LEAST SQUARES AND BEST UNBIASED ESTIMATES[1]

By T. W. Anderson

Columbia University

1. Introduction. The Gauss-Markov Theorem states that least squares estimates are best linear unbiased estimates. A probability model for the assertion specifies that each observable variable can be written

$$y_\alpha = \sum_{i=1}^{p} \beta_i x_{i\alpha} + v_\alpha, \qquad \alpha = 1, \cdots, n, \tag{1}$$

where $\beta_1, \cdots, \beta_p$ are parameters to be estimated, the set $x_{i\alpha}$ are known numbers, forming a matrix of rank $p(\leqq n)$ and $v_1, \cdots, v_n$ are (unobservable) random variables with means 0, variances σ^2 and are uncorrelated. *Best* means minimum variance among unbiased estimates. In this paper we raise the question of the extent to which the qualification *linear* can be omitted from the statement of the theorem.

We shall assume that the errors in (1), $v_1, \cdots, v_n$, are independently distributed with means 0 and common variances σ^2 and in the first part of the note that they are identically distributed. Then *unbiased* means unbiased identically in the values of $\beta_1, \cdots, \beta_p$ and the common error distribution; *minimum variance* means uniformly with respect to these parameters and the distribution. We consider estimates of every nontrivial linear combination $\sum_1^p \theta_i\beta_i$. The least squares estimate of the linear combination is $\sum_1^p \theta_i b_i$, where

$$(b_1, \cdots, b_p) = b' = \sum_1^n y_\alpha x'_\alpha \left(\sum_1^n x_\alpha x'_\alpha\right)^{-1}$$

and $x'_\alpha = (x_{1\alpha}, \cdots, x_{p\alpha})$. For convenience we assume throughout the paper that the rank of the matrix $(x_{i\alpha})$ is p.

2. Case of identically distributed errors. A particular linear combination of the parameters is $\sum_1^p \beta_i\bar{x}_i = \mu$, say, where $\bar{x}_i = \sum_1^n x_{i\alpha}/n$; this is the expected value of the sample mean $\bar{y} = \sum_1^n y_\alpha/n$. In general a least squares estimate is a linear combination of the observations, say $\sum_1^n c_\alpha y_\alpha$, and each coefficient is a linear combination of the corresponding "independent" variates, say $c_\alpha = \sum_1^p \phi_i x_{i\alpha}$. The sample mean $\bar{y}$ is the least squares estimate of μ if there exist p numbers, $\phi_1, \cdots, \phi_p$, such that $1/n = \sum_1^p \phi_i x_{i\alpha}$. Then the regression function $\mathcal{E}y_\alpha$ can be written $\mu + \sum_1^{p-1} \eta_i w_{i\alpha}$, where $(1, w_{1\alpha}, \cdots, w_{p-1,\alpha})$ is a linear transform of $(x_{1\alpha}, \cdots, x_{p\alpha})$ and $(\mu, \eta_1, \cdots, \eta_{p-1})$ is the inverse linear transform of $(\beta_1, \cdots, \beta_p)$.

Proposition 1. *If $\bar{y}$ is a least squares estimate, it is the best unbiased estimate of $\sum_1^p \beta_i\bar{x}_i$.*

Received July 16, 1961.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034, and Contract Number Nonr-3279(00).