T_3 y_{..} \qquad (4.14)$$

$$y_{i.} = \frac{1}{m} \sum_j y_{ij}, \qquad y_{..} = \frac{1}{mn} \sum_{i,j} y_{ij}$$

have variance $(1-\rho_1 - \rho_2)\sigma^2$ and zero covariances.

One application of the transformation (4.13) is related to experiments with split-plot design. Usual procedure in analyzing data from such experiments is to perform analysis of variance and test hypotheses concerning the main effects and interactions. Instead of the experimental design model, however, one may wish to investigate the functional relationship between the observed outputs and the treatments if the latter are quantitative. The transformation (4.13) will remove the correlations among the observations belonging to the same whole plot and enables the use of simple least squares procedure. Since

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} \quad \text{and} \quad \frac{1-\rho}{1 + (n-1)\rho} = \frac{\sigma_b^2}{\sigma_a^2 + n\sigma_b^2}$$

the constant K is estimated as

$$K = \sqrt{\frac{\text{Error B}}{\text{Error A}}} - 1$$

where errors A and B are the whole plot and sub-plot error mean squares obtained from the analysis of variance table.

An example of the application of the transformation

(4.14) has been given by Fuller (8).

C. The Growth Process of Rotations and the Exponential Model

We have suggested that rotation effects be viewed as the realization of growth processes. The limiting value of the growth process of a rotation is clearly the information sought in a rotation experiment because it indicates the long-term performance one may expect from the rotation. The rate of change, on the other hand, is useful in determining the duration required for the experiment to provide reliable information on the limiting value. The question therefore arises as to how one may estimate the limiting value and the rate at which a rotation approaches it.

As we have noted earlier, a linear statistical model is assumed in the conventional analysis of rotation experiments. The usual assumption is that, apart from random variation, the yield of a rotation is the sum of the true mean and trend effects, the latter being approximated by polynomials of various degrees. The true mean is often estimated as the average yield of several cycles. Whether such an estimate is informative in suggesting the long-term performance of a rotation is questionable. The trend as approximated by polynomials may provide some additional information. However, as a polynomial does not have a limiting value, its inclusion gives little indication as to what limiting value the

rotation may eventually attain.

In view of the discussion above, it is logical to formulate a statistical model that will provide a limiting value for rotation effects. Since we consider rotation effects as growth processes, a natural choice for a statistical model is the exponential model discussed earlier in this manuscript. Such a model, with parameters for the limiting value and the rate of change, offers a more meaningful explanation of the mechanism of rotation effects. It may also have the following advantage.

As the ultimate purpose of rotation practice is invariably to maximize farming profits on a long-term basis, there is no doubt that economic analysis on the net return will be a helpful guide in selecting the most profitable rotation. That such study has not been reported in literature may partly be accounted for by the lack of a suitable model for rotation effects. Although we shall not pursue this matter further, we should point out that, perhaps by the use of exponential models, economic study may be encouraged or may eventually become an integral part of rotation study.

1. Formulation of a rotation model

We shall now discuss the analysis of rotation experiments with exponential models. We first express our model in the general functional form

$$y = f(x, \theta) \qquad (4.15)$$

where y is an observed yield, $f(x,\theta)$ may be a simple exponential model or a combination of exponential and linear models. With appropriate form Equation 4.15 will be referred to as a rotation model.

While we hypothesize growth processes for rotation effects, there seems to be no reason to assume an error structure different from that of conventional analysis. We shall, therefore, assume that the error associated with the observation in (4.15) consists of the plot error and plot x year error defined in (4.3). The residual seasonal variations assumed in (4.2) will be accounted for either as contrasts in the model, or as deviations from the model. This is possible because the functional relationship (4.15) fits rotation experiments remarkably well.

Our first step then is to estimate the plot and plot x year error variances. Although plot variations have been found to exist (22, for example), it is interesting to note that a small negative value was estimated for the plot error variance for two rotation experiments conducted by Iowa Agricultural Experiment Station. If plot variance is non-zero, the transformation (4.11) is carried out before further analysis.

A rotation experiment may have as treatments several rotations and several fertilization practices. Although our development herein will be for the analysis of experiments of

this type, the application of the procedure elsewhere will present little difficulty.

A common type of fertilization practices consists of different levels of applied nitrogen. The treatments form factorial combinations, and, when a split-plot design is used, the rotations are often handled as whole plot treatments. The transformation (4.13) is useful in analyzing such experiments to remove the correlated errors introduced by the split-plot design.

