ESTIMATION BY LEAST SQUARES AND BY MAXIMUM LIKELIHOOD

JOSEPH BERKSON
MAYO CLINIC AND MAYO FOUNDATION*

We are concerned with a functional relation:

$$P_i = F(x_i, \alpha, \beta) = F(Y_i) \tag{1}$$

$$Y_i = \alpha + \beta x_i \tag{2}$$

where P_i represents a true value corresponding to x_i, α, β represent parameters, and Y_i is the linear transform of P_i. At each of $r \geqq 2$ values of x, we have an observation of p_i which at x_i is distributed as a random variable around P_i with variance σ_i^2. We are to estimate the parameters as $\hat{\alpha}$, $\hat{\beta}$ for the predicting equation

$$\hat{p}_i = F(x_i, \hat{\alpha}, \hat{\beta}). \tag{3}$$

By a least squares estimate of α, β is generally understood one obtained by minimizing

$$\sum \frac{1}{\sigma_i^2}(p_i - \hat{p}_i)^2. \tag{4}$$

Although statements to the contrary are often made, application of the principle of least squares is not limited to situations in which p is normally distributed. The Gauss-Markov theorem is to the effect that, among unbiased estimates which are linear functions of the observations, those yielded by least squares have minimum variance, and the independence of this property from any assumption regarding the form of distribution is just one of the striking characteristics of the principle of least squares.

The principle of maximum likelihood, on the other hand, requires for its application a knowledge of the probability distribution of p. Under this principle one estimates the parameters α, β so that, were the estimates the true values, the probability of the total set of observations of p would be maximum. This principle has great intuitive appeal, is probably the oldest existing rule of estimate, and has been widely used in practical applications under the name of "the most probable value." If the p_i's are normally distributed about P_i with σ_i^2 independent of P_i, the principle of maximum likelihood yields the same estimate as does least squares, and Gauss is said to have derived least squares from this application.

In recent years, the principle of maximum likelihood has been strongly advanced under the influence of the teachings of Sir Ronald Fisher, who in a renowned paper of 1922 and in later writings [1] outlined a comprehensive and unified system of mathematical statistics as well as a philosophy of statistical inference that has had profound and wide development. Neyman [2] in a fundamental paper in 1949 defined a family of estimates,

* The Mayo Foundation, Rochester, Minnesota, is a part of the Graduate School of the University of Minnesota.