

Optimizing Dependent Variables and Multi-Responses Regression Model Development

Asst.Prof: Dr. Nawzad. M. Ahmed

Statistics & Informatics Department - College of Administration & Economy –

University of Sulaimani –Sulaimani -Iraq

nawzadstat@yahoo.com

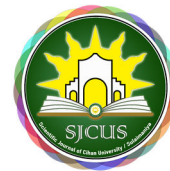
Abstract

This study is concerned as an extension and development of the results for a previous study which was done by the researcher, recommended to optimize the response variables for the (fitted multi-responses regression model, MRRM). This is tried to be done in this study, by building a linear programming systems for each response separately which produced in (MRRM), and solving these systems using linear programming procedure solution methodology named by (Simplex Method). In order to build these systems one can put each response (dependent variable) as an (*objective*) variable that required optimizing, and the independent variables (explanatory variables) are concerned as inputs to construct and building the (*constraints*). Then after solving these systems, it's able to determine the optimal (maximum value for the objectives), and the values of inputs (independent variables), which maximize the objectives.

Keywords: multi-responses regression model (MRRM), linear programming system (LPS), simplex system solution method (SSSM), Response surface methodology (RSM), Linear programming problem (LPP).

ملخص

هذه الدراسة معنية بتوسيع وتطوير النتائج لدراسة سابقة حيث أجريت من قبل الباحث و كانت إحدى توصياتها هي تعظيم قيم متغيرات الاستجابات لمنظومة أنموذج الانحدار متعددات الاستجابة (Multi-Response Regression Model)، وهذه العملية (Optimization) والتي تمت المحاولة لأنجازها في متن هذا البحث، وذلك من خلال بناء منظومات برمجة خطية (Building linear Programming Systems) لكل متغير استجابة بمفردها والمنتجة أصلا من منظومة أنموذج انحدار-متعددات الاستجابة سابقة الذكر، ومن ثم إيجاد الحلول العددية لهذه المنظومات ولكل متغير استجابة بمفردها وتحديد أمثل أقل قيمة (minimum and optimal) للمتغيرات القيود (constraints variables) وذلك باستخدام طريقة سمبلكس واسعة الاستخدام لحل هذه المنظومات في البرمجة الخطية. لأجل بناء منظومات البرمجة الخطية في هذه الدراسة، تم وضع كل متغير استجابة كمتغير دالة الهدف (variable of objective function) المطلوب تعظيم (Maximize) قيمتها بعملية تسمى الأمثلية (Optimization). كما وان المتغيرات المستقلة أو التفسيرية في كل أنموذج تمت استخدامها كمداخلات لمتباينات القيود في كل منظومة وكما مطلوب أيضا إيجاد أقل قيمة عضى لهذه المتغيرات. وبعد إيجاد الحلول العددية للمنظومات يصبح من الممكن تحديد القيم المثلى للمدخلات التي تجعل من متغيرات الهدف تصل الى أعلى قيمها.



پوخته

نهم توپژینه وهیه ههولیکه بوفراوانکردن و پهره پیدانی دهرنه نجامهکانی توپژینه وهیه که پیشتره مان توپژهر نه نجامی داوه که تیایدا داواکاربوه به ههئسان به نه نجامدانی توپژینه وهیه که بویه رزکردنه وهی نرخى ژمارهیی پهرچه گۆروهکان (Maximizing Response Variables) بومۆدیلیک به ناوی (Multi-response Regression Model) که نه توپژینه وهی پیشودا هاتوه .

