Optimizing Multi-response Regression model

A.P. Dr. Nawzad. M. Ahmed

Statistics & Information Dept.

Administration & Economic Collage

Sulaimani University

General linear model

$$Y = X\beta + U \quad , U \sim Normal(\mu, \sigma^2)$$

Y : vector of Response variable (single vector) contains n observation(rows), and X is a matrix of explanatory variable (independents, factors), and also U is a vector of random variable is required to be distributed as normal, the form in details is

$$\boldsymbol{y_i = \beta_o + \beta_1\, x_{1i} + \beta_2\, x_{2i} + \ldots\ldots + \beta_k\, x_{ki} + u_i}$$

$$\boldsymbol{u_i} \sim N\ (\mu, \sigma^2)$$

The set of the parameters $\hat{\boldsymbol{\beta}}$=($\boldsymbol{\beta}_o$,$\boldsymbol{\beta}_1$,..,$\boldsymbol{\beta}_k$) can be estimated by ordinary least square (OLS)method as:

$\hat{\boldsymbol{\beta}}=(x'x)^{-1}(x'y)$, $(x'x)$: **fisher inform. matrix**

If the response variable increases from one dependent variable to a set of finite dependent variables, (r, vectors r= 2,3,…….., R), then the multiple regression model might be an another type of regression called **(Multi-response regression model, MRRM),** hence the (**OLS**) is not capable to estimate the parameter matrix of the new model, is called (MRREM).

which was proposed by (Len Beirman, & Freidman 1997) has better performance to detect effects and patterns for the factors(X_1,X_2,,.....X_k)that are introduced to the (MRRM) system of the responses (r.vs) all together. This type of estimation model requires the linear dependency among responses (Yi's) and, independencies(Multi-colinearity)among factors(Xi's).

$$Y_i = Z_i B_i + E_i \qquad \text{----------- (1)}$$

Where:

Y_i is an (Nx1) vector of observation on (i^{th}) response.

Z_i is an (NxP_i) matrix of rank (P_i) with known function of standardized variables.

B_i is a (P_i x 1) vector of unknown constant parameters.

E_i is a random error vector associated with the (i^{th}) response.

Also we assume that:

$E(\epsilon_i) = 0$

$Var(\epsilon_i) = \delta_{ii} I_N, \quad i = 1,2,3,\ldots,r$

$Cov(\epsilon_i, \epsilon_j) = \delta_{ij} I_N \quad i,j = 1,2,3,\ldots,r \quad , i \neq j$ Then the (rxr) matrix whose $(i \text{ , and } j)^{th}$ element is (δ_{ij}) will be denoted by($\sum$).

Also we can rewrite the equation (1) as follow:

$\tilde{Y} = Z B + \epsilon$ ---------- (2)

Where:

$\tilde{Y} = (\acute{Y}_1 : \acute{Y}_2 : \ldots\ldots : \acute{Y}_r)'$

$B = (B'_1 : B'_2 : \ldots\ldots : B'_r)'$

$\epsilon = (\epsilon'_1 : \epsilon'_2 : \ldots\ldots : \epsilon'_r)'$

$Z = diag(Z_1, Z_2, \ldots\ldots, Z_r)$

From equation (2) we can see that (ϵ) has the variance-Covariance matrix:

$$\textbf{Var-Cov}\ (\epsilon) = \sum \otimes I_N \qquad \text{---------- (3)}$$

$$\underline{\hat{B}} = (Z'\Delta^{-1}Z)^{-1} Z'\Delta^{-1}\tilde{Y} \qquad \text{--------- (4)}$$

$\otimes$: Kronecker product matrix. and Equation (2) is a (MRRM).

$\underline{\hat{B}}$ is a matrix of estimators for (MRRM).

From a previous study which was done by the researcher about an agricultural experiment (feeding sweet corn with(3) types of fertilizers (**N, P, and K**) distributed in (3) – blocks.

and recording responses

Y1: average number of leaves.

Y2:average height .

Y3:average circumference.

Y4: average weight of sweet corn.

The multi-response model estimated by the estimated matrix of parameters**(**$\underline{\hat{B}}$**)** as below after standardizing :

$$\underline{\hat{B}} = \begin{bmatrix} -0.36446 & 0.170529 & -0.49591 & 1.008166 \\ 0.318848 & 0.613716 & 0.620585 & 0.281032 \\ -0.23337 & 0.273046 & -0.80172 & 0.187157 \\ -0.0932 & -0.06955 & 1.603784 & -0.76545 \end{bmatrix}$$

first row are estimated parameters for (Y1), second row for (Y2),… and so on. Then we get the origin (MRRM) as given below:

$$\hat{Y}_1 = 14.460 + 0.2520\ Z_1 + 0.0980\ Z_2 + 0.0370\ Z_3$$
$$\hat{Y}_2 = 150.413 + 0.9690 Z_1 + 0.3390\ Z_2 + 0.0530\ Z_3 \qquad \text{------- (5)}$$
$$\hat{Y}_3 = 2.37900 + 0.0370 Z_1 + 0.0120\ Z_2 + 0.0070\ Z_3$$
$$\hat{Y}_4 = 153.564 + 1.0230 Z_1 + 0.5720\ Z_2 + 0.1700\ Z_3$$

The Samuelsson optimization formula was represented by the following (LPS) :

$$optimize\ \ Z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

$$subject\ to\ the\ linear\ constraints$$

$$a_{11}x_1 \ + a_{12}x_2 + \cdots + a_{1n}x_n \ (\leq , \geq)\ b_1$$

$$a_{21}x_1 \ + a_{22}x_2 + \cdots + a_{2n}x_n \ (\leq , \geq)\ b_2$$

$$\text{---}$$

$$. \qquad . \qquad . \qquad\qquad \text{------------} \ (6)$$

$$. \qquad . \qquad .$$

$$. \qquad . \qquad .$$

$$a_{m1}x_1 \ + a_{m2}x_2 + \cdots + a_{mn}x_n \ (\leq , \geq)\ b_m$$

$$x_1 , x_2, \ldots \ldots \ldots , x_n \geq 0$$

This system may be solved by using **simplex method**.

For Example: in order to optimize the average weight of a sweet corn per block, this can be done by (maximizing the regression model for Y4, and taking the remainders of regression models for each Y3, Y2, and Y1 as a constrains:

The first (LPS) is concerned with the average weight of sweet corn flower ($\hat{Y}_4$) per block.

$$\textbf{Maximize } \hat{Y}_4 = 1.023Z_1 + 0.572Z_2 + 0.17Z_3$$

St: $0.252Z_1 + 0.098Z_2 + 0.037Z_3 <= 18.540$

$0.969Z_1 + 0.339Z_2 + 0.053Z_3 <= 72.587$

0.037Z1+0.012Z2+0.007Z3<=3.2210

Z1, Z2, Z3 >= 0

The solution of the system (8) is given by:

Z_2= 189.183685 and zero value for each of (Z_1, and Z_3), by substituting these values in the objective function (Maximize $\hat{Y}_4$) in the system equation (5) we get:

Maximize $\hat{Y}_4$ = intercept of (Regression $\hat{Y}_4$) +0.5720 Z_2

Maximize $\hat{Y}_4$ = 153.564000 + 108.21310 = 261.7771 grams, is the optimal value of weight average for the sweet corn flower per block, after feeding it with an optimal weight (189.183), grams of phosphorus, per block. In the same manner we can optimize each remainder response(Y1, Y2, Y3), by maximizing their regression model from multi-response model system, and taking the remaining regression as a constraint for the (LPS).