ON THE EFFICIENCY OF EXPERIMENTAL DESIGNS[1]

By Sylvain Ehrenfeld

Columbia University

1. Introduction. Many models used in statistical investigations can be formulated in terms of least-square theory. The models that will be discussed can be stated as follows. Let $y_1, \cdots, y_N$ be N independently and normally distributed random variables with common variance σ^2. It is assumed that the expected value of y_α is given by

$$E(y_\alpha) = \beta_1 x_{\alpha 1} + \beta_2 x_{\alpha 2} + \cdots + \beta_p x_{\alpha p}, \qquad \alpha = 1, \cdots, N, \tag{1.1}$$

where quantities $x_{\alpha i}$ for $i = 1, \cdots, p$; $\alpha = 1, \cdots, N$ are known constants and $\beta_1, \cdots, \beta_p$ are unknown constants. The coefficients $\beta_1, \cdots, \beta_p$ are the population regression coefficients of y on $x_1, x_2, \cdots, x_p$ respectively. In matrix notation the above model can be expressed as

$$E(y) = X\beta \tag{1.2}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}, \qquad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix},$$

The p column vectors in X will be denoted by $x_1, x_2, \cdots, x_p$ where $x_j' = (x_{1j}, x_{2j}, \cdots, x_{Nj})$. In some cases the experimenter has some amount of freedom in the choice of the p vectors x_i. The efficiency and sensitivity of the design may be very much affected by the choice of the design matrix X. The choice of this matrix is equivalent to that of p vectors in N-dimensional Euclidean space.

A simple illustration is furnished by the following example. Suppose y_α are independent random variables with equal variance σ^2, where $E(y_\alpha) = \beta_1(x_\alpha - \bar{x}) + \beta_2$. The x's are assumed to be fixed constants. Suppose, furthermore, that we have N pairs of observations $(x_1, y_1), \cdots, (x_N, y_N)$ and we want to estimate β_1. It is known that the variance of the least square estimate of β_1 is inversely proportional to $\sum_\alpha (x_\alpha - \bar{x})^2$. Hence, if we could choose values $x_1, \cdots, x_N$ in a domain T, we would choose them such that $\sum_\alpha (x_\alpha - \bar{x})$ is as large as possible.

In Section 2 we will prove a theorem about quadratic forms which, with the aid of other considerations, will motivate a criterion for the efficiency of a design matrix X. In Section 3 two theorems will be proved to aid in the applica-

Received July 3, 1953, revised May 7, 1954.

[1] Work partially sponsored by the Office of Naval Research.