نهم کرداری (optimization) ه که نهم توپژینه وهیه دا ههولیکه بومۆدراوه به بنیاتنانی سیسته میکی بهرمه جهی هیلی ده بیئت (Building linear programming system) پهرچه گۆراویک به ته نیا که دیاری کراوه نه مۆدیلی ناماژه پیکراو لای سه ره وه، وه دۆزینه وهی شیکاری ژمارهیی بومۆ نهم سسته مانه بومۆ ههر پهرچه گۆرویک به دۆزینه وهی به رزترین و باشتترین (Optimal Maximum) بومۆ نهم گۆراوانه وه ههروه ها دیاری کردنی باشتترین که مترین نرخى ژمارهیی (Optimal Minimum) بومۆ ههریه که نه گۆراوهکانی ناهاوتای ده سته سته کان (Constraints Inequalities Variables) به کارهینانی ریگای ماتماتیکی ناوبراو به (Simplex Method) که به کارهینانی زور به ربلوه نه شیکارکردنی نهم جوړه سسته مانه . بومۆ درست کردنی نهم سسته مانه توپژهر ههر پهرچه گۆرویکى داناهه به گۆراوی نه خشی نامانج (Variable of the Objective Function) که پیویسته بگه یه نرینه به رزترین باشتترین نرخ . هاوکات گۆراوه سه ره به خوکانی ناو ههر مۆدیلیکی داناهه به (Inputs Variables) بومۆ دیاری کردنی ناهاوتاکانی ده سته سته کان بومۆ ههر سسته میک (Constraints Inequalities Variables) که پیویسته نهم گۆراوانه باشتترین که مترین نرخیان بومۆ بدوژرینه وه که وا نه نرخى گۆراوی نه خشی نامانج بکات و باشتترین گه وره ترین (Optimal Maximum) نرخى پییبه خشیئت.

1- Introduction (1, 3, 8):

Response surface methodology (RSM) was introduced first by (Box, and Wilson (1951). The main goal of this methodology is to find the optimum operating conditions, or is to find the set of values of the input variables (factors, or explanatory variables), which result in the most desirable response values. (Hill and Hunter, 1966) made several studies in which multi-responses are investigated.

In multi-responses situation several response variables are considered, and the optimization problem is more complex than in a single response case. The main difficulty is from the fact that two or more responses are under investigation together. It is rarely the case where all responses achieve their respective optima at the same set of conditions, this is because the optimization becomes unclear and not capable since there is no unique way to arrange multivariate values of a multi-response function. In addition to the conditions which are optimal for a unique response may be far from optimal for the other responses. Derringer and Suich (1980) have derived a procedure of simultaneous optimization of several response variables for the same set of input variables. Khuri and Conlon (1981) developed an algorithm for the simultaneous optimization of several response functions that depends on the same of controllable variables and are adequately represented by polynomial regression model for the same degree, here the linear dependencies among the responses must be exist then they are be chosen in developing a function that measures the distance of the vector of estimated responses from the estimated optimum. (Raissi and Frarsani, 2009) are presented a paper on statistical optimization through multi-responses surface methodology to solve more than two responses through the suggested procedure. In this study the optimization of multi-responses regression was done by using the procedure that used by (Anil Kumar Pant & Vinod Kumar, 2011).

2- Optimization of Multi-Responses Regression model ^(3, 4, 5):

In this study we try to apply the technique of (LPS), to solve the (LPP), which is represented by maximizing the responses which were produced in previous study that concerned with fitting a (MRRM) for an agricultural experiment under consideration, thus our target is to find out the amount (doses) of the (Nitrogen, N, phosphorus, P, and Potassium, K) as a factors, for which total yield of each:

Y_1 : The average number of leaves. Y_2 : The average height of corn plant (cm).

Y_3 : The average circumferences. Y_4 : The average weight of sweet corn (grams).

Samuelson defines linear programming as “*The analysis of problems in which linear function of a number of variables is maximized when those variables are subject to a number of constraints in the form of linear inequalities*”.

The general form of a (LPP) with (n) decision variables (factors), and (m) constraints can be stated in the following form:

$$\begin{array}{l}
 \text{optimize } Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \\
 \text{subject to the linear constraints} \\
 \begin{array}{l}
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n (\leq, \geq) b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n (\leq, \geq) b_2 \\
 \text{-----} \\
 \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \\
 \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \\
 \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \quad \cdot \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n (\leq, \geq) b_m \\
 x_1, x_2, \dots, x_n \geq 0
 \end{array}
 \end{array}
 \quad \left. \vphantom{\begin{array}{l} \text{optimize } Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \text{subject to the linear constraints} \end{array}} \right\} \text{----- (1)}$$

This system can be solved by using simplex method.