In the following development we shall assume that all required transformations have been performed for y and x. Therefore we may write our model as

$$y = f(x,\theta) + e \tag{4.16}$$

where the errors are independent, identically distributed normal variables with mean zero and variance σ^2.

Let us first consider a rotation without fertilization treatments. We may express the yield by

$$y = \alpha_0 + \beta_0 e^{-ax} \tag{4.17}$$

or

$$y = \alpha + \beta(1 - e^{-ax}) \tag{4.18}$$

where $x = 1,2,\ldots$ denote cycles of the rotation. The models (4.17) and (4.18) are functionally related in that $\alpha_0 = \alpha+\beta$ and $\beta_0 = -\beta$. Model (4.18) has a more convenient form, for if we have t rotations, we may write

$$y_j = \alpha + \beta_j(1 - e^{-ax_j}), \quad j = 1,\ldots,t. \tag{4.19}$$

Then all rotations start with the same yield α, but rotation j has $\alpha + \beta_j$ as its limiting value or yield. The term $\beta_j(1 - e^{-ax_j})$ represents the rotation effect.

Next, suppose that in a rotation experiment s levels of nitrogen fertilizer are applied to each rotation. Since the combination of a particular rotation and a fixed level of nitrogen can be considered as an integral part of a treatment, or a unique rotation practice, we would be inclined to regard the effect of such a treatment combination as rotation effect. It follows that model (4.19) will suffice for rotations with the same nitrogen level. To have a common model for all treatment combination, however, we shall need some modifications.

Consider the response of a crop to s nitrogen levels. Following the general procedure of response surface study, we may describe the response function by a polynomial. Alternatively, we may represent the response with an exponential model, i.e.

$$y_k = K + A\, e^{-BN_k} \qquad k = 1,\ldots,s \tag{4.20}$$

where N_K is the quantity of nitrogen fertilizer applied, and y_k the corresponding response. It is true that as N_k increases, y_k will attain a maximum and then decreases. However, within the limit of nitrogen levels that are practical or of interest in the present stage of experimentation, it has been found that Equation 4.20 provides satisfactory fit to responses. In fact it is generally true that

(4.20) approximate the response over a wider range than does the quadratic.

We now combine (4.19) and (4.20) into a single model

$$y_{jk} = \alpha + Ae^{-BN_k} + \beta_{jk}(1-e^{-ax_j}) \qquad (4.21)$$

where N_k is replaced by N_{jk} if the nitrogen levels vary with rotations. With this model, treatment jk starts with the yield $\alpha + Ae^{-BN_k}$ and has a limiting yield of $\alpha + Ae^{-BN_k} + \beta_{jk}$. Note that instead of β_j we have β_{jk} as the limiting rotation effect. This is because that the effect of a rotation may vary with the applied nitrogen levels. By using β_{jk} we have included the rotations x nitrogen interaction in the rotation effects, and when we speak of a rotation, we shall often refer to it as rotation jk.

There has been no allowance for year effects in the rotation models above. Because of the lack of a more satisfactory approach, the usual procedure is to assign dummy variables for years and estimate year effects by the coefficients of these variables (3, 22, 40), with a similar procedure for years by treatments interaction. Thus, for example, in the analysis of variance for a 12-year rice-pasture experiment by Yates (40), there are 11 degrees of freedom for the variation due to years ignoring treatments. In another example, Yates (40) partitions the year to year variations into two portions: one is the years within series variation (For any rotation, plots with different phases

occurring in the same year are said to belong to different series. A rotation of length three has three series.), and the other is the among series variation. It should be noted that the years within series variation also measures, under the concept of growth process, the effects of rotations. Little note has been made of this in the literature. A conventional analysis one seems to be more concerned with the comparison of rotation performance averaged over all cycles, rather than the change from cycle to cycle and the limiting values of various rotations.

If a rotation has gone through c cycles of length n, one can estimate part of the year effect by estimating variation among years within cycles, with c(n-1) degrees of freedom. The concept of growth process implies that yields of the same cycle receive the same rotation effect so that this portion of year variation is free from rotation effects. One might therefore be tempted to use cycle averages for statistical analysis. Unfortunately the use of cycle average will reduce considerably the number of parameters one can estimate, and more seriously it rules out the possibility of investigating year x treatment interactions, which are often present in a rotation experiment. When cycles are not complete, or when rotations of unequal length are involved, even the estimation of that portion of year effects that is free from rotation effects may be impossible. In fact, unless all cycles are represented each year, rotation and

year effects will at best be partially confounded.

