EXAMINATION OF RESIDUALS

F. J. ANSCOMBE
PRINCETON UNIVERSITY AND THE UNIVERSITY OF CHICAGO

1. Introduction

1.1. Suppose that n given observations, $y_1, y_2, \cdots, y_n$, are claimed to be independent determinations, having equal weight, of means $\mu_1, \mu_2, \cdots, \mu_n$, such that

$$\mu_i = \sum_r a_{ir}\theta_r, \tag{1}$$

where $\mathbf{A} = (a_{ir})$ is a matrix of given coefficients and (θ_r) is a vector of unknown parameters. In this paper the suffix i (and later the suffixes j, k, l) will always run over the values $1, 2, \cdots, n$, and the suffix r will run from 1 up to the number of parameters (θ_r).

Let $(\hat{\theta}_r)$ denote estimates of (θ_r) obtained by the method of least squares, let (Y_i) denote the fitted values,

$$Y_i = \sum_r a_{ir}\hat{\theta}_r, \tag{2}$$

and let (z_i) denote the residuals,

$$z_i = y_i - Y_i. \tag{3}$$

If A stands for the linear space spanned by $(a_{i1}), (a_{i2}), \cdots$, that is, by the columns of $\mathbf{A}$, and if $\bar{A}$ is the complement of A, consisting of all n-component vectors orthogonal to A, then (Y_i) is the projection of (y_i) on A and (z_i) is the projection of (y_i) on $\bar{A}$. Let $\mathbf{Q} = (q_{ij})$ be the idempotent positive-semidefinite symmetric matrix taking (y_i) into (z_i), that is,

$$z_i = \sum_j q_{ij}y_j. \tag{4}$$

If A has dimension $n - \nu$ (where $\nu > 0$), $\bar{A}$ is of dimension ν and $\mathbf{Q}$ has rank ν. Given A, we can choose a parameter set (θ_r), where $r = 1, 2, \cdots, n - \nu$, such that the columns of $\mathbf{A}$ are linearly independent, and then if $\mathbf{V}^{-1} = \mathbf{A}'\mathbf{A}$ and if $\mathbf{I}$ stands for the $n \times n$ identity matrix (δ_{ij}), we have

$$\mathbf{Q} = \mathbf{I} - \mathbf{AVA}'. \tag{5}$$

The trace of $\mathbf{Q}$ is

$$\sum_i q_{ii} = \nu. \tag{6}$$

Research carried out at Princeton University and in part at the Department of Statistics, University of Chicago, under sponsorship of the Logistics and Mathematical Statistics Branch, Office of Naval Research.