3-Data Description & Experiment: Recall the detail of the previous study as below:

The plot of land that was experienced upon owns specifics of agricultural land in general and the land area of experiment was (450m²). It is located in the west of Sulaimani, and area called Farmanday-Gshty. In order to apply the experiment, the area was divided into (36) equal blocks. The area of each block was (12m²), and a large area had been chosen to avoid the interaction between the pieces of the experiment units. The factors are (3) different levels of nitrogen (N), (4) different levels of phosphorus (P), and (3) different levels of potassium (K). We took all combinations of three outputs between them which are equal to (36) combinations, each combination consists of three levels of (N, P, and K) was gave to a piece of land, We had to measure *the average number of leaves as (Y_1), average height of plant as (Y_2), average circumference as (Y_3), and the average weight of sweet corn flower as (Y_4).* all per block of the plants. Each experimental block contained (30) plants, the response variables indicates the average of each experimental block, the circumference of each plant was taken from three different points which are *bottom, middle* and *top*, then taking average of them for each



plant in the experimental block. The experience duration time was (60) days. The data that had been collected as described in previously are used to perform appropriate (MRRM), and analysis.

An adequate multi-responses regression model for the above experiment was fitted in the previous study under subhead:

- 1- Testing the linear dependencies among the responses.
- 2- Testing the luck of fit (Luck of Fit:lof) of suggested multi-responses ion model.
- 3- Testing the significance of the fitted model.

Then the suggested regression model was founded as follow ^(1, 2, 8)

Calculating (\hat{B}) by using the mode $\underline{Y} = \underline{w} \underline{\Gamma} + \underline{\delta}$ (2)

Where:

$\underline{Y} = [Y_1 : Y_2 : \dots : Y_r]$, $\underline{w} = [Z_1 : Z_2 : \dots : Z_r]$, $\underline{\delta} = [\epsilon_1 : \epsilon_2 : \dots : \epsilon_r]$, and

$\underline{\Gamma} = [B_1, B_2, \dots, B_r]$, and let $\underline{C} = (C_1, C_2, \dots, C_r)'$ be non-zero $r \times 1$ vector.

(\hat{B}) ($Y_{u1} : Y_{u2} : \dots : Y_{ur}$)' = C , $u=1,2,3,\dots,N$ (3)

B is an $(m \times r)$ matrix of rank $(m < r)$ of constant coefficients

($Y_{u1} : Y_{u2} : \dots : Y_{ur}$) is the (u^{th}) row of the $(N \times r)$ data matrix (Y).

C is an $(m \times 1)$ vector of constants. Then eq.(4) can be shown as :

$B \hat{Y} = 1'_N \otimes C$ (4)

We can detect linear dependencies by Eigen value analysis. Let's suppose that rounding errors in the response values exists and they are distributed independently and uniformly over interval $(-\delta, \delta)$ the quantity (δ) is equal to one half of the last digit reported when all the multi-responses values are rounded to the same number of significant.

The matrix below is the estimated parameters of the four models together (MRRM)*.

$$\hat{B} = \begin{pmatrix} -0.36446 & 0.170529 & -0.49591 & 1.008166 \\ 0.318848 & 0.613716 & 0.620585 & 0.281032 \\ -0.23337 & 0.273046 & -0.80172 & 0.187157 \\ -0.0932 & -0.06955 & 1.603784 & -0.76545 \end{pmatrix} \text{ ----- (5)}$$

(*) The matrix in relation (5) above is the result of a study was done by the researcher ⁽⁹⁾.

