APPENDIX - B

STATISTICAL FORMULAE

PARTIAL AND MULTIPLE CORRELATION COEFFICIENTS

Consider a set of (p+1) variates X_1, X_2,........X_p, X_{p+1} on which n observations have been made. Treating X_{p+1} as the dependent variable and the rest as independent variables, the regression equation is of the form;

$$X'_{p+1} = a+b_1X_1 + b_2X_2 \ldots\ldots\ldots +_pX_p$$

where b_1, b_2, ..., b_p are known as regression coefficients and X'_{p+1} is the predicted value of X_{p+1}.

Defining a new set of variates,

$x_1, x_2, \ldots\ldots x_p, x_{p+1}$ where the variate x_i

represents the deviation from the mean of

X_i, i.e. $x_i = x_i - \bar{x}_i$, the regression equation can be written as ;

$$x'_{p+1} = b_1x_1 + b_2x_2 + \ldots\ldots\ldots + b_px_p$$

The values of b_1 b_2, ..., b_p are obtained by minimising

$$(x_{p+1} - x'_{p+1})^2$$

This leads to the following simultaneous equations known as Normal Equations.

$$b_1 \quad x_1 + b_2 \quad x_1 x_2 + \ldots\ldots + b_p \quad x_1x_p = \quad x_1 x_{p+1}$$

$$b_1 \quad x_2x_1 + b_2 \quad x_2^2 + \ldots\ldots + b_p \quad x_2x_p = \quad x_2x_{p+1}$$

....

....

II

Let S be the p x p matrix (S_{ij}) i, j = 1,2 .....p, where $S_{ij} = x_i x_j$, b the column vector $(b_1, b_2, \ldots, b_p)$ and Y the column $(x_1 x_{p+1}, x_2 x_{p+1}, \ldots, b_p)$. The the above p normal equations can be written in the form of the following matrix relation

$$sb = Y$$

Thus, $b = S^{-1}Y$ where S^{-1} is the inverse of S.

The values of $b_1, b_2, \ldots b_p$ can be obtained from the above.

The simple correlation coefficient r_{ij} between the variable x_i and x_j is given by

$$r_{ij} = \frac{S_{ij}}{\sqrt{s_{ii}\, s_{jj}}}$$

Let C be the Correlation matrix (r_{ij}) and C' its inverse. The partial correlation coefficients are then obtained from

$$r_{(p+1)i\,\cdot\,123\ldots(i-1)(i+1)\ldots p} = \frac{-C'_{(p+1)i}}{\sqrt{C'_{(p+1)(p+1)}\, C_{ii}}}$$

Where C'_{ij} is the element in the matrix C' corresponding to r_{ij} in the matrix C. The significance of a partial correlation coefficient r is tested by calculating 't' from the equation;

$$t = \frac{r\sqrt{n-p-1}}{\sqrt{1-r^2}}$$

The simple correlation between the actual and predicted values of x_{p+1} is known as Multiple Correlation Coefficient R and is given by

$$R^2 = \frac{x'^2_{p+1}}{x^2_{p+1}}$$

A test of significance of the multiple regression or the multiple correlation coefficient is given by calculation of F from the equation

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - p - 1}{p}$$