ESTIMATION OF THE INTERCEPT OF A LINEAR REGRESSION FUNCTION

by

WM. JACKSON HALL

Institute of Statistics, University of North Carolina

Special report to the Office of Naval Research
of work at Chapel Hill under Project NR 042 031
for research in probability and statistics.

Institute of Statistics
Mimeograph Series No. 79

August 1953

1. Introduction. By performing experiments at the levels $x_1, x_2, \ldots$, we obtain observations $y_i = \alpha + \beta x_i + \Delta_i$ $(i=1,2,\ldots)$, the Δ's being independent $N(0,\sigma^2)$ random variables, and α and β the unknown linear regression coefficients. The x's are at the disposal of the experimenter, subject to the restriction of being within a given range, say $x' \leq x \leq x''$. It is desired to design an experiment--i.e., to specify the levels at which the experiment is to be performed--to estimate that x, say θ, for which $E(y)$ equals some specified value; we choose this value to be zero without loss of generality. Then, we are to estimate $\theta = -\alpha/\beta$. We shall assume that $x' < \theta < x''$; moreover, we transform the x-axis so that $x' = -1$, $x'' = +1$. (See the diagrams, pg. 13 .)

We propose $t_N = -a_N/b_N$ as an estimate of θ, a_N and b_N being the least squares estimates of α and β, respectively, based on N observations, assuming the x's are determined before experimentation begins. In this case, t_N is the maximum likelihood estimate of θ since a_N and b_N are maximum likelihood estimates. If the x's are chosen sequentially so that they are random variables, we use the estimate t_N, computing a_N and b_N from the same formulas. From regression theory, we have

$$(1) \qquad b_N = \frac{\Sigma(x-\bar{x})y}{\Sigma(x-\bar{x})^2} = \beta + \frac{N\Sigma x\Delta - \Sigma x\Sigma\Delta}{N\Sigma(x-\bar{x})^2}$$

$$(2) \qquad a_N = \bar{y} - b_N\bar{x} = \alpha + \frac{\Sigma x^2\Sigma\Delta - \Sigma x\Sigma x\Delta}{N\Sigma(x-\bar{x})^2} .$$

(Sometimes we shall omit the subscripts on a_N, b_N, and t_N without fear of confusion.)

We shall give designs of experiments for estimating θ based on certain "optimum" criteria in both the non-sequential and sequential cases. Properties of both the non-sequential and sequential estimates t_N are discussed, including, in the non-sequential case, an approximation to the distribution of t_N and a confidence interval for θ. Finally, some examples have been constructed to illustrate the character of the designs.

2. Non-sequential estimation. We assume the x's to be fixed by the experimenter in advance of the experimentation. From regression theory, we know that (a,b) has a bivariate normal distribution with means α and β, variances ${\sigma_a}^2$ and ${\sigma_b}^2$, and covariance σ_{ab}, where

$${\sigma_a}^2 = \frac{\Sigma x^2}{N\Sigma(x-\bar{x})^2}\,\sigma^2, \qquad {\sigma_b}^2 = \frac{1}{\Sigma(x-\bar{x})^2}\,\sigma^2, \qquad \sigma_{ab} = \frac{-\Sigma x}{N\Sigma(x-\bar{x})^2}\,\sigma^2 .$$

For later reference, if we replace σ^2 above by s^2, an estimate of σ^2 on N-2 d.f., we obtain the estimates ${s_a}^2$, ${s_b}^2$, and s_{ab}, respectively, which are independent of a and b.

It may be shown that the ratio of two normally distributed variables has no finite moments (except in special cases); hence, $t = -a/b$ has no finite moments. This would not be so if b could not take on values in a small interval about zero; hence, if the coefficient of variation of b, v_b, is sufficiently small, such an event would not occur in practice. Therefore, we may suppose that if we give expansions in powers of v_b for the mean and variance of t, that the first few terms in the expansions will give, in practice, reasonable measures of location and scale, respectively. The symbols for mean and variance of t used below are to be understood in this light.