The first column of (\hat{B}) matrix is the estimated parameters of the average number of leaves in plant per block, the second, third and fourth columns are the estimated parameters of the average height plant per block, average circumference plant per block and average weight of sweet corn flower of plant per block respectively. Then the multi-response regression model was obtained as:

$$\left. \begin{aligned} \hat{Y}_{1i} &= -0.36446 + 0.318848 Z_1 - 0.23337 Z_2 - 0.0932 Z_3 \\ \hat{Y}_{2i} &= 0.170529 + 0.613716 Z_1 + 0.273046 Z_2 - 0.06955 Z_3 \\ \hat{Y}_{3i} &= -0.49591 + 0.620585 Z_1 - 0.80172 Z_2 + 1.603784 Z_3 \\ \hat{Y}_{4i} &= 1.008166 + 0.281032 Z_1 + 0.187157 Z_2 - 0.76545 Z_3 \end{aligned} \right\} \text{ ----- (6)}$$

Model (6) was obtained after normalizing all variables under consideration, because it's necessary if the values of observations for dependents, and explanatory (factors) are not collected under the same scale. See the following table (descriptive statistics for variables under consideration.

	Y1	Y2	Y3	Y4	Z1	Z2	Z3
Mean	24.13889	184.7056	3.777778	196.9611	26.66667	21.25	23.33333
S.D	5.899892	22.70104	0.90998	25.03274	20.83952	18.41486	20.83952

Renormalizing model (6) to the origin using the normalization equation below:^(1, 4, 5)

$w_i = \frac{2x_i - (x_{il} + x_{ih})}{x_{ih} - x_{il}}$, Where (w_i) , is a standardized observation (i), and (x_{il}, x_{ih}) are the low and high level of (origin data for variables Y_i 's or Z_i 's), $i = 1, 2, \dots, r$. And (r), is the number of response variables which can be measured for each setting of a group of (k) coded variables (w_1, w_2, \dots, w_k) .

Then we get the origin (MRRM) as given below:

$$\begin{aligned} \hat{Y}_1 &= 14.460 + 0.2520 Z_1 + 0.0980 Z_2 + 0.0370 Z_3 \\ \hat{Y}_2 &= 150.413 + 0.9690 Z_1 + 0.3390 Z_2 + 0.0530 Z_3 \\ \hat{Y}_3 &= 2.37900 + 0.0370 Z_1 + 0.0120 Z_2 + 0.0070 Z_3 \\ \hat{Y}_4 &= 153.564 + 1.0230 Z_1 + 0.5720 Z_2 + 0.1700 Z_3 \end{aligned} \quad \text{----- (7)}$$

5-Finding the optimum operating conditions, and building (LPS)⁽³⁾:

In order to optimize a response among $(\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \text{ and } \hat{Y}_4)$, from the system equation (7) let us to say (\hat{Y}_4) , firstly we must maximize the regression model of (\hat{Y}_4) from the multi-response regression models system in equation system (7) as an objective function for the suggested (LPS), and consider the remaining models for all $(\hat{Y}_1, \hat{Y}_2, \hat{Y}_3)$ in (7) without their intercepts as a constraint. The right hand side for these constraints (as it determined from the SLP in equation system (1) are obtained by subtracting the intercept of each remainder regression model from the maximum value of its response. This technique might have repeated in constructing and building (LPS) for each response.

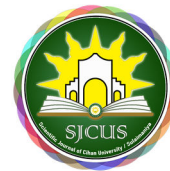
6- Building and Solving (LPS) for the under consideration systems^{(1), (8)}:

In order to solve these (LPS's), the software (Lindo, V6.1) has been used. The first (LPS) is concerned with the average weight of sweet corn flower (\hat{Y}_4) per block.

$$\begin{aligned} \text{Maximize } \hat{Y}_4 &= 1.023Z_1 + 0.572Z_2 + 0.17Z_3 \\ \text{S.t: } 0.252Z_1 + 0.098Z_2 + 0.037Z_3 &\leq 18.540 \\ 0.969Z_1 + 0.339Z_2 + 0.053Z_3 &\leq 72.587 \\ 0.037Z_1 + 0.012Z_2 + 0.007Z_3 &\leq 3.2210 \end{aligned} \quad \text{----- (8)}$$

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

$Z_2 = 189.183685$ and zero value for each of $(Z_1, \text{ and } Z_3)$, by substituting these values in the objective function (Maximize \hat{Y}_4) in the system equation (8) we get:



Maximize $\hat{Y}_4 = \text{intercept of (Regression } \hat{Y}_{4i}) + 0.5720 Z_2$

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.