Denoting the effect of year i, i = 1,2,...,M, by a_i, the year x nitrogen interaction by $(aN)_{ik}$, and the variable for year x rotation effect interaction by $a_i\beta_{jk}(1 - e^{-\gamma x_j})$(9), model (4.21) is extended to

$$y_{ijk} = \alpha + a_i + Ae^{-BN_k} + \beta_{jk}(1-e^{-\gamma x_j}) + (aN)_{ik} + C[a_i\beta_{jk}(1-e^{-ax_j})] \qquad (4.22)$$

The model (4.22) will not be a suitable model for all rotation experiments. For example, if the inspection of data indicates the presence of years x quadratic nitrogen interaction, the term $(aN^2)_{ik}$ should be included in the model. The statement holds true for other factors, such as meteorological information, cultivation practices, and many others, that are known to contribute to the variation in yield. In brief, with due caution any pertinent factor may be considered and included in the model in a convenient but appropriate form.

An example of the use of the rotation model (4.22) is the analysis of a rotation-fertility experiment by Fuller and Cady (9), who used a transformation different from that of (4.13). The experiment was conducted at Carrington-Clyde Experimental Farm by Iowa Agricultural Experiment Station over a period of twelve years, 1952 to 1963. Since a sizeable increase in stand levels occurred during the experimental period, a term denoted by S_iM_j was added to

model (4.22) for the linear nitrogen by stand interaction.

The last nine years were used in their analysis since all treatments were present. Model (4.20) was first fitted to the average yields for the nine years, to give the response function for cycle zero

$$y_{jk} = 114 + Ae^{-0.4N_{jk}} . \qquad (4.23)$$

The full model was then

$$y_{ijk} - 114 = a_i + Ae^{-0.4N_{jk}} + \beta_{jk}(1-e^{-ax_j}) + (aN)_{ik} + C[a_i\beta_{jk}(1-e^{-ax_j})] + D(S_iM_j). \qquad (4.24)$$

The models (4.23) and (4.24) were estimated by iterative least squares procedure, under the restrictions

$$\sum_i a_i = 0, \qquad \sum_i (aN)_{ik} = \sum_k (aN)_{ik} = 0,$$

$$\sum_i S_i = 0, \qquad \sum_j \sum_k \beta_{jk} = 0.$$

2. An alternative model

In this section we shall construct a different rotation model and demonstrate the statistical procedure by analyzing a rotation-fertility experiment at Clarion-Webster Experimental Farm conducted by the Iowa Agricultural Experiment Station.

The Clarion-Webster experiment was initiated in 1954 with two replicates, five basic rotations of length four, and four sub-plot nitrogen treatments at levels 0, 30, 60 and 120 pounds per acre. The basic rotations are C-C-C-O, C-S-C-O, C-C-O-M and C-O-M-M and continuous corn. The notations

represent C - corn, O - oats, S - soybean and M - meadow. The corn crops in different basic rotations as well as those at different phases in the same basic rotations are considered as different 'rotations' or tests. For example, the first and second year corn in C-C-O-M are different rotations and denoted by C-C-O-M and C-C-O-M respectively. Continuous corn was considered as four rotations in the preliminary analysis, but was treated as one rotation in the final analyses.

The 1958-64 data were included in the analysis. The plot errors were assumed to be zero because estimated plot error variance was a small negative number. The whole plot and sub-plot mean squares obtained from regular analysis of variance are

Whole plot mean square = 182.98

Sub-plot mean square = 91.15.

All analyses were carried out on transformed data and results were converted to individual observation basis for presentation. The analysis of variance for the transformed data is shown in Table 8.

With this background information, we now proceed to discuss our hypotheses and models.

In the study of a cause-and-response relationship of natural occurring phenomena or experimental results, the first requirement of a model is that it provides close fit to the data. For a more basic study, it is also desirable that the

Table 8. Analysis of variance for the Clarion-Webster rotation-fertility experiment, 1958-1964

Source	d.f.	S.S.	M.S.
Years	6	66986	11164
Rotations	11	75153	6832
Years x Rotations	66	19366	293
Nitrogen	3	71037	23679
Years x Nitrogen	18	7253	403
Rotations x Nitrogen	33	50330	1525
Years x Rotations x Nitrogen	198	18722	95
Error	335	30534	91

model gives meaningful approximation to the mechanism of the behavior under investigation. With this in mind, our proposal of rotation model has been an attempt to bring to light the more basic aspect of a rotation experiment, the growth process of rotation effects. Further, one may formulate a certain hypothesis and seek to verify it by constructing a suitable model that will fit the data well when the hypothesis is valid. It is in this latter sense our model in this section is unique and different from that of the previous section.