Using the series expansion for the expectation of the ratio of two normally distributed variables given by Rao [1, pp. 153-4], we have

$$E\,t = \theta - (\bar{x}-\theta)\sum_{\nu=1}^{\infty} \frac{(2\nu)!}{2^{\nu}\,\nu!} v_b^{2\nu}$$

where $v_b = \sigma_b/\beta = \sigma/\beta\Sigma(x-\bar{x})^2$. By a development similar to that in Rao, we obtain

$$\mathrm{Var}\,t = \frac{1}{N}\Sigma(x-\bar{x})^2 \sum_{\nu=1}^{\infty} \frac{(2\nu)!}{2^{\nu}\,\nu!} v_b^{2\nu} + (\bar{x}-\theta)^2 \sum_{\nu=1}^{\infty} c_\nu v_b^{2\nu}$$

where $c_\nu = 2\nu \frac{(2\nu)!}{2^{\nu}\,\nu!} - \frac{1}{2^{\nu}} \sum_{\mu=1}^{\infty} \frac{(2\mu)!}{\mu!}\frac{(2\nu-2\mu)!}{(\nu-\mu)!}$. Hence

(3) $$E\,t = \theta - \frac{\bar{x}-\theta}{\Sigma(x-\bar{x})^2}\frac{\sigma^2}{\beta^2} + O(v_b^4) .$$

(4) $$\mathrm{Var}\,t = \left[\frac{1}{N} + \frac{(\bar{x}-\theta)^2}{\Sigma(x-\bar{x})^2}\right]\frac{\sigma^2}{\beta^2} + O(v_b^4) .$$

(Further justification may be given for these expansions: (1) If we assume b to have a truncated (at b=0) normal distribution, the variance of t is finite and the first two terms in the expansion (4) may be shown to be correct up to the proper order [2, pp. 353-4, 358]. (2) If we expand t as a function of $\underline{\Delta} = (\Delta_1, \ldots, \Delta_N)$ in a Maclaurin series and take expectation and variance term-wise, we obtain the same first two terms as given in (3) and (4), even if the Δ's are not assumed normally distributed.)

We can reduce the bias of t and the variance of t, as given by (3) and (4), simultaneously by choosing the x's so that $\bar{x}-\theta$ is small and $\Sigma(x-\bar{x})^2$ is large. We shall choose the x's so that $\Sigma(x-\bar{x})^2$ will be maximized subject to $\bar{x}$ being fixed close to θ; this is accomplished by performing all experiments at one of the two extreme levels: n at $x'(=-1)$ and N-n at $x''(=+1)$ where n is chosen so as to make the following approximation as close as possible: $\bar{x} = (N-2n)/N \doteq \theta$; i.e., n is the closest integer to $\frac{1}{2}N(1-\theta)$. These will be termed the optimum criteria in the non-sequential case. Thus a good design will require some a priori knowledge of θ; if none is available, it would appear reasonable to retain the two-level design with $n \doteq N/2$.

(As further justification for the optimum criteria, we note that: (1) Maximizing the term corresponding to θ in the inverse of the information matrix for (β,θ) leads to the same criteria. (2) According to Fieller (see Section 5), we may obtain a confidence interval for θ with length

$$\frac{2t_o}{b^2-t_o^2 s_b^2}\sqrt{a^2 s_b^2 - 2abs_{ab} + b^2 s_a^2 - t_o^2(s_a^2 s_b^2 - s_{ab}^2)}$$

where t_o is a constant. If we replace (a,b,s^2) by (α,β,σ^2), minimization of the length of the confidence interval is essentially accomplished by the above criteria.)

If n observations are taken at x = -1 and N-n at x = +1, it may be shown that

$$a_N = \frac{1}{2}(\bar{y}+\bar{z}), \qquad b_N = -\frac{1}{2}(\bar{y}-\bar{z}), \qquad t_N = (\bar{y}+\bar{z})/(\bar{y}-\bar{z}),$$

where here $\bar{y}$ denotes the average of the observations at x=-1 and $\bar{z}$ the average of those at x=+1.