The 2nd (LPS) is concerned with the average circumference for plant per block (\hat{Y}_3).

Maximize $\hat{Y}_3 = 0.037Z_1 + 0.012Z_2 + 0.007Z_3$

Such that:

$$0.252Z_1 + 0.098Z_2 + 0.037Z_3 \leq 18.540$$

$$0.969Z_1 + 0.339Z_2 + 0.053Z_3 \leq 72.587$$

$$1.023Z_1 + 0.572Z_2 + 0.170Z_3 \leq 90.036$$

The solution of the system (9) is given by:

----- (9)

$Z_3 = 501.081085$, and zero value for each of (Z_1 , and Z_2), by substituting these values in the objective function (Maximize \hat{Y}_3) in the system equation (9) we get:

Maximize $\hat{Y}_3 = \text{intercept of (Regression } \hat{Y}_{3i}) + 0.007 Z_3$

Maximize $\hat{Y}_3 = 2.3790 + 0.007(501.081085) = 5.8454$ cms, is the optimal value of average circumference (Y_3), after feeding it with an optimal weight (501.081085 grams) of potassium (K) per block.

The 3rd (LPS) is concerned with the average height of plant per block (\hat{Y}_2).

Maximize $\hat{Y}_2 = 0.969Z_1 + 0.339Z_2 + 0.053Z_3$

Such that:

$$0.252Z_1 + 0.098Z_2 + 0.037Z_3 \leq 18.540$$

$$0.037Z_1 + 0.012Z_2 + 0.007Z_3 \leq 3.221$$

$$1.023Z_1 + 0.572Z_2 + 0.170Z_3 \leq 90.036$$

The solution of the system (10) is given by:

----- (10)

$Z_1 = 73.571426$, and zero value for each (Z_2 , and Z_3), by substituting these values in the objective function (Maximize \hat{Y}_2) in the system equation (10) we get:

Maximize $\hat{Y}_2 = \text{intercept of (Regression } \hat{Y}_{2i}) + 0.9690Z_1$

Maximize $\hat{Y}_2 = 120.413 + 0.9690(73.571426) = 191.703$ cm is the optimal average height of plant per block, after feeding it with an optimal weight (Nitrogen, N 73.571426 grams) per block.

The 4th (LPS) in (11) is concerned with the average No. of leaves of plant per block (\hat{Y}_1).

Maximize $\hat{Y}_1 = 0.252Z_1 + 0.098Z_2 + 0.037Z_3$

Such that:

$$0.969Z_1 + 0.339Z_2 + 0.053Z_3 \leq 72.587$$

$$0.037Z_1 + 0.0120Z_2 + 0.0070Z_3 \leq 3.221$$

$$1.023Z_1 + 0.572Z_2 + 0.170Z_3 \leq 90.036$$

The solution of the system (11) is given by:

$Z_1 = 65.225044$, $Z_2 = 13.172618$ and $Z_3 = 92.800262$, by substituting these values in the objective function (Maximize \hat{Y}_1) in the system equation (11) we get:

$$\text{Maximize } \hat{Y}_1 = \text{intercept of (Regression } \hat{Y}_1) + 0.2520Z_1 + 0.0980Z_2 + 0.0300Z_3$$

Maximize $\hat{Y}_1 = 14.46 + 0.252(68.48079) + 0.098(13.17261) + 0.030(92.80026) = 36$ leaves approximately, is the optimal average number of leaves for plant per block, after feeding it with an optimal weights ($Z_1 = 65.225044$, $Z_2 = 13.172618$, and $Z_3 = 92.80026$ grams) of Nitrogen (N), phosphorus (P), and Potassium (K) per block, respectively.