It is clear that limiting value varies with rotations. It is however not clear whether the differences in rotation effects result from differences in the nutrients supplied to crop plant or other factors. Rotations affect crop yield in a number of ways. In addition to providing some control over diseases and pests, they improve soil fertility by

improving water and nutrient availability, and, as meadow is often included in a rotation system, further by returning a large amount of organic matter and nutrients to the soil. If soil fertility is the major factor that produces rotation effects, an interesting question is therefore: Can we use a common growth process to describe the differences?

To answer this question it is hypothesized that all rotation effects follow a common growth curve and that rotation effects can be expressed in fertilizer nitrogen equivalence.[1] We express this in the model

$$y_{jk} = \alpha(1-e^{aN}) \tag{4.25}$$

where N is the total nitrogen available to crop plant when the rotation effect tends to its limiting value. Let N_j denote the sum of the initial soil fertility and the gain in soil fertility due to rotation practice, we must have

$$N = N_j + N_k.$$

Hence

$$\begin{aligned} y_{jk} &= \alpha - \alpha e^{a(N_j + N_k)} \\ &= \alpha - (\alpha e^{aN_j})e^{aN_k} \\ &= \alpha + \beta_j e^{aN_k} \end{aligned} \tag{4.26}$$

where $\beta_j < 0$. When no fertilizer nitrogen is applied, the limiting value of rotation j is then $\alpha + \beta_j$.

---

[1]This was first hypothesized by Dr. W. D. Shrader and communicated to Dr. W. A. Fuller and Dr. F. B. Cady (31).

Our hypothesis of a common growth process implies that all rotations have the same value for the parameter a in model (4.26), although β_j may be different. In estimating the parameters, the model

$$y_{jk} = \alpha + \sum_m \beta_m x e^{aN_k} \qquad \begin{array}{l} x = 1 \text{ if } m = j \\ = 0 \text{ if } m \neq j \end{array} \qquad (4.27)$$

is therefore fitted.

The results obtained by fitting the rotation model (4.27) for the Clarion-Webster experiment are shown in Tables 9 and 10. The N_k were coded in the analysis as $N_1 = 0$, $N_2 = 1$, $N_3 = 2$ and $N_4 = 4$. Note that the corrected S.S. for y_{jk} consists of the variations due to treatments (rotations, nitrogen and their interaction) with a total of 47 d.f. In fitting model (4.27), these are partitioned into variation that is accounted for by the model and that by deviations from the model. An F-test can be performed to test the adequacy of the model by dividing the deviations M.S. by the error M.S. obtained in the analysis of variance. A non-significant result at a predetermined significant level would lead to the acceptance of the model.

In the model we have set $\beta_8 = 0$ for the rotation C-O-M-M because a small positive value was first estimated for β_8. Therefore ten parameters were estimated and the corrected model S.S. has nine d.f., leaving 38 d.f. for deviations S.S. The F value computed as 1.14 is not significant at 5% level.

Table 9. Analysis of variance for the model $y_{jk} = \alpha + \Sigma\beta_j e^{aN_k}$ for the Clarion-Webster experiment, 1958-1964

Source	d.f.	S.S.	M.S.
Treatments	47	196520	
Model	9	192584	21398
Deviations	38	3936	104
Error (Table 8)	335	30534	91

Table 10. Estimates of nitrogen model for the Clarion-Webster experiment, 1958-1964

Rotation	Parameter	Estimate	Standard deviation
	α	115.7	0.8
C-C-C-O	β_1	-14.6	2.8
C-C-C-O	β_2	-43.7	2.8
C-C-C-O	β_3	-57.3	2.8
C-S-C-O	β_4	-10.0	2.8
C-S-C-O	β_5	-17.3	2.8
C-C-O-M	β_6	- 3.0	2.8
C-C-O-M	β_7	-20.9	2.8
C-O-M-M[a]	β_8	0.0	--
Cont. corn	β_9	-66.3	1.6
	a	- 0.38	0.024

[a] β_8 was set equal to zero.

We have accepted model (4.27), or equivalently model (4.25), to describe the common growth process for all rotations. Before taking a closer examination of this result, however, we choose to complete the analysis by investigating the years by treatments interaction.

Model (4.26) when used for an individual year is

$$y_{ijk} = \alpha_i + \beta_{ij} e^{aN_k} \qquad (4.28)$$

where y_{ijk} is yield of rotation jk in year i. If we fit model (4.28) for each year and find the sum of regression S.S., the quantity obtained will be equal to the model S.S. plus β_j x years interaction S.S. Hence the β_j x years interaction S.S. can be obtained by difference.

For the Clarion-Webster experiment, the sum of regression S.S. was computed to be 213419. Since eight β_{ij} were estimated for each of the seven years, the eight β_j were estimated in model (4.27), the degrees of freedom for β_j x years interaction are 48.

It is seen from the analysis of variance in Table 6 that there are a total of (66 + 18 + 198) = 282 d.f. for years by treatments interaction. When Equation 4.28 is used to represent the interaction, that portion of variation that belongs to the 282 d.f. but is not accounted for by Equation 4.28 may be called deviations from Equation 4.28. To avoid confusion the deviations from model (4.27) will be referred to as deviation-1, and the deviations from Equation 4.28,

deviation-2. Deviation-2 S.S. is obtained by substracting the β_j x years S.S. from the sum of interaction S.S. associated with the 282 d.f. of years by treatments interactions. It can also be obtained as the difference between the sum of the seven residual S.S. and deviation-2 S.S. An F-test of deviation-2 M.S. against the error M.S. is not significant at 5% level, indicating that years by treatments interaction is sufficiently accounted for by β_j x years interaction.

The complete analysis of variance based on models (4.27) and (4.28) is given in Table 11.

While the β_j x years interaction provides a satisfactory explanation for the years by treatments interaction, one would want to explore further the pattern of the interaction. The first possible source of variation one may consider is whether the changes in β_j over years are in the same direction, namely, all β_j are larger in some years and smaller in others. To obtain this information we fit the following model for multiplicative effect

$$y_{ijk} = \alpha_i + \delta_i \beta_i e^{aN_k} \qquad (4.29)$$

with the estimated $\beta_j e^{aN_k}$ as independent variables. The sum of regression S.S. for fitting the multiplicative effect model, estimated to be 203571, is the sum of the model (4.27) S.S. and the multiplicative effect S.S. The latter with 6 d.f. is therefore obtained by difference.

It is seen that the 6 d.f. multiplicative effect

Table 11. Complete analysis of variance for the Clarion-Webster experiment, 1958-1964

Source	d.f.	S.S.	M.S.
Years	6	66986	11165
Treatments	47	196520	4181
Model	9	192584	21398
Deviation-1	38	3936	104
Treatments x years	282	45342	161
β_j x years	48	20835	434
Multiplicative	6	8717	1453
Remainder	42	12118	289
Deviation-2	234	24507	105
Error	335	30534	91

explained a large share of the β_j x years interaction. The remainder S.S. may be further partitioned if other constrasts are included in the model (4.29).

We now return to the model (4.25) we fitted for the Clarion-Webster experiment. The model is

$$y_{jk} = 115.7 - 115.7\ e^{-0.38N}, \tag{4.30}$$

where N is in 40 pound units of fertilizer nitrogen. The curve is plotted in Figure 6.

The acceptance of model (4.27) indicates that we can place all rotations on the same growth curve. The position of a treatment combination on the curve is determined by its limiting yield, the estimate of $\alpha+\beta_j e^{aN_k}$. For example,

$$y_{1k} = 115.7 - 14.6\ e^{-0.38N_k}. \tag{4.31}$$

The limiting yield in pounds per acre for the first year corn in the rotation C-C-C-O with zero level of nitrogen is

$$y_{11} = 115.7 - 14.6 = 101.1$$

which is the ordinate of the point on the curve that is associated with rotation 11.

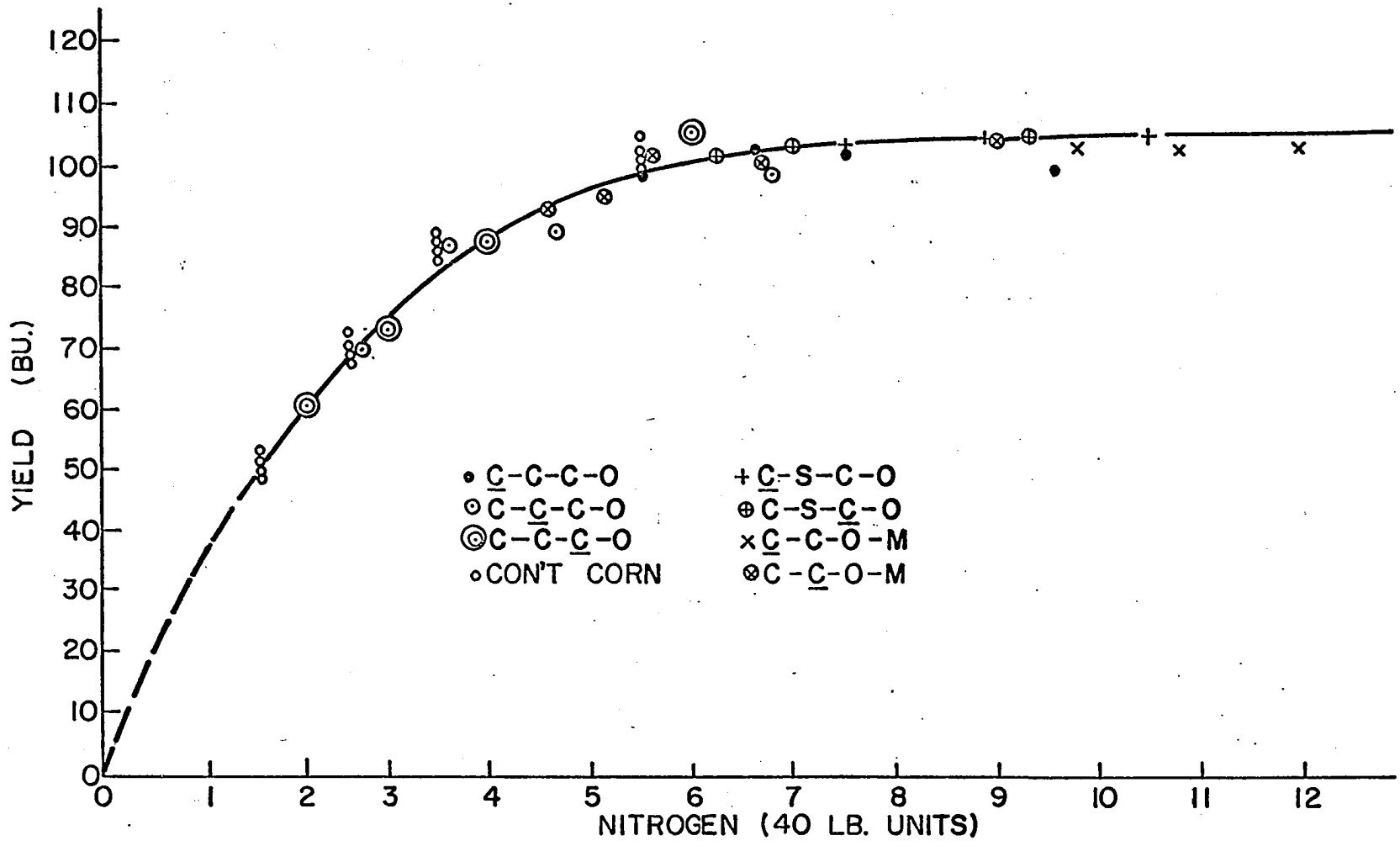
We can also locate the same point by first determining the abscissa. Recognizing from (4.25) and (4.27) that for rotation jk

$$\alpha\ e^{aN} = -\beta_j\ e^{aN_k},$$

we find the abscissa to be

$$N = N_k + \frac{1}{a}\,[\log(-\beta_j) - \log\,\alpha], \tag{4.32}$$

Figure 6. Corn-nitrogen response for the Clarion-Webster experiment, 1958-1964


YIELD (BU.)
120
110
100
90
80
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10 11 12
NITROGEN (40 LB. UNITS)
C-C-C-O
C-C-C-O
C-C-C-O
CON'T CORN
C-S-C-O
C-S-C-O
C-C-O-M
C-C-O-M

where log has base e. For rotation 11, we see that N = 5.5.

It is interesting to note that estimates of N_j can be obtained from (4.32) by letting $N_k = 0$. These estimates are given in Table 12.

Table 12. Estimates of soil fertility in fertilizer nitrogen equivalence, in 40-pound units

Rotation	Soil fertility
C-C-C-O	5.37
$\overline{C}$-C-C-O	2.52
C-$\overline{C}$-C-O	1.82
C-S-$\overline{C}$-O	6.35
$\overline{C}$-S-C-O	4.97
C-C-$\overline{O}$-M	9.47
$\overline{C}$-C-O-M	4.44
Cont. corn	1.44

The estimates of N_j given in Table 12 were obtained from (4.32) by letting $N_k = 0$. These estimates measure soil fertility in 40-pound units of fertilizer nitrogen when limiting rotation effects are reached.

We have seen that the results of the analysis strongly support the hypothesis that rotations can be placed on a common growth curve and that the measurable effect of rotations is nitrogen effect. The close fit of the model (4.30) to the long-term experiment is illustrated in Figure 6, where the rotation average yields were plotted along with their common growth curve. We were also able to show that the source of

treatments by years interaction is mainly the change of β_j from year to year, and that a good share of this variation can be accounted for by only 6 d.f. of multiplicative effect. The evidence so far seems to indicate that the approach adopted offers an analysis for long-term rotation experiments which is relatively simple computationally and yields meaningful and interpretable results.

V. SUMMARY

Standard statistical procedures do not always utilize the prior information on parameters under investigation. Development of theory to incorporate such information into statistical procedures has not yet yielded results that are readily applicable to a large class of situations. If a research worker acquires the prior information through experience in the area of study, the information may be vague and fragmentary, or may be rather well-defined. If the information is a direct consequence of the statistical model adopted, as in the case of the non-linear parameter ρ of an exponential model, the prior information can be formulated with certainty.

Assume that the parameter θ is uniformly distributed over the interval $(0,K)$, and that the least squares estimate $\hat{\theta}$ given θ is $N(\theta,1)$. It was shown that a gain in the efficiency in interval estimation in terms of expected length of confidence intervals can be achieved by constructing confidence intervals based on some interpolation tests. The interpolation tests are two-sided tests, the two rejection regions of which are of unequal size.

The parameter space of the non-linear parameter ρ of the exponential model is the interval $(0,1)$. If r is a linear ratio estimator of ρ, it was shown that Fieller's theorem can be modified to give real, finite confidence

intervals for ρ based on r, that contains values of ρ belonging to the interval (0,1) only. Two graphical methods were suggested for the determination of such confidence intervals. For a particular region of the sample space, these procedures can be modified by using interpolation tests to give shorter average confidence intervals. Discussion was also made on the modification of the technique required when parameter space is (0,K) for $K \neq 1$.

One class of long-term experiments in agriculture is the class of rotation experiments for the study of the effects of various systems of crop rotation. Rotation experiments have been conducted in many agricultural experiment institutes but the analysis of experimental results has not always been satisfactory.

A characteristic of rotation effect is that it tends to level off in the long run. By hypothesizing a growth process for rotation effects, it was possible to include the exponential model as part of the statistical model for rotation effects. An alternative approach also using the exponential model was used to build a statistical model under the assumption that all rotation effects can be described by a common growth curve. Analysis based on these models is relatively simple computationally and offers meaningful and interpretable results. An example was given for the analysis of a rotation-fertility experiment at the Clarion-

Webster Experimental Farm conducted by the Iowa Agricultural Experiment Station.

VI. LITERATURE CITED

1. Beale, E. M. L. Confidence regions in non-linear estimation. Royal Statistical Society Journal Series B, 22: 41-88. 1960.

2. Box, G. E. P. and Lucas, H. L. Design of experiments in non-linear situations. Biometrika 46: 77-90. 1959.

3. Cochran, W. G. Long-term agricultural experiments. Royal Statistical Society Journal, Supplement 6: 104-148. 1939.

4. Cramér, H. Mathematical methods of statistics. Princeton University Press, Princeton, N.J. 1946.

5. Fieller, E. C. The distribution of the index in a normal bivariate population. Biometrika 24: 428-440. 1932.

6. Fieller, E. C. A fundamental formula in the statistics of biological assay. Quarterly Journal of Pharmacy and Pharmacology 17: 117-123. 1944.

7. Finney, D. J. The efficiencies of alternative estimators for an asymptotic regression equation. Biometrika 45: 370-388. 1958.

8. Fuller, W. A. Some topics in the analysis of production experiments. To appear in Dillon, J. L., editor. Essays in honor of E. O. Heady. Iowa State University Press, Ames, Iowa. [To be published in 1966]

9. Fuller, W. A. and Cady, F. B. Estimation of asymptotic rotation and nitrogen effects. Agronomy Journal 57: 299-302. 1965.

10. Halperin, M. Confidence interval estimation in non-linear regression. Royal Statistical Society Journal Series B, 25: 330-333. 1963.