3. Properties of non-sequential estimates. For designs in which $\bar{x}-\theta$ is very small and/or $\Sigma(x-\bar{x})^2$ is large (in particular, for optimum non-sequential designs), Var $t \doteq \sigma^2/N\beta^2$. This may be estimated by s^2/Nb^2. We find that s^2/Nb^2 has no finite moments, but by arguments similar to those above, we obtain

$$E\frac{s^2}{Nb^2} = \frac{1}{N}\, Es^2\, E\frac{1}{b^2} = \frac{\sigma^2}{N\beta^2}\left[1 + 3v_b^2 + O(v_b^4)\right] .$$

Hence, s^2/Nb^2 is a consistent estimate of $\sigma^2/N\beta^2$ if $\Sigma(x-\bar{x})^2$ tends to infinity with N; moreover, it is a conservative estimate in that the bias is positive.

t_N, being the maximum likelihood estimate of θ, has all the well-known properties of such estimates, in particular, consistency and asymptotic normality, assuming $\Sigma(x-\bar{x})^2$ tends to infinity with N. The asymptotic variance is $\sigma^2/N\beta^2$.

4. The distribution of the non-sequential estimate. Geary [3] gives an approximation to the distribution of the ratio of two normal variables, the error being small if the coefficient of variation of the denominator variable is small. Applying his work, with some refinement, to the variable $t = -a/b$, we have, denoting the c.d.f. of t_N by F_N,

$$F_N(t) = \Phi[u(t)] + R(t)$$

where Φ is the standard normal c.d.f.,

$$u(t) = (\alpha+\beta t)/\sqrt{\sigma_b^2 t^2 + 2\sigma_{ab} t + \sigma_a^2}$$

$$= (t-\theta)\frac{\beta}{\sigma}\sqrt{\frac{\Sigma(x-\bar{x})^2}{t^2 - 2\bar{x}t + \Sigma x^2/N}} ,$$

and

$$R(t) = \sqrt{2/\pi}\int_{\beta/\sigma_b}^{\infty} \left(\frac{x\sigma_b^3 t - \beta\sigma_b^2 t - \alpha\sigma_b^2 + \beta\sigma_{ab}}{\sigma_b\sqrt{\sigma_a^2\sigma_b^2 - \sigma_{ab}^2}}\right) e^{-x^2/2}\, dx - (-\beta/\sigma_b) .$$

Now $R(t)$ is a monotone increasing function of t and varies from $-\Phi(-1/v_b)$ to $+\Phi(-1/v_b)$. Hence, for small v_b,

$$F_N(t) \doteq \Phi[u(t)] .$$

In particular, if $v_b \leq 0.430$, $|R(t)| \leq 0.01$.

For two-level designs,

$$u(t) = (t-\theta)\frac{\beta}{\sigma}\sqrt{\frac{4n(N-n)}{Nt^2 - 2(N-2n)t + N}}$$

and the normal approximation is valid within 1 % if $v_b \leq 0.430$ or $n(N-n)/N \geq 1.353\ \sigma^2/\beta^2$.

5. Confidence interval for θ (non-sequential case). Fieller [4] develops exact confidence intervals for the ratio of two normal variables based on the Student-Fisher t-distribution. He reasons as follows: Since $a + \theta b$ is normal and $s_a^2 + 2\theta s_{ab} + \theta^2 s_b^2$ is an independent estimate of its variance on N-2 d.f., it follows that

$$z = (a+\theta b)/\sqrt{s_a^2 + 2\theta s_{ab} + \theta^2 s_b^2}$$

has a t-distribution with N-2 d.f. Therefore, for given δ, if t_o is chosen so that $\Pr(|z| \leq t_o) = \delta$, we have

$$\delta = \Pr(z^2 \leq t_o^2) = \Pr[(a+\theta b)^2 \leq t_o^2(s_a^2 + 2\theta s_{ab} + \theta^2 s_b^2)]$$

$$= \Pr[(a^2 - t_o^2 s_a^2) + 2\theta(ab - t_o^2 s_{ab}) + \theta^2(b^2 - t_o^2 s_b^2) \leq 0] .$$

Fieller shows that if b is sufficiently different from zero (specifically, if $b^2/s_b^2 > t_o^2$, which is likely if v_b is small), then we have a confidence interval for θ:

$$\delta = \Pr\left(\frac{ab - t_o^2 s_{ab} - t_o c}{b^2 - t_o^2 s_b^2} \leq \theta \leq \frac{ab - t_o^2 s_{ab} + t_o c}{b^2 - t_o^2 s_b^2}\right)$$

where $c = a^2 s_b^2 - 2ab s_{ab} + b^2 s_a^2 - t_o^2 (s_a^2 s_b^2 - s_{ab}^2)$.

6. Sequential estimation. We now suppose the experiments to be performed sequentially, the level of each experiment being determined on the basis of the previous observations. Specifically, we suppose observations to be taken in groups of k (a positive integer), the levels in the $(m+1)^{st}$ group being determined on the basis of the observations in the first m groups; thus

$$(5) \quad x_{km+i} = f(y_1, y_2, \ldots, y_{km}) = g(\Delta_1, \Delta_2, \ldots, \Delta_{km}), \quad i=1,\ldots,k; \quad m=0,1,\ldots$$

say. Hence, the y's are no longer independent, and the previous results do not necessarily hold. We propose the same estimate, $t_N = -a_N/b_N$, where a_N and b_N are defined by (1) and (2) though now they may not be normally distributed.

Expanding t_N as a function of $\underline{\Delta} = (\Delta_1, \ldots, \Delta_N)$, from equations (1), (2), and (5), in a Maclaurin series, we obtain:

$$t_N = \theta + \sum_i \left.\frac{\partial t_N}{\partial \Delta_i}\right|_{\underline{\Delta}=0} \Delta_i + \frac{1}{2} \sum_{i,j} \left.\frac{\partial^2 t_N}{\partial \Delta_i \partial \Delta_j}\right|_{\underline{\Delta}=0} \Delta_i \Delta_j + \ldots$$

Assuming that we may take expectation and variance term-wise (though they may not be existent as in the non-sequential case), we have

$$(6) \qquad E\, t_N = \theta + \frac{1}{2}\sigma^2 \sum_i \frac{\partial^2 t_N}{\partial \Delta_i^2}\Big]_{\underline{\Delta}=0} + \ldots$$

$$(7) \qquad \text{Var}\, t_N = \sigma^2 \sum_i \Big(\frac{\partial t_N}{\partial \Delta_i}\Big]_{\underline{\Delta}=0}\Big)^2 + \ldots$$

$$= \Big[\frac{1}{N} + \frac{(\bar{\xi}-\theta)^2}{\Sigma(\xi-\bar{\xi})^2}\Big]\frac{\sigma^2}{\beta^2} + \ldots$$

where $\xi_i = x_i\Big]_{\underline{\Delta}=0}$.

Consider a sequential plan such that $\bar{x} = t_\ell$ for some $\ell < N$; we call all such plans in which observations are taken in groups of k "designs of type D_k." Then $\bar{\xi} = \bar{x}\Big]_{\underline{\Delta}=0} = t_\ell\Big]_{\underline{\Delta}=0} = \theta$, and $\text{Var}\, t_N = \sigma^2/N\beta^2 + \ldots$. By evaluating the second term in equation (6), we find it to be zero for designs of type D_k. Hence, for such designs, the bias and the variance, as given by (6) and (7), are simultaneously reduced. By evaluation of some of the higher order terms, it may be shown that $\Sigma(\xi-\bar{\xi})^2$ always appears with a negative exponent. Hence, as in the non-sequential case, a design in which all but the last group of observations are taken at $x = \pm 1$, and the last group is so allocated that $\bar{x} = t_{N-k}$, is optimum in the above sense. We call such a design a "truncated two-level design". Thus, for

$k = 1$, the first $N-1$ observations are to be taken at -1 and $+1$, keeping the average close to the previous estimate of θ so that the last observation may be taken at some x between -1 and $+1$ in such a way that $\bar{x} = t_{N-1}$. Explicitly, the sequential design is :

$$(8)\quad \begin{cases} x_1 = -1 \qquad x_2 = +1 \\ x_{m+1} = \text{sgn}(t_m - \frac{m-2n}{m+1}),\ 1 < m < N-2 \quad (\text{after the } m^{th} \text{ stage, we} \\ \qquad \text{suppose } n \text{ observations have been taken at } -1 \text{ and } m-n \text{ at } +1) \\ x_{N-1} = \min(Nt_{N-2} - \sum_{i=1}^{N-2} x_i + 1,\ \ +1) \\ x_N = Nt_{N-1} - \sum_{i=1}^{N-1} x_i \ , \end{cases}$$

where $\text{sgn}(u) = +1$ if $u \geq 0$, -1 if $u < 0$.

(If we wish to add further single observations with the possibility of terminating the experiment at any step but yet retaining a design of type D_1, we can take

$$x_{n+1} = (n+1)t_n - nt_{n-1} \quad (n \geq N) .$$

Then, at any stage, we have $\frac{1}{n+1} \sum_{i=1}^{n+1} x_i = t_n$. Moreover, such observations will permit a check on the linearity of the regression line, if desired.)

7. Properties of sequential estimates. In the two-level sequential, $F_N(t)$ -- the distribution of non-sequential t in Section 4 -- is the conditional distribution of t given that n observations were taken at -1 and $N-n$ at $+1$.

In the sequential case, we have no assurance that t_N is a maximum likelihood estimate; however, we do prove consistency: Consider an arbitrary sequential design in which $\Sigma(x-\bar{x})^2$ tends in probability to infinity with N. (For two-level designs, this assumption is that n and N-n tend in probability to infinity with N.) Now x_i is independent of Δ_j for $i \leq j$, so that $E(x_i\Delta_i)(x_j\Delta_j) = E(x_i\Delta_i x_j)E\Delta_j = 0 \ (i < j)$; thus

$$E\Sigma\Delta = 0, \qquad E(\Sigma\Delta)^2 = E\Sigma\Delta^2 = N\sigma^2 ,$$
$$E\Sigma x\Delta = 0, \qquad E(\Sigma x\Delta)^2 = E\Sigma x^2\Delta^2 \leq E\Sigma\Delta^2 = N\sigma^2 ;$$

moreover, $|\Sigma x| < N$ and $\Sigma x^2 < N$. By application of Tchebycheff's Inequality on $\frac{1}{N}\Sigma\Delta$ and $\frac{1}{N}\Sigma x\Delta$ and Slutsky's Theorem [2, pg. 255], the consistency of a_N and b_N as estimates of α and β follows from equations (1) and (2) and the above remarks, irrespective of the distribution of the Δ's. Further application of Slutsky's Theorem proves the consistency of t_N as an estimate of θ under the stated conditions.

8. Examples. Two samples have been constructed by assigning the values $\alpha = 1$, $\beta = 4$ $(\theta = -1/4)$, $\sigma^2 = 1$, $x' = -1$, $x'' = +1$, and taking values of the Δ's from Mahalanobis' "Tables of a random sample from a normal population" [5]. (Sample 1 consists of the first 20 values and Sample 2 the second 20 in Plate 1.) Four designs were used with each sample, two non-sequential and two sequential, and estimates were computed on the first 10 observations as well as on the total of 20. The designs used were:

Design 1: $x_1 = x_2 = -1$, $x_3 = x_4 = -1/2$, $x_5 = x_6 = 0$, $x_7 = x_8 = +1/2$, $x_9 = x_{10} = +1$; similarly for x_{11} to x_{20}.

Design 2: two-level non-sequential design with $n = N/2$, $x_i = (-1)^i$.

Design 3: $x_1 = -1$, $x_2 = +1$, $x_n = nt_{n-1} - (n-1)t_{n-2}$ $(n > 2)$ (see the last paragraph in Section 6).

Design 4: "optimum" sequential design, given by (8). (The designs for the samples of 10 were not truncated.)

Before presentation, the data were transformed by the transformation $x^* = 1 + 4x$ so that $\alpha^* = 0$, $\beta^* = 1$, $\sigma^2 = 1$, $x^{*\prime} = -3$, $x^{*\prime\prime} = +5$; hence t_N^* is an estimate of $\theta^* = 0$. (See the diagrams below.) The estimates of θ^* from each design and variance estimates from Designs 1 and 4 are given in Table I below; the levels of the designs are given in Table II; confidence intervals for θ^* are given in Table III.

9. Acknowledgement. Acknowledgement is hereby made to Professor H. Hotelling who proposed this problem and gave many helpful suggestions, and to Professor S. N. Roy for suggestions after reading the manuscript.

DIAGRAMS

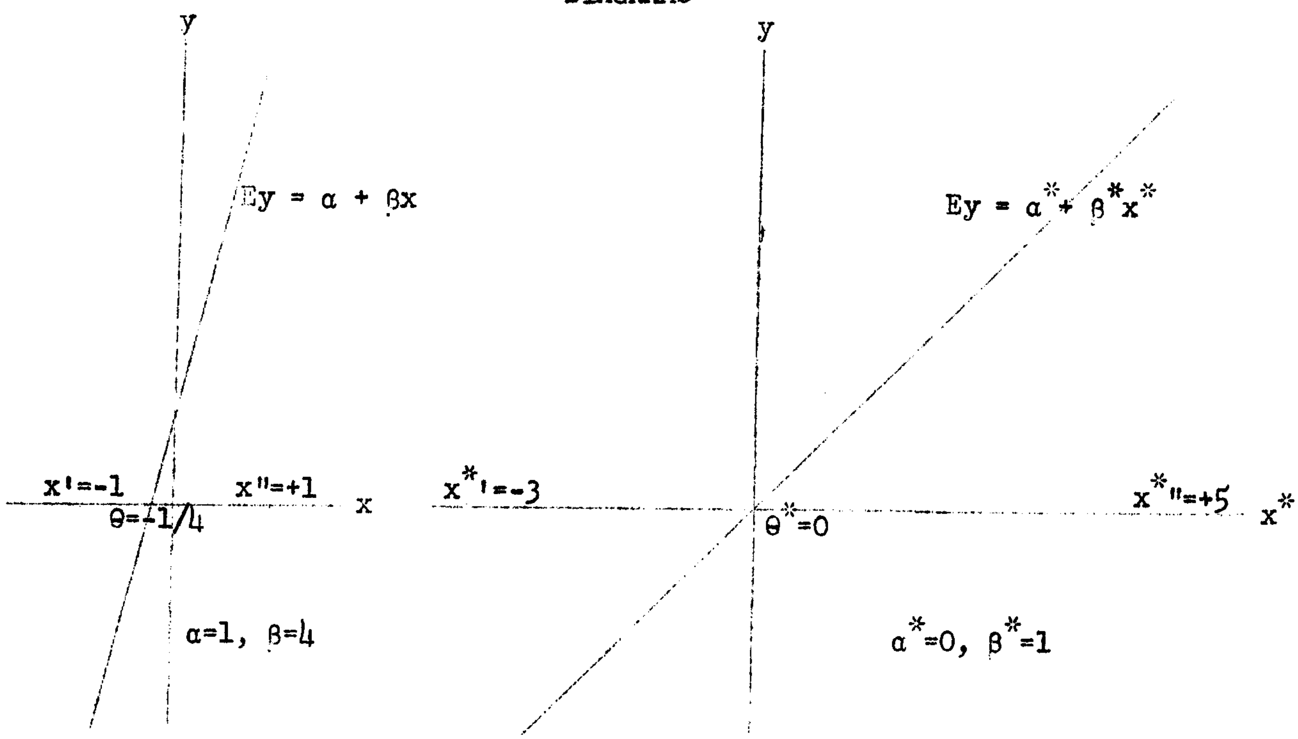

TABLE I

t_N^* (Estimate of $\theta^* = 0$) and Variance Estimates
(after the transformation $x^* = 1 + 4x$)

		Design 1	Design 2	Design 3	Design 4	$\sigma/\sqrt{N}\beta$	$s/\sqrt{N}b$ Design 1	$s/\sqrt{N}b$ Design 4
Sample 1	N=10	+.199	-.071	+.016	+.008	.316	.214	.355
	N=20	+.158	-.070	+.001	-.004	.224	.224	.201
Sample 2	N=10	-.901	-.640	-.619	-.578	.316	.447	.404
	N=20	-.349	-.218	-.239	-.228	.224	.277	.261

TABLE II

Levels of the Experiments (x_i^*, i=1,...,20)
(after the transformation $x^* = 1 + 4x$)

i	Design 1	Design 2	Design 3		Design 4	
			Sample 1	Sample 2	Sample 1	Sample 2
1	-3	-3	-3.000	-3.000	-3	-3
2	-3	+5	+5.000	+5.000	+5	+5
3	-1	-3	-0.803	-4.792	-3	-3
4	-1	+5	+0.598	-0.745	+5	-3
5	+1	-3	+1.827	-1.749	-3	-3
6	+1	+5	+0.793	-0.446	+5	+5
7	+3	-3	-0.347	-2.475	-3	-3
8	+3	+5	-1.840	-1.422	-3	-3
9	+5	-3	+0.922	+2.062	+5	-3
10	+5	+5	-1.828	+1.855	-3	+5
$\bar{x}^*$	+1.000	+1.000	+0.132	-0.571	+0.200	-0.600
11	-3	-3	-1.142	-1.093	-3	-3
12	-3	+5	-0.190	+1.901	-3	+5
13	-1	-3	+1.385	-1.271	+5	-3
14	-1	+5	-0.516	+0.977	-3	+5
15	+1	-3	+0.855	-1.284	+5	-3
16	+1	+5	-0.134	-1.185	-3	-3
17	+3	-3	+0.119	+0.801	+5	+5
18	+3	+5	+0.676	+2.534	-3	-3
19	+5	-3	-0.500	+0.010	-3	-3
20	+5	+5	-0.984	-0.887	+1.745	+4.274
$\bar{x}^*$	+1.000	+1.000	+0.045	-0.260	+0.037	-0.236

TABLE III

Confidence Interval for θ^*
(after the transformation $x^* = 1 + 4x$)

δ	N	Design 2	
		Sample 1	Sample 2
0.95	10	-0.782, +1.027	-0.342, +1.827
	20	-0.375, +0.540	-0.335, +0.823
0.99	10	-1.170, +1.541	-0.772, +2.518
	20	-0.536, +0.724	-0.526, +1.048

REFERENCES

[1] C. R. Rao, Advanced Statistical Methods in Biometric Research, John Wiley and Sons, New York (1952).

[2] Harald Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton (1946).

[3] R. C. Geary, "The frequency distribution of the quotient of two normal variates", Journal of the Royal Statistical Society, Vol. 93 (1930), pp. 442-6.

[4] E. C. Fieller, "A fundamental formula in the statistics of biological assay, and some applications", Quarterly Journal of Pharmacy, Vol. 17 (1944), pp. 117-23.

[5] P. C. Mahalanobis, et.al., "Tables of a random sample from a normal population," Sankyhä, Vol. 1 (1934), pp. 289-328.