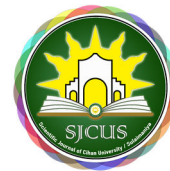
Results Analysis and conclusions:

After optimizing the average weight of the sweet corn flower (\hat{Y}_4), the value of average circumference (\hat{Y}_3), the average height of plant (\hat{Y}_2), and the average number of leaves for plant (\hat{Y}_1) each per blocks. The following important conclusions are presented.

I / during optimizing (\hat{Y}_4), the only factor (Z_2 : phosphorus, P) has a main effect, and the others (Z_1 : Nitrogen, N), and (Z_3 : Potassium, K) have a null effect (zero values). For optimizing (\hat{Y}_3), the factor (Z_3 : Potassium, K) alone is an effective, and the others also have a null effect, and optimizing (\hat{Y}_2) only needs the factor (Z_1 : Nitrogen, N), but optimizing (\hat{Y}_1), needs collection of all (Z_1 : Nitrogen (N), Z_2 : phosphorus (P), and Z_3 : Potassium (K) per block, respectively.

II /from the fact that the factors (Z_1 , Z_2 , and Z_3) are stochastically linearly independent, it is clearly can be seen from optimizing each of (\hat{Y}_4 , \hat{Y}_3 , and \hat{Y}_2), one and only one of these factors has a positive value, and the reminders are zeros. This leads us to say that “optimizing (\hat{Y}_4 , \hat{Y}_3 , and \hat{Y}_2) needs only one of fertilizers (Z_2 , Z_3 , or Z_1) one to one correspondent.

III / concerning with optimizing average number of leaves per block (\hat{Y}_1), the factors (Z_1 , Z_2 , and Z_3) are not stochastically independent, this can be clearly seen that optimizing (\hat{Y}_1), produced under the collection effects of all fertilizers instead of only one of them, then their results, optimizing (\hat{Y}_1), and (Z_1 , Z_2 , and Z_3) are no confidante as they were appeared in optimizing (\hat{Y}_4 , \hat{Y}_3 , and \hat{Y}_2), so its recommended to remove the optimization of (\hat{Y}_1). Or it's recommended to make a sensitive analysis to substitute another weight for more effective one of (Z_1 , Z_2 , and Z_3) instead of reminder to optimize (\hat{Y}_1), also this process can be done if and only if the chemical and agricultural contents for these fertilizers and tests them allowed to do it.



IV- As a recommendation: The results of this study can be generalized and applied on several sectors (productizes, or employee) to reach the economic major goal that is minimizing (cost, time, efforts) and maximizing efficiency.

References:

- 1/ Anil, K. P. and Vinod, K. (2011). Statistical Analysis of Multi-Response Data. Lap Lambert Academic Publishing. Germany.
- 2/ Box, G. E. P, and Draper, N. R. (1978). Statistics for Experiments: An Introduction to Design Data Analysis, and Model Building. New York: John Wiley.
- 3/Khuri, A. I. and Conlon, M. (1981). Simultaneous Optimization of Multiple Response Represented by Polynomial Regression Functions. American Statistical Association and American society for Quality.
- 4/Kumar, V. and Pant, A. K. (2011). Statistical Analysis of Multi-Response Plant Growth Data, Int. J. Agricult, Stat. Sci., Vol. 7, No. 1. 289-300.
- 5/ Liao, H. C. (2004). A Data Envelopment Analysis Method for Optimizing Multi-response problem with Censored Data in The Tagochi- Method, Technical Report, Department of Health Services Administration, Chang Chan Medical University.
- 6/ Mead, R. and Pike, D. J (1975). A Review of Response Surface Methodology from a Biometric Point of View, Biometrics, 31: 803-851.
- 7/Montgomery, Douglas, C. (2005). Design and Analysis of Experiments: Response Surface Method and Designs, John Wiley and Sons, New York.
- 8/Raissi, S., and Farzani, R-Eslami (2009). Statistical Processes through Multi-Response Surface Methodology. World Academy of Science Engineering and Technology, 51.
- 9/ Nawzad, M. A. (2018).” Multi-response Regression Modelling for an Agricultural Experiment ⁽¹⁾”. (Submitted for publishing in Scientific Journal, University of Human Development).