11. Hartley, H. O. The estimation of non-linear parameters by "internal least squares". Biometrika 35: 32-45. 1948.

12. Hartley, H. O. Exact confidence regions for the parameters in non-linear regression laws. Biometrika 51: 347-353. 1964.

13. Hartley, H. O. The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. Technometrics 3: 269-280. 1961.

14. Hartley, H. O. and Booker, A. Nonlinear least squares estimation. Annals of Mathematical Statistics 36: 638-650. 1965.

15. Hirons, R. W. A further note on the extension of Stevens' tables for asymptotic regression. Biometrics 18: 123. 1962.

16. Johnston, B. O. and Grandage, A. H. E. A sampling study of estimators of the nonlinear parameter in the exponential model. North Carolina State College Institute of Statistics Mimeograph Series No. 329. 1962.

17. Levenberg, K. A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics 2: 164-168. 1944.

18. Linhart, H. Tables for W. L. Stevens' "asymptotic regression". Biometrics 16: 125. 1960.

19. Lipton, S. On the extension of Stevens' tables for asymptotic regression. Biometrics 17: 321. 1961.

20. Marquardt, D. L. An algorithm for least squares estimation of non-linear parameters. Society for Industrial and Applied Mathematics Journal 11: 431-441. 1963.

21. Mitscherlich, E. A. Die Bestimmung des Düngerbedürfnisses des Bodens. Paul Parey, Berlin. 1925.

22. Patterson, H. D. The analysis of a non-replicated experiment involving a single four-course rotation of crops. Biometrics 15: 30-59. 1959.

23. Patterson, H. D. The analysis of the results of rotation experiment on the use of straw and fertilizer. Journal of Agricultural Science 43: 77-88. 1953.

24. Patterson, H. D. A simple method for fitting an asymptotic regression curve. Biometrics 12: 323-329. 1956.

25. Patterson, H. D. Theory of cyclic experiments. Royal Statistical Society Journal Series B, 26: 1-45. 1964.

26. Patterson, H. D. The use of autoregression in fitting an exponential curve. Biometrika 45: 389-400. 1958.

27. Patterson, H. D. and Lipton, H. An investigation of Hartley's method for fitting an exponential curve. Biometrika 46: 281-292. 1959.

28. Pimentel-Gomes, F. The use of Mitsherlich's regression law in the analysis of experiments with fertilizers. Biometrics 9: 498-516. 1953.

29. Portman, R. A study of the Monroe estimator of the non-linear parameter in the exponential model. Unpublished M.S. thesis. Library, North Carolina State College, Raleigh, North Carolina. 1961.

30. Scheffé, H. The analysis of variance. John Wiley and Sons, Inc., New York, New York. 1959.

31. Shrader, W. D., Fuller, W. A. and Cady, F. B. Estimation of a common nitrogen response function for corn in different crop rotations. [To be published in Agronomy Journal in 1966]

32. Stevens, W. L. Asymptotic regression. Biometrics 7: 247-267. 1951.

33. Tang, H. A sampling study of the linear parameters in the exponential model using various estimators of the nonlinear parameter. Unpublished M.S. thesis. Library, North Carolina State College, Raliegh, North Carolina. 1962.

34. Turner, M. E., Monroe, R. J. and Lucas, J. L. Generalized asymptotic regression and non-linear path analysis. Biometrics 17: 120-143. 1961.

35. Wey, I. T. Application of a priori information to the estimation of parameters in a linear regression model. Unpublished M.S. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa. 1965.

36. White, R. F. Procedures for fitting a class of exponential models. Unpublished M.S. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa. 1956.

37. Wilks, S. S. and Daly, J. F. An optimum property of confidence regions associated with the likelihood function. Annals of Mathematical Statistics 10: 225-235. 1939.

38. Williams, E. J. Exact fiducial limits in non-linear estimation. Royal Statistical Society Journal Series B, 24: 125-139. 1962.

39. Williams, E. J. Regression analysis. John Wiley and Sons, Inc., New York, New York. 1959.

40. Yates, F. The analysis of experiments containing different crop rotations. Biometrics 10: 324-346. 1954.

VII. ACKNOWLEDGEMENTS

The author wishes to express deepest gratitude and appreciation for the guidance, encouragement, and inspiration of Dr. W. A. Fuller.

Thanks are also due Dr. T. A. Bancroft and Dr. F. B. Cady for their kindness, interest and assistance during the course of this study and the author's stay at Iowa State University.

Acknowledgement is made for the support of this study by Iowa Agriculture and Home Economics Experiment Station.