A THEOREM IN LEAST SQUARES

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

NOTATIONS AND PRELIMINARY RESULTS

(0.1) A row vector $\alpha = (a_1, a_2, \ldots, a_m)$ is always represented by a Greek letter and its elements by lower case italics. A matrix (a_{ij}) consisting of n rows and m columns is denoted by a capital letter. Thus $A = (a_{ij})$.

(0.2) The row vectors in A are represented by $\alpha_1, \ldots, \alpha_n$ and the column vectors by $\beta_1, \ldots, \beta_m$. The rank of the matrix A is equal to the number of independent vectors in the α set which is same as that in the β set.

(0.3) Any vector of m elements can be represented by a linear combination of a set of m independent vectors each of m elements.

(0.4) If the number of independent vectors in the set $\alpha_1, \ldots, \alpha_n$ is r then there exist vectors $\gamma_1, \ldots, \gamma_{m-r}$ such that the vector products

$$\gamma_i . \gamma_j = 0, \ i \neq j, \ \gamma_i . \gamma_i = 1$$
$$\alpha_i . \gamma_j = 0 \text{ for all } i \text{ and } j.$$

Similarly there exist vectors $\delta_1, \ldots, \delta_{n-r}$ such that

$$\delta_i . \delta_j = 0, \ i \neq j, \ \delta_i . \delta_i = 1$$
$$\beta_i . \delta_j = 0 \ \text{ for all } i \text{ and } j.$$

(0.5) If $\phi_1, \ldots, \phi_k$ are k independent vectors of m elements then they can be replaced by another set of k independent vectors $\psi_1, \ldots, \psi_s, \psi_{s+1}, \ldots, \psi_k$ obtained as linear combinations of $\phi_1, \ldots, \phi_k$ such that every linear combination of $\psi_1, \ldots, \psi_s$ can be expressed in terms of $\alpha_1, \ldots, \alpha_n$ and no linear combination of $\psi_{s+1}, \ldots, \psi_k$ can be so expressed. The value of $s = k$ minus the rank of the matrix $(\phi_i . \gamma_j)$, where γ's are as defined in (0.4).

These are well known results in vector algebra and they follow from definitions of vector and orthogonality conditions. To these we add two well known results in statistics.

(0.6) A set of linear functions of normally distributed variables is distributed as multivariate normal.

(0.7) If $z_1, \ldots, z_s$ are distributed as

$$\text{const. } e^{-\frac{1}{2}\zeta D^{-1}\zeta'} dz_1 \ldots dz_s$$

where $\zeta = (z_1, \ldots, z_s)$, then $\zeta D^{-1}\zeta'$ is distributed as χ^2 on s degrees of freedom. If $\nu = (z_1, \ldots, z_s, z_{s+1}, \ldots, z_t)$ has the distribution

$$\text{const.}\ e^{-\frac{1}{2}\nu D^{-1}\nu'}\ dz_1 \ldots\ldots dz_t$$

then $\nu D^{-1}\nu' - \zeta D^{-1}\zeta'$ is distributed as χ^2 on $t-s$ degrees of freedom.

THE PROBLEM

Let $\eta = (y_1, \ldots, y_n)$ be the vector of n independent observations from normal populations defined by

$$E(y_i) = \tau . \alpha_i$$

$$V(y_i) = \sigma^2 \text{ independent of } i$$

where $\tau = (t_1, \ldots, t_m)$ is a vector containing m unknown parameters and α's are given vectors. The expectation vector can be simply written as

$$E(\eta) = \tau A'$$

where A' is the transpose of $A = (a_{ij})$. The least value of

$$\Sigma(y_i - \alpha_i . \tau)^2 = (\eta - \tau A').(\eta - \tau A') \qquad \ldots \quad (1.1)$$

when minimised with respect to $t_1, \ldots, t_m$ subject to k independent restrictions

$$\phi_i . \tau = g_i, \quad i = 1, \ldots, k, \qquad \ldots \quad (1.2)$$

is denoted by R^2. The problem is to find the distribution of R^2. The solution is given in an earlier paper (Rao, 1946) where it was shown that R^2/σ^2 is distributed as χ^2 on $(n-r+s)$ degrees of freedom where r is the rank of A and s is as defined in (0.5). Since this result is of fundamental importance in the theory of distributions a simple and an independent proof is presented here. First we note that the value of R^2 remains the same when the restrictions (1.2) are replaced by

$$\psi_i . \tau = h_i, \qquad i = 1, \ldots, k, \qquad \ldots \quad (1.3)$$

where ψ's are defined in (0.5) and h_i are the corresponding linear combinations of g_i. Thus R^2 is the minimum value of (1.1) when t's are subject to (1.3).

THE DISTRIBUTION OF R^2

Since (0.3) is true η can be expressed as

$$\eta = c_1\beta_1 + \ldots + c_m\beta_m + d_1\delta_1 + \ldots + d_{n-r}\delta_{n-r} \qquad \ldots \quad (2.1)$$

where $\delta_1, \ldots \delta_{n-r}$ are defined in (0.4). Also

$$E(\eta) = t_1\beta_1 + \ldots + t_m\beta_m$$

Therefore

$$[\eta - E(\eta)]^2 = d_1^2 + \ldots + d^2_{n-r} + \Sigma\,\Sigma(c_i - t_i)(c_j - t_j)\beta_i.\beta_j$$
$$= d_1^2 + \ldots + d^2_{n-r} + (\xi - \tau)A'A(\xi - \tau)'$$

where $\xi = (c_1, \ldots, c_m)$. This shows that the unconditional minimum is $d_1^2 + \ldots + d^2_{n-r}$ when $\xi = \tau$. Also

$$\eta.\delta_i = c_1\beta_1.\delta_i + \ldots + d_i\delta_i.\delta_i + \ldots = d_i$$
$$E(d_i) = E(\eta.\delta_i) = t_1\beta_1.\delta_i + \ldots + t_m\beta_m.\delta_i = 0$$
$$V(d_i) = V(\eta.\delta_i) = \delta_i^2\sigma^2 = \sigma^2$$
$$C(d_i d_j) = C(\eta.\delta_i, \eta.\delta_j) = \delta_i.\delta_j\sigma^2 = 0.$$

Thus d_i are linear functions of normal variates and R^2_0 the unconditional minimum is the sum of squares of $n - r$ independent normal variates with zero mean and variance σ^2. Therefore R_0^2/σ^2 is distributed as χ^2 on $(n - r)$ degrees of freedom.

Starting with the representation (2.1) and multiplying by $\beta_1, \ldots, \beta_m$ we obtain the equations

$$\eta A = \xi A'A$$

Let λ be such that $\lambda A'A = \psi$, then

$$E(\xi.\psi) = E(\xi A'A\lambda') = E(\eta A\lambda') = \tau A'A\lambda' = \tau.\psi$$
$$V(\xi.\psi) = V(\eta A\lambda') = \lambda A'A\lambda'\sigma^2 = \psi.\lambda\sigma^2$$

Also $$C(\xi.\psi, d_i) = (\delta_i.\lambda A')\sigma^2 = 0.$$

If λ_1, λ_2 are two vectors such that

$$\lambda_1 A'A = \psi_1, \quad \lambda_2 A'A = \psi_2$$

then $C(\xi.\psi_1, \xi.\psi_2) = \lambda_1.\psi_2\sigma^2 = \lambda_2.\psi_1\sigma^2$. Since $\psi_1, \ldots \psi_s$ can be expressed as linear combinations of vectors in A it follows that there exist vectors $\lambda_1, \ldots, \lambda_s$ such that

$$\lambda_i A'A = \psi_i, \; i = 1, \ldots, s.$$

Define $z_i = (\xi.\psi_i - \tau.\psi_i)$, then the dispersion matrix of $\zeta = (z_1, \ldots, z_s) = \sigma^2 D = (\lambda_i.\psi_j)\sigma^2$ in which case

$$\frac{\zeta D^{-1}\zeta'}{\sigma^2} \qquad \ldots \quad (2.2)$$

is distributed as χ^2 on s degrees of freedom independently of the d's.

Now $R^2 = R_0^2 + R^2_1$ where R_1^2 is the minimum value of $(\xi - \tau)\,A'A(\xi - \tau)'$ when t's are subject to the conditions (1.3). Introducing Lagrangian multipliers $l_1, \ldots, l_k$ we consider the expression

$$(\xi - \tau)A'A(\xi - \tau)' - 2l_1\psi_1.\tau - \ldots - 2l_k\psi_k.\tau.$$

The minimising equations are

$$(\xi-\tau)A'A - l_1\psi_1 - \ldots - l_k\psi_k = 0$$

or

$$[(\xi-\tau)A']A - l_1\psi_1 - \ldots - l_s\psi_s = l_{s+1}\psi_{s+1} + \ldots + l_k\psi_k$$

Since no linear combination of $\psi_{s+1}, \ldots, \psi_k$ can be expressed (refer 0.5) in terms of the vectors in A it follows that $l_{s+1} = \ldots = l_k = 0$. Thus the equations reduce to

$$(\xi-\tau)A'A = l_1\psi_1 + \ldots + l_s\psi_s$$

Multiplying by $\lambda_1, \lambda_2, \ldots$ defined above we find

$$(\xi-\tau)A'A\lambda'_1 = z_1 = l_1\psi_1.\lambda_1 + \ldots + l_s\psi_s.\lambda_1$$

$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots$$

$$(\xi-\tau)A'A\lambda'_s = z_s = l_1\psi_1.\lambda_s + \ldots + l_s\psi_s.\lambda_s$$

yielding the solution

$$(l_1, \ldots, l_s) = \zeta D^{-1}$$

The minimum value R^2_1 is $l_1\psi_1.(\xi-\tau) + \ldots + l_k\psi_k.(\xi-\tau) = \zeta D^{-1}\zeta'$. Hence R_1^2/σ^2 is distributed as χ^2 on s degrees of freedom (2.2). We observe that the z's are uncorrelated with the d's so that R_0^2/σ^2 and R_1^2/σ^2 are independently distributed, hence their sum R^2/σ^2 is a χ^2 on $(n-r+s)$ degrees of freedom.

If R_3^2 is the minimum value of (1.1) when the restrictions are increased in number then R_3^2/σ^2 is distributed as χ^2 on $(n-r+t)$ degrees of freedom where $t > s$. Because of (0.7) it follows that $(R_3^2 - R_2^2)/\sigma^2$ is also a χ^2 on $(t-s)$ degrees of freedom independently of R_2^2.

It will be observed that the conditions (1.2) have been replaced by (1.3) only for convenience of proving the result. For practical computations no such step is necessary. Any method of determining the least sum of squares can be followed.

REFERENCE

RAO, C. R. (1946): On the linear combination of observations and the general theory of least squares. *Sankhyā* 7, 237.

Paper received: October, 1950.