t	$B(t)/N$	t	$B(t)/N$
0	.0000	10	.1049
1	.0016	11	.1043
2	.0103	12	.1028
3	.0279	13	.1006
4	.0486	14	.0990
5	.0714	15	.0994
6	.0867	16	.1009
7	.0980	17	.1013
8	.1039	18	.0992
9	.1066	19	.0999
		20	.0993

In conclusion the author wishes to thank Professor S. S. Wilks for various suggestions he has made in connection with this note.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ESTIMATES OF PARAMETERS BY MEANS OF LEAST SQUARES

BY EVAN JOHNSON, JR.

As a criterion for comparing estimates of a parameter of a universe, of known type of distribution, the use of the principle of least squares is suggested. A criterion may be stated in rather general terms. Its application to any given problem presumes a knowledge of the distribution functions of the estimates considered. In the present paper a criterion is set up and application of it is made in the estimation of the mean and of the square of standard deviation of a normal universe.

We shall use the symbol θ to represent a parameter to be estimated. It is to be remembered that θ is a constant throughout any problem, that it represents an unknown value, and that observations and functions of observations (called estimates) are the only variables that occur. We shall use the symbols x_i, $i = 1, 2, \cdots, n$, to represent observed values of the variable x of the universe, and the symbol F to represent a given function of the observations x_i.

If we choose to consider a given function F as an estimate of θ, we are then interested in the error $F - \theta$. This quantity differs from the so-called residual of least square theory, since we are here interested in the difference between computed and true values, rather than in the difference between observed and computed values. To avoid any possible confusion we shall refer to $F - \theta$ as the *error*. Over the set of all samples of n observations, x_i, the distribution of the errors $F - \theta$ is expressed by means of the distribution function $f(F)$, which may be computed from the known distribution function of the universe. We shall assume that the function $f(F)$ has been normalized, so that $\int_\alpha^\beta f(F)\,dF = 1$, where the interval from α to β includes all possible values of F. The integral $I = \int_\alpha^\beta (F - \theta)^2 f(F)\,dF$, associated with a given estimate F, may be thought of as the average square error over the set of all samples.

In this notation we shall state a criterion for the judgment of estimates in either of the two following forms:

DEFINITION 1. *Let f_1 be the distribution function of F_1, and f_2 that of F_2. The estimate F_1 of θ will be judged better than the estimate F_2 if*

$$\int_\alpha^\beta (x - \theta)^2 f_1(x)\,dx < \int_\alpha^\beta (x - \theta)^2 f_2(x)\,dx.$$

DEFINITION 2. *From a given class of functions, of which F is a member, F will be called the best estimate if*

$$I = \int_\alpha^\beta (F - \theta)^2 f(F)\,dF \tag{1}$$

is less than the corresponding integral for all other functions of the class.

It is to be observed that the integral I is a function of the quantities θ and f. From this is seen at once the distinction between the present problem of minimizing the average square error and the similar problem of finding that point around which the mean square value of the deviations of a variable is a minimum. In the problem under consideration we wish to find the function F, or more precisely its distribution function $f(F)$, for which I takes its minimum with a fixed value of θ. In the alternative problem we have a given distribution f and we wish to find the minimum of I with respect to θ.

A second observation to be made is that the integral I can not be usefully minimized in the sense of the general conditions of the calculus of variations. The problem would be of the isoperimetric variety, with the side condition $\int_\alpha^\beta f(x)\,dx = 1$. A solution might be expressed as the limit, as a approaches zero, of functions $f(x)$ with proper continuity conditions, such that

$$f(x) \begin{cases} = 0 \text{ when } |x - \theta| \geqq a, \\ > 0 \text{ when } |x - \theta| < a, \text{ and } \int_{\theta-a}^{\theta+a} f(x)\,dx = 1. \end{cases}$$

Such a solution would be meaningless in practical statistical theory. Solutions are to be expected, therefore, only in those cases where the class of functions, from which F is to be selected, is sufficiently restricted.

The two following examples illustrate both restrictions and possible application of the theory.

As a first example let us consider the problem of finding an estimate F of the mean, $\bar{x}$, of a normal universe. The mean of a distribution is a symmetric linear function of the variates of the distribution. For the class of functions from which to select an estimate F of $\bar{x}$, let us take the class of all symmetric homogeneous linear functions of the observations x_i. Let

$$F = a(x_1 + x_2 + \cdots + x_n). \tag{2}$$

We wish to find the value of a, if any, for which I is a minimum.

F is the sum of n normally distributed independent variables, ax_i, each with standard deviation $a\sigma$. F, therefore, has a distribution function

$$f = C \cdot \exp\left(\frac{-(F - an\bar{x})^2}{2a^2 n\sigma^2}\right),$$

where C is so chosen that $\int_{-\infty}^{\infty} f\,dF = 1$. A discussion of general distribution functions may be found in Dunham Jackson's article, "Theory of Small Samples," in the *American Mathematical Monthly*, Volume XLII, 1935. In this case it can be shown without particular difficulty that

$$\begin{aligned} I &= C \int_{-\infty}^{\infty} (F - \bar{x})^2 \cdot \exp\left(\frac{-(F - an\bar{x})^2}{2a^2 n\sigma^2}\right) dF \\ &= a^2 n\sigma^2 + \bar{x}^2(an - 1)^2. \end{aligned}$$

To determine the minimum of I with respect to a, we set

$$\frac{\partial I}{\partial a} = 2\,an\sigma^2 + 2\bar{x}^2(an - 1)n = 0,$$

and obtain

$$\begin{aligned} a &= \frac{\bar{x}^2}{n\bar{x}^2 + \sigma^2} = \frac{1}{n}\,\frac{1}{1 + \sigma^2/n\bar{x}^2} \\ &= \frac{1}{n}\left(1 - \frac{\sigma^2}{n\bar{x}^2} + \cdots\right). \end{aligned} \tag{3}$$

It is seen that for even such a simple example as the estimation of the mean there is no estimate of the form of equation (2), with a independent of the parameter to be estimated, for which I takes its minimum value.

For a distribution in which $\bar{x} \neq 0$, and $\sigma^2/n\bar{x}^2$ is small, a is given as a first approximation by $1/n$. The function F is merely the mean of the sample observations. If $\bar{x} = 0$, the required solution is $a = 0$, and there is no best least square estimate of the type of equation (2).

In the case where $\sigma^2/\bar{x}^2$ is not small, as is apt to be the case when $\bar{x}$ is near zero, the determination of a desirable estimate by least squares requires a knowledge of the ratio $\sigma^2/\bar{x}^2$, which may perhaps be judged approximately in a special problem. If this value is assumed known, the required value of a may be found most easily by rewriting equation (3) in the form

$$a = \frac{1}{n + \sigma^2/\bar{x}^2}. \tag{4}$$

The second example to be considered is the determination of an estimate of σ^2 of a normal universe. A comparison with the definition of σ^2 suggests the use of a function F given by the equation

$$F = a \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2 \}, \tag{5}$$

where $\bar{x}$ is the mean of the n observations. The value of a is, of course, to be determined by minimizing the integral I.

F is the sum of the squares of n normally distributed but not independent variables. It may be shown, however, (Jackson, *loc. cit.*) to be expressible as the sum of the squares of $n-1$ independent normally distributed variables, each with standard deviation $\sqrt{a}\sigma$. The distribution function for F takes the form

$$f(F) = C\,(F)^{(n-3)/2} e^{-F/2a\sigma^2}, \tag{6}$$

F taking only positive values, and C is again chosen to normalize $f(F)$. The integral I may be written

$$I = C \int_0^\infty (F - \sigma^2)^2 (F)^{(n-3)/2} e^{-F/2a\sigma^2}\, dF.$$

The integration is most easily accomplished by replacing F by u^2, and in terms of u

$$I = C' \int_0^\infty (u^2 - \sigma^2)^2 u^{n-2} e^{-u^2/2a\sigma^2}\, du.$$

The various steps in the integration will differ for even and odd values of n, but in each case the final result is the same. It is found that

$$I = \sigma^4 \{ a^2(n^2 - 1) - 2a(n - 1) + 1 \}. \tag{7}$$

The value of a which minimizes I is determined from the relation

$$\frac{\partial I}{\partial a} = \sigma^4 \{2a(n^2 - 1) - 2(n - 1)\} = 0.$$

Dividing by $(n-1)$, which is not zero in a sample of two or more observations, we obtain

$$a = \frac{1}{n + 1}. \tag{8}$$

In contrast to the previous example we have here an absolute minimum of I with respect to all estimates of the type of equation (5). The best least square estimate of this type is, therefore,

$$F = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n + 1}. \tag{9}$$

PENNSYLVANIA STATE COLLEGE,
STATE COLLEGE